CAP-RNAseq: AN ONLINE PLATFORM FOR RNA-Seq DATA CLUSTERING, ANNOTATION AND PRIORITIZATION BASED ON GENE ESSENTIALITY AND CONGRUENCE BETWEEN mRNA AND PROTEIN LEVELS

A DISSERTATION SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE OF

BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

NEUROSCIENCE

By
MERVE VURAL ÖZDENİZ
April, 2024

CAP-RNAseq: AN ONLINE PLATFORM FOR RNA-Seq DATA CLUSTERING, ANNOTATION AND PRIORITIZATION BASED ON GENE ESSENTIALITY AND CONGRUENCE BETWEEN mRNA AND PROTEIN LEVELS

By Merve Vural Özdeniz

April 2024

We certify that we have read this dissertation and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

Özlen Konu Karakayalı (Advisor)

Michelle Marie Adams

Onur Çizmecioğlu

Öznur Taştan Okan

Ertuğrul Dalgıç

Approved for the Graduate School of Engineering and Science:

Orhan Arıkan

Director of the Graduate School of Engineering and Science

Abstract

CAP-RNAseq: AN ONLINE PLATFORM FOR RNA-Seq DATA CLUSTERING, ANNOTATION AND PRIORITIZATION BASED ON GENE ESSENTIALITY AND CONGRUENCE BETWEEN mRNA AND PROTEIN LEVELS

Merve Vural Özdeniz
Ph.D. in Neuroscience
Supervisor: Özlen Konu Karakayalı
April, 2024

In recent years, there has been a remarkable growth in the application of RNA-seq in both clinical and molecular biology research contexts. The analysis and interpretation of these RNA-seq data demands a good knowledge of bioinformatics. Many different applications are available to perform the analysis, but more comprehensive applications are needed, especially for researchers without coding experience. Therefore, I developed an all-in-one novel RNA-seq analysis tool, CAP-RNAseq (http://konulabapps.bilkent.edu.tr:3838/CAPRNAseq/), which provide valuable analysis for co-expression cluster prioritization and annotation. CAP-RNAseq in particular performs **clustering** of the genes based on their expression patterns, **annotates** mirror clusters that display inverse patterns with a network-based visualizations before **prioritization** of clusters and/or genes based on "gene essentiality", protein levels and the degree of congruence between mRNA and protein levels of genes. Furthermore, for illustration of the use of CAP-RNAseq in this thesis, I reanalyzed a number of published

RNA-seq datasets and identified novel pathways modulated by NTRK2 overexpression (GSE136868) in neural stem cells and also showed significance of the essential genes/pathways in senescent cell clearance focusing on NTRK2 (fibroblast; GSE190998) and THBD (Huh7, GSE228941) siRNA models. In addition, I analyzed our lab's novel RNA-seq data obtained from breast cancer cell lines in CAP-RNAseq; and the findings revealed a) the complex associations between steroid hormones; Drospirenone, Aldosterone, and Estrogen in hormone positive T47D and mineralocorticoid receptor-overexpressing MCF-7 cells; and b) significant differences in essential and non-essential gene expression of the isogenic MCF7 cells overexpressing wildtype or mutant TP53. I also studied a public breast cancer dataset (GSE201085) demonstrating CAP-RNAseq's ability to identify novel breast cancer markers exhibiting high mRNA-protein level correlations. In conclusion, this thesis not only demonstrates the use and power of CAP-RNAseq as a tool to identify essential genes and pathways by analyzing RNA-seq data, but also provides new insights into the roles of essential genes in glioma, senescence and breast cancer.

Keywords: Clustering, Annotation, Prioritization, Transcriptome, RNA-seq, Cancer

Özet

CAP-RNAseq: GEN ESANSİYELLİĞİNE VE mRNA VE PROTEİN DÜZEYLERİ ARASINDAKİ UYUMA DAYALI RNA-Seq VERİ KÜMELEME, ANNOTASYON VE ÖNCELİKLENDİRME İÇİN ÇEVRİMİÇİ BİR PLATFORM

Merve Vural Özdeniz
Ph.D. in Neuroscience
Supervisor: Özlen Konu Karakayalı
Nisan, 2024

Son yıllarda, RNA-seq'in hem klinik hem de moleküler biyoloji araştırma bağlamlarında uygulanmasında kayda değer bir büyüme olmuştur. Bu RNA-seq verilerinin analizi ve yorumlanması, iyi bir biyoinformatik bilgisi gerektirmektedir. Analizi gerçekleştirmek için birçok farklı uygulama mevcuttur, ancak özellikle kodlama deneyimi olmayan araştırmacılar için daha kapsamlı uygulamalara ihtiyaç vardır. Bu nedenle, ortak ifade kümesi önceliklendirme ve annotasyon için değerli analizler sağlayan, hepsi bir arada yeni bir RNA-Seq analiz aracı olan CAP-RNAseq'i (http://konulabapps.bilkent.edu.tr:3838/CAPRNAseq/) geliştirdim. CAP-RNAseq özellikle genlerin ifade profillerine göre kümelenmesini gerçekleştirmekte ve birbiri ile ters ifade profiline sahip kümelerin annotasyonunu network tabalı görseller ile gösterir. Ardından, "gen esansiyelliği", protein seviyeleri ve genlerin mRNA ve protein seviyeleri arasındaki uyum derecesine göre kümelerin ve/veya genlerin önceliklendirilmesini yapmaktadır. Ayrıca, CAP-RNAseq'in bu tezde kullanımını göstermek için, bir dizi

yayınlanmış RNA-seq veri kümesini yeniden analiz ettim ve nöral kök hücrelerde NTRK2 aşırı ekspresyonu (GSE136868) tarafından modüle edilen yeni yolakları belirledim ve ayrıca NTRK2 (fibroblast; GSE190998) ve THBD (Huh7, GSE228941) siRNA modellerine odaklanarak yaşlanan hücre temizliğinde temel genlerin / yolakların önemini gösterdim.

Buna ek olarak, CAP-RNAseq'te laboratuvarımızda meme kanseri hücre hatlarından elde edilen yeni RNA-seq verilerini analiz ettim; ve bulgular a) hormon pozitif T47D ve mineralokortikoid reseptörünü aşırı eksprese eden MCF-7 hücrelerinde steroid hormonları; Drospirenone, Aldosterone ve Estrogen arasındaki karmaşık ilişkileri; ve b) yabanıl tip veya mutant TP53'ü aşırı eksprese eden izojenik MCF7 hücrelerinin temel ve temel olmayan gen ekspresyonundaki önemli farklılıkları ortaya çıkardı.

Ayrıca CAP-RNAseq'in yüksek mRNA-protein seviyesi korelasyonları sergileyen yeni meme kanseri belirteçlerini tanımlama yeteneğini gösteren halka açık bir meme kanseri veri kümesi (GSE201085) üzerinde de çalıştım. Sonuç olarak, bu tez CAP-RNAseq'in RNA-seq verilerini analiz ederek temel genleri ve yolakları tanımlamak için bir araç olarak kullanımını ve gücünü göstermekle kalmıyor, aynı zamanda glioma, senesans ve meme kanserinde temel genlerin rolleri hakkında yeni bilgiler sağlıyor.

Anahtar Kelimeler: Kümeleme, Annotasyon, Önceliklendirme, Transkriptom, RNA-seq, Kanser

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude to my Ph.D. supervisor, Ozlen Konu, for her unwavering support and guidance throughout the course of this research. Her expertise and insightful criticism were invaluable not only for this thesis but also for my academic development.

My sincere thanks also go to the members of my TIK committee, Professor Michelle Adams and Dr. Öznur Taştan, for their valuable insights and detailed comments on my project. I would like to thank Dr. Onur Çizmecioğlu and Dr. Ertuğrul Dalgıç for their time and efforts in reviewing my thesis and participating in my defense committee. I would also like to thank Dr. Serap Erkek and her lab members for hosting me in their lab at Izmir Biomedicine and Genome Center (IBG) and teaching me ChIP procedures.

I would like to thank all the formers Konu lab members. I am especially grateful to Kübra Çalışır who has not only been a colleague but also a great friend. Her support and camaraderie have been invaluable to me. I would also like to thank Seniye Targen, Ronaldo Leka, Farid Ahadli, Melike Tombaz and Siber Güneş Kılınç for their support and help in my studies and for what they taught me. Additionally, I am grateful to Burcin Irem Arıcı and Güneş Tok for their friendship and contributions to my project.

My heartfelt appreciation goes to all current Konu lab members, Rana Acar, Mert Demirdizen, Moiz Aftab, Muazzez Çelebi Çınar, Seda Aktürk, Zehra Fırat, Nil Gürgül, Algı Demirtaş, Betül Metin and Emel Ahmad. My special thanks also go to Rana Acar and Mert Demirdizen, who began as interns under my mentorship and have become both colleagues and friends. Their constant support and willingness to help throughout my

research and their friendship were invaluable to me and I am so grateful to have them in my life.

I would like to thank to my family, Gülsüm, İsmet, Aysun, Arzu and my little brother Sefa who always makes me laugh for their love and support by all means. I reserve my deepest gratitude and love for my husband, Engin Özdeniz, who has been my faithful partner throughout this journey. His constant support, patience and encouragement have sustained me through the most challenging times. His belief in my capabilities and his constant reassurance have lifted me up, allowing me to focus on my research with a clear mind and a calm heart. I thank him so much for understanding and putting up with all my ups and downs. I am incredibly lucky to have such a supportive and loving partner by my side. I am also extremely grateful that the joy in our lives is about to double as we are expecting twins. This incredible personal milestone has given me even more motivation and joy in these times.

I would like to thank Bilkent University Neuroscience and MBG departments for providing us such an excellent academic opportunity. This study has been funded by a research grant from TUSEB (Project no: 4405 to OK) and also received funding from the European Horizon's research and innovation program HORIZON-HLTH-2022-STAYHLTH-02 under agreement No 101095679 (to OK), funded by the European Union. I also received a scholarship from TUBITAK 219Z029 (to OK) in the context of TP53 wildtype mutant RNA-seq data analysis.

Finally, I would like to extend my appreciation to all those who were involved in any way in this work and contributed to my academic journey.

To my precious husband and children

Table of Contents

List of Figures xvi

List of Tables xxii

CHAPTER 1 1

INTRODUCTION 1

1.1 Deciphering Transcriptomes with RNA-Sequencing 1

1.1.1 RNA-seq Basics 2

1.1.2 Navigating the RNA-seq Landscape: Challenges and Opportunities 7

1.1.3 The Significance of Clustering in RNA-seq Analysis 9

1.1.4 Methods of Clustering Gene Expression Data 11

1.1.5 Gene Selection Methods and Their Impact on Clustering Performance 15

1.1.6 Annotation of Clusters 18

1.1.7 DepMap and its Potential as an Annotation Tool 19

1.1.8 Genomic Exploration of Essential Genes 23

1.1.9 Correlation Between the mRNA and Protein Levels in the Cell 27

1.1.10 Prioritization of Genes 28

1.1.11 Online RNAseq Analysis Tools 34

1.1.11.1 iDEP 34

1.1.11.2 GENAVi 35

1.1.11.3 DEBrowser 36

1.1.11.4 RNfuzzy 36

1.1.11.5 BEAVR 37

1.1.11.6 WebMeV 37

1.1.11.7 3D RNA-seq 38

1.1.11.8 Omics Playground 38

1.1.11.9 FungiExpresZ 39

1.2 Senescence and Gene Essentiality 43

1.3 Essential Genes and Cancer in General 46

1.3.1 Essential Genes in Brain Cancers and Brain Stem Cells 48

1.3.2 Essential Genes in Breast Cancer 50

1.4 Brain Cancers 52

1.4.1 Glioblastoma Multiforme (GBM) 54

1.4.2 Oncogenic Signaling 56

1.4.3 Neurotrophin Receptor Splicing in GBM 57

1.5 Breast Cancer and Steroid Hormones 60

1.5.1 Steroid Hormone and Nuclear Receptor Signaling 62

1.5.1.1 Steroid Hormones 62

1.5.1.1.1 Estrogen 62

1.5.1.1.2 Progesterone 64

1.5.1.1.3 Aldosterone 69

1.5.1.1.4 Combination of Drsp with E2 as Treatment 72

1.5.1.2 Nuclear Receptor Signaling 74

1.5.1.2.1 Estrogen Signaling and Its Relationship with Breast Cancer 75

1.5.1.2.2 Progesterone Signaling and Its Relationship with Breast Cancer 80

1.5.1.2.3 MR Signaling and Its Relationship with Breast Cancer 86

1.5.1.2.4 Drsp and Its Associations with MR, ER, and PR in Breast Cancer 89

Aims and Rationale 92

CHAPTER 2 98

MATERIALS AND METHOD 98

2.1 Development of CAP-RNAseq R Shiny Application 98

2.1.1 Publicly Available Datasets Integrated into CAP-RNAseq 100

2.1.2 Data Processing in CAP-RNAseq 102

2.1.3 Clustering in CAP-RNAseq 103

2.1.4 Mirror Clusters 104

2.1.5 Cluster Prioritization 104

2.1.6 Differential Gene Expression and Functional Analysis 107

2.1.7 Gene Prioritization 107

2.1.8 Search a Gene/Geneset 108

2.1.9 Primer Design 109

2.1.10 Publicly Available Demo Datasets 109
2.1.11 Web Server Construction 111
2.2 Hormone Studies in Breast Cancer 113
2.2.1 Materials 113
2.2.1.1 Solutions, Kits and Reagents Used in the Experiments 113
2.2.1.2 Equipments 117
2.2.1.3 Plasmids 118
2.2.2 Methods 118
2.2.2.1 Cell Culture Experiments 118
2.2.2.2 Transfection of Cells with MR 118
2.2.2.3 Hormone Treatments 119
2.2.2.4 Collection of Cell Pellets 120
2.2.2.5 RNA Extraction 121
2.2.2.6 cDNA Synthesis 122
2.2.2.7 qRT-PCR Protocol 122
2.2.2.8 Western Blot 124
2.2.2.9 RNA Sequencing Procedure 128
2.2.2.10 Statistical Analysis 128
CHAPTER 3 130
RESULTS 130
3.1 General Architecture of the CAP-RNAseq App 130
3.1.1 CAP-RNAseq Demonstration Using Demo 1 Dataset 130
3.1.2 Gene Set Annotation of Demo 1 133
3.1.3 Essential Gene Analysis of Demo1 Dataset 137
3.1.4 Analysis of Correlation Between mRNA-Protein Levels of Demo 1 Dataset 140
3.1.5 Gene Prioritization for Demo1 Dataset 141
3.1.6 Protein Expression of Prioritized Genes in Demo1 Data Using HPA Module 144
3.2 Analysis of Senescence and “Essential Genes” 147
3.2.1 Analysis of Specific Clusters in Terms of Essentiality 150

3.2.2 Gene Set Annotation for Essential and Non-Essential Genes 152

3.2.3 Specific Analysis of Cluster 2 153

3.2.4 The Effect of NTRK2 Inhibition in Senescent Cells 156

3.2.5 Cross-Dataset Mapping 157

3.3 Essential Genes and Recurrence Free Breast Cancer 159

3.3.1 Gene Set Annotation for All, Essential and Non-essential Genes 160

3.3.2 Possible Biomarkers for the Recurrence 163

3.4 Drospirenone, Estrogen, Aldosterone Driven Expression Changes in T47D Cells Analyzed by Mirror Clusters 166

3.4.1 RNA-seq Quality 166

3.4.2 Analysis of T47D Data in CAP-RNAseq 168

3.4.3 Mirror Clusters and Essentiality Analysis of Clusters 169

3.4.4 Annotation Networks of Genes Based on Essentiality in Mirror Clusters 171

3.4.5 Identification of Genes Affected by Drsp in Mirror Clusters 4 and 9 173

3.4.6 Synergistic Effect of Drsp and E2 in Mirror Clusters 5 and 11 180

3.4.7 E2 Related Genes in Mirror Cluster 2 and 6 185

3.4.8 Essential Genes Downregulated by E2 Treatment 188

3.4.9 Genes Affected by DEA But Not Aldo Alone in Mirror Clusters 1 and 8 190

3.4.10 Genes Affected by D_E2 and DEA But Not Solo Treatments in the Mirror Cluster 7 and 10 194

3.5 Hormone Driven Expression Changes in MCF-7 Cells 199

3.5.1 RNA-seq Quality 199

3.5.2 Analysis of MCF-7 Data in CAP-RNAseq 200

3.5.3 Gene Set Enrichment Analysis of All Clusters 203

3.5.4 Genes Affected by Drsp 205

3.5.5 Genes Affected by Aldosterone 209

3.5.6 Genes Affected by E2 211

3.5.7 Aldosterone effect on D_E2 Treatments 214

3.5.8 Cross-Dataset Mapping for the Genes Increased by Drsp in T47D Data . 215

3.6 Hormone Studies in Breast Cancer Cell Lines 217

3.6.1 Investigation of the Effect of Drsp in T47D Cells 218

3.6.2 Validation of Genes 222

3.6.2.1 qRT-PCR Results in MCF-7 Cells 224

3.6.2.2 Western Blot Results in MCF-7 Cells 227

3.6.3 Comparison of Genes in both MCF-7 and T47D Cell Lines 229

3.6.4 ChIPseq Analysis 233

3.7 Ongoing Studies 241

3.7.1 Investigation of Associations Between Senescence and Liver Fibrosis and Finding Markers 241

3.7.2 Understanding the Involvement of Essential Genes in TP53 Mutant Breast Cancer Cells 244

CHAPTER 4 247

DISCUSSION, CONCLUSION AND FUTURE PERSPECTIVES 247

4.1 A Summary of Novelties Brought by CAP-RNAseq 247

4.2 Novel Associations Discovered by CAP-RNAseq for NTRK2 Signaling in Neural Stem Cells and Fibroblasts 251

4.3 Senescence and Essential Gene Relationship 254

4.4 The Effect of NTRK2 Inhibition 258

4.5 Genes Affecting the Recurrence of Cancer and the Identification of Biomarkers 261

4.6 Intricate Crosstalk between Hormones in T47D Cells 263

4.6.1 Pathway Enrichment of Essential and Non-essential Genes 263

4.6.2 Genes Affected by Drsp Alone or in Combination with Another Treatment 267

4.6.3 Synergistic Effects of Drsp and E2 271

4.6.4 Prioritized Genes and Pathways Affected by E2 274

4.6.5 Genes Affected by DEA 276

4.6.6 Effect of Aldosterone to Hormone Therapy 278

4.6.7 Genes Affected by D_E2 and DEA But Not D, E2, or Aldo 281

4.7. Effects of Steroid Hormones in MR-Overexpressing MCF-7 Cells 284

4.7.1 Essentiality and Enrichment of the Clusters 284

4.7.2 Prioritized Genes Modulated by Drsp 286

4.7.3 Prioritized Genes Modulated by Aldosterone 288

4.7.4 Prioritized Genes Modulated by E2 289

4.7.5 Cross-Dataset Mapping for the Genes Modulated by Drsp 290

4.8 Evaluating the Interactions of Drsp, Aldo, and E2 in Breast Cancer Cell Lines by Wet-Lab Experiments 291

4.9 Modulatory Effects of E2 and Drsp on Estrogen Receptor Binding Sites in Breast Cancer 295

4.10 Conclusions 296

4.11 Future Perspectives 299

Appendix A. Assessing the Transient Transfection Stability and Duration 301

Appendix B. MTT Assays 303

Appendix C. The Effect of High-dose Drsp and MR Overexpression in the Presence of Aldo 311

Appendix D. Investigation of the Effect of Drsp Doses in T47D Cells 315

Appendix E. Chromatin Immunoprecipitation Sequencing (ChIP-Seq) 317

Appendix F. Analysis of GSE74620 dataset 324

REFERENCES 325

COPY RIGHT PERMISSIONS 390

List of Figures

Figure 1. Flow chart of RNA-seq analysis..... 6
Figure 2. Mechanisms of cellular senescence and its consequences. 44
Figure 3. Structure of drospirenone and spironolactone. 67
Figure 4. Mechanisms for estrogen signaling. In the graph, estrogen, or E2 (orange circle E), binds to ERα/ERβ and GPER1 to exert its genomic and non-genomic effects. Orange highlights the genomic impact, which is caused by the E2-receptor complex binding to EREs as it enters the nucleus. The non-genomic impact is displayed in blue. E2 binds to ERs in the membrane-like GPER1 and controls expression by adjusting the activation of associated enzymes or the opening of ion channels. E; estrogen or E2, ERα; estrogen receptor alpha, ERβ; estrogen receptor beta, GPER1; G protein-coupled estrogen receptor 1, ERE; estrogen response elements. 77
Figure 5. Classical and non-classical signaling of progesterone. A) Classical signalling of progesterone is mediatied by the diffusion of hormone through membrane and binding to receptor inside the cell. B) Rapid response in non-classical mechanism is mediated through the activatiton of c-Src, MAPK and AKT. C) The non-classical signaling can be initated by the binding of progesterone to mPRs and PGRMC1. 83
Figure 6. The workflow of CAP-RNAseq. 99
Figure 7. Analysis for finding the optimal number of clusters. A) Cluster dendrogram of Hierarchical Clustering cut for 8 clusters. B) The table shows the number of well-clustered clusters at each cluster number after performing a silhouette analysis for cluster numbers from 4 through 15. C) The graphical visualization of average silhouette width for cluster numbers until 15. 131
Figure 8. Expression patterns in GSE136868 presented by k -means clustering. The cluster centroid pattern is shown by red lines. 132
Figure 9. Dissimilarity heatmap and a table for mirror clusters. A) Heatmap shows the correlation values, in which negative values (blue) indicate the lowest similarities and positive values (red) indicate the highest similarities. B) The mirror clusters are shown in a table to be easily identified. 133
Figure 10. Analysis of clusters 1, 2, 6 and 8. A-B-C-D) Expression pattern of clusters. E) The network visualization of enrichments of genes in each cluster using Hallmark gene set collection of MSigDB. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between TrkB.T1 and TrkB.FL samples. 135
Figure 11. Analysis of clusters 3, 4, 5 and 7. A-B-C-D) Expression pattern of clusters. E) The network visualization of enrichments of genes in each cluster using Hallmark gene-set collection of MSigDB. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between TrkB.T1 and TrkB.FL samples. 136
Figure 12. Analysis of cluster 1 and cluster 2 along with their mirror clusters 8 and 6, respectively. Each graph displays the log2 of average CPM values on the x-axis and CV of CPM values on the y-axis of genes of cluster 1 (A), cluster 8 (C), cluster 2 (E), and cluster 6 (G). The size of each dot displays a gene's log2 of range CPM value. The median CV for both essential and non-essential genes is displayed using the Mann Whitney U test in cluster 1 (B), cluster 8 (D), cluster 2 (F), and cluster 6 (H). 139
Figure 13. Correlation result between mRNA-protein expression values of genes of cluster 6 in glioblastoma disease. 141

Figure 14. Expression pattern and mRNA ~Protein correlation of each gene. (A, C, D) Graphs showing the normalized log2 expression of genes in each sample. (B, D, F) mRNA ~Protein correlation of each gene in all lineages using DepMap data. 143
Figure 15. Protein expression levels of CCND2 (A, B), TNFSF10 (C) and AK4 (D, E) in all cancers and normal tissues using the HPA dataset. 145
Figure 16. The correlation of NTRK2 mRNA expression with CCND2 expression in LGG and GBM patients. 147
Figure 17. Expression patterns in GSE190998 presented by k -means clustering. The cluster centroid pattern is shown by red lines. 148
Figure 18. Dissimilarity heatmap and a table for mirror clusters in Demo 2 data. A) Heatmap shows the correlation values, in which negative values (blue) indicate the lowest similarities and positive values (red) indicate the highest similarities. B) The mirror clusters are shown in a table to be easily identified. 149
Figure 19. Examination of cluster 2 and cluster 5. Each graph presents the log2 of average CPM values on the x-axis and CV of CPM values on the y-axis for the genes of cluster 2 (A) and cluster 5 (C). The log2 of range CPM values are displayed by the size of the dots. The median CV is displayed using the Mann Whitney U test for both essential and non-essential genesin cluster 2 (B) and cluster 5 (D)..... 151
Figure 20. MSigDB networks for Hallmark gene set collection. Network visualization with enriched terms having q-values less than 0.05 for essential genes (A) and non-essential genes (B) in the clusters. The number of genes that overlap with each gene set collection was shown by edge widths. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between siCtrl_S and siCtrl_P samples. 153
Figure 21. Expression plots of the CCND2 gene in each sample of Demo 1 data (A) and Demo 2 data (B)..... 155
Figure 22. MSigDB networks for Hallmark terms having q-values less than 0.05 for all genes. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between siNTRK2 and siCtrl_S samples. .. 157
Figure 23. Cross-comparison of gene expression from Demo 1 and Demo 2 datasets. A) The Demo 2 dataset's genes from cluster 5 were searched and mapped to the Demo 1 analysis. B) Gene expression patterns within cluster 6 and C) within cluster 8 of the Demo 1 dataset. 158
Figure 24. K-means clustering of GSE201085 data using group means values and presented as box plot. Red lines indicate the pattern of clusters. 160
Figure 25. MSigDB analysis of clusters for Hallmark gene set collection with network visualization of enriched terms having q-values less than 0.05 for all genes. Red nodes indicate the increase in gene expression in pCR samples compared to RDR samples. 161
Figure 26. MSigDB networks for Hallmark gene set collection. Network visualization with enriched terms having q-values less than 0.1 for essential genes (A) and non-essential genes (B) in the clusters. Red nodes indicate the increase in gene expression in pCR samples compared to RDR samples, whereas blue node indicate a decrease. 162
Figure 27. Analysis of IL1R2 gene. A) Plot of scaled logCPM values of IL1R2 gene against its cluster centroid. B) Correlation plot for mRNA and protein levels in breast cancer obtained from DepMap data. C) Protein levels of IL1R2 gene in all cancers by HPA. D) Protein level of IL1R2 gene in normal breast tissues. 164
Figure 28. Analysis of ENPP5 gene. A) Plot of scaled logCPM values of ENPP5 gene against its cluster centroid. B) Correlation plot for mRNA and protein levels in breast cancer obtained from DepMap data. C) Protein levels of IL1R2 gene in all cancers by HPA. D) Protein level of ENPP5 gene in normal breast tissues..... 165

Figure 29. PCA result of two RNA-seq data after combat. 168
Figure 30. k-means clustering result of 1,571 genes in T47D RNA-seq data with 12 clusters. The cluster centroid pattern is shown by red lines. 169
Figure 31. Dissimilarity heatmap for 12 clusters to find the mirror clusters. A) Heatmap visualization. B) Table visualization of mirror clusters..... 170
Figure 32. MSigDB networks for Hallmark gene set collection. Network visualization with enriched terms having q-values less than 0.05 for essential genes (A) and non-essential genes (B) in the clusters. The number of genes that overlap with each gene set collection was shown by edge widths. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between D_E2 and control samples. 172
Figure 33. Cluster 9 and 4. A) Expression pattern of cluster 9. B) Expression pattern of cluster 4. C) MSigDB analysis of cluster 9 using Hallmark and REACTOME gene sets with q-value< 0.1 for top 7 terms. D) Scatter plot showing the log2 of average CPM in the x-axis and coefficient of variation of CPM in the y-axis of genes in cluster 9. The size of the dots shows the log2 of range CPM..... 174
Figure 34. Log2 expression graphs and HPA results of PLLP (A-B-C), FKBP5 (D-E-F) and MYO10 (G-H-I) genes..... 177
Figure 35. Log2 expression graphs and HPA results of SOX2 (A-B-C), and CLU (D-E-F) genes..... 180
Figure 36. Expression patterns of mirror clusters 5 (A) and 11 (B). 181
Figure 37. Analysis of clusters 5 and 11. A) MSigDB analysis using Hallmark and REACTOME gene sets with q-value< 0.1 for top 10 terms. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between D_E2 and control samples. B) Scatter plots for cluster 5 and C) cluster 11..... 183
Figure 38. Log2 expression graphs and HPA results of FOXA1 (A-B-C), and FKBP4 (D-E-F) genes. 184
Figure 39. Expression patterns of mirror clusters 2 (A) and 6 (B). 185
Figure 40. MSigDB analysis of clusters 2 and 6 using Hallmark and REACTOME gene sets with q-value< 0.1 for top 10 terms. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between D_E2 and control samples. 186
Figure 41. Log2 expression graph and HPA results of PDZK1 (A-B-C) and NDRG1 (D-E-F). 187
Figure 42. Analysis of cluster 3. A) Visualization of the expression patterns of cluster 3. B) Scatter plot showing the Selectivity values in the x-axis and log2 of average CPM in the y-axis of genes in cluster 3. The size of the dots shows the coefficient of variation of CPM..... 188
Figure 43. Log2 expression graph and HPA results of ERBB2 (A-B-C)..... 190
Figure 44. Expression patterns of mirror clusters 1 (A) and 8 (B). 191
Figure 45. Log2 expression graphs and HPA results of BCL2L1 (A-B-C)..... 192
Figure 46. Log2 expression graph and HPA results of CEP85 (A-B-C)..... 194
Figure 47. Expression patterns of mirror clusters 7 (A) and 10 (B). 194
Figure 48. MSigDB analysis cluster 7 and 10 using Hallmark and REACTOME gene sets with q-value< 0.1 for top 10 terms. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between DEA and control samples. 195
Figure 49. Log2 expression graphs and HPA results of RUNX1 (A-B-C) and ARF6 (D-E-F) genes..... 197

Figure 50. Log2 expression graphs and HPA results of CITED2 (A-B-C), and TPM4 (D-E-F) genes. 198
Figure 51. PCA result of MCF-7 RNA-seq data. 200
Figure 52. k-means clustering result of 1,994 genes in MCF-7 RNA-seq data with 16 clusters. Red lines indicating the cluster centroids show the cluster pattern. 201
Figure 53. Dissimilarity heatmap for 16 clusters to find the mirror clusters. A) Heatmap visualization. B) Table visualization of mirror clusters. 201
Figure 54. MSigDB analysis using Hallmark gene sets with q-value< 0.05 for top 10 terms. A) Only essential genes were used. B) Only non-essential genes were used. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between D_E2 and control samples. 205
Figure 55. Analysis of clusters 4 and 11. Visualizations of cluster 4 (A) and cluster 11 (B). Scatter plots displaying the log2 of average CPM values in the x-axis and CV of CPM in the y-axis of genes in cluster 4 (C) and cluster 11 (D). The size of each dot displays a gene's log2 of range CPM value. 206
Figure 56. Log2 expression graphs and HPA results of CHI3L2 (A-B-C), ENO3 (D-E-F), CYP1B1 (G-H-I), and TXNRD1 (J-K-L) genes. 208
Figure 57. Log2 expression graphs and HPA results of RELB (A-B-C), PREX1 (D-E-F) and BCL6 (G-H-I) genes. 210
Figure 58. MSigDB analysis using Hallmark, REACTOME and GO gene sets with q-value< 0.1 for top 7 terms. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between D_E2 and control samples. 212
Figure 59. Log2 expression graphs and HPA results of CD44 (A-B-C) and PGR (D-E-F) genes. 213
Figure 60. Analysis of mirror clusters, 11 and 5. A) The expression pattern of clusters 11. B) The expression pattern of clusters 5. C) MSigDB analysis for essential genes using Hallmark gene sets with q-value< 0.05 for top 10 terms. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between D_E2 and control samples. 214
Figure 61. Cross dataset mapping. A) Visualization of cluster 9 of T47D data. B) Visualization of cluster 11 of MCF-7 data. C) The result of the cross-mapping when the genes of cluster 9 in the T47D data were searched in the MCF-7 data. 216
Figure 62. qRT-PCR results with RNAs from T47D cells. A) SGK1, B) MR, C) S100P and D) CXCL12 genes. One-way ANOVA with Tukey's multiple comparisons test was used (*: p<0.05, **: p<0.01, ***: p<0.001, ****:p<0.0001). 220
Figure 63. Checking the expression of PGR, ZBTB16 and S100P genes by qRT-PCR with RNAs from T47D cells. One-way ANOVA with Tukey's multiple comparisons test was used(*: p<0.05, **: p<0.01, ***: p<0.001, ****:p<0.0001). 221
Figure 64. qRT-PCR results with RNAs collected from MCF-7 cells transfected with OV alone and exposed to 5 different drug combinations. (*: p<0.05, **: p<0.01, ***: p<0.001, ****:p<0.0001). ANOVA results are as follows: F(5,6)= 8.77, p=0. 00994 for PGR; F(5,6)= 1.184, p=0. 415 for MR; F(5,6)= 286.9, p= 4.72e-07 for ZBTB16; F(5,6)=0.633, p= 0.684 for TFF1; F(5,6)=10.01, p= 0.0071 for GR; and F(5,6)=6.408, p= 0.0213 for PER1. 226
Figure 65. Effect of Drsp, Aldo, E2, and combination treatment on protein levels of MCF-7 cells transfected with MR expressing vector (OV) for 24 hours. A) Western blot images of the protein levels of the groups after the application and GAPDH protein used as loading control and B) MR, GR, CHK1, C) BAX, and BCL-2 protein levels. (* : p<0.05) 228

Figure 66. Expression pattern of FKBP5 and its validation results in qRT-PCR. One-way ANOVA with Tukey's multiple comparisons test was used. ANOVA results in MCF-7 cells are as follows: : F(5,6)=10.58, p=0.00616. ANOVA results in T47D cells are as follows: F(5,6)=15.78, p=0.00213. (*: p<0.05, **: p<0.01, ***: p<0.001, ****:p<0.0001)..... 230
Figure 67. Expression pattern of NDRG1 and its validation results in qRT-PCR. One-way ANOVA with Tukey's multiple comparisons test was used. ANOVA results in MCF-7 cells are as follows: : F(5,6)=4.771, p=0.0417. ANOVA results in T47D cells are as follows: : F(5,6)=18.03, p=0.00148. (*: p<0.05, **: p<0.01, ***: p<0.001, ****:p<0.0001)..... 231
Figure 68. Expression pattern of BCL6 and its validation results in qRT-PCR. One-way ANOVA with Tukey's multiple comparisons test was used. ANOVA results in MCF-7 cells are as follows: : F(5,6)=5.619, p=0.0289. ANOVA results in T47D cells are as follows: : F(5,6)=12.75, p=0.00378. (*: p<0.05, **: p<0.01, ***: p<0.001, ****:p<0.0001)..... 232
Figure 69. Uncut and input chromatin samples run on the agarose gel..... 233
Figure 70. qPCR analysis by Percent input method for A) Control sample, B) DE2 sample, C) E2 sample. D) by Fold Enrichment method normalized to negative control 3. 234
Figure 71. Quality control results of all samples checked at Galaxy..... 235
Figure 72. Location and intensity of peaks over the whole genome and Venn diagram A) Control peaks; B) D_E2 peaks and C) E2 peaks. D) Venn diagram representation for common and individual peaks of samples after removal of blacklist regions..... 236
Figure 73. Peak profile plots and heatmap. A) Read count frequency of control, D_E2 and E2 peaks by TSS region, B) heatmap image of the information in A..... 237
Figure 74. Genomic annotation of the samples. A) Distribution of genomic features in control, DE2 and E2 samples. B) Distribution of binding portions of control, DE2 and E2 peaks according to their distance from the TSS..... 238
Figure 75. Comparison of KEGG pathways in DE2 and E2 samples. 239
Figure 76. Peak and coverage result of our own ChIP data and GSE104983 data in the PGR (A) and GREB1 (B) region in IGV..... 240
Figure 77. Analysis of GSE228941 data using two clusters. A) Expression patterns of genes in clusters. B) MSigDB analysis using Hallmark for non-essential genes. C) MSigDB analysis using Hallmark for essential genes. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between treatment and control samples. 243
Figure 78. Cluster and functional annotation analysis of empty vector (EV), wildtype (WT), R175H and R273H TP53 mutant expressing MCF7 cells using CAP-RNAseq. A) Expression patterns of genes in 8 clusters. B) MSigDB analysis using Hallmark for essential genes. C) MSigDB analysis using Hallmark for non-essential genes. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between WT and EV samples..... 246
Figure 79. Western blot results with proteins extracted from cells collected on days 2, 3 and 4 post-transfections. A) Western blot results of cells collected 2, 3 and 4 days after transfection under MR and GAPDH control. B) The result on the left shows the amount of MR protein in cells in complete media collected on days 2, 3 and 4, while the panel on the right shows the results of serum-deprived cells using MR and STAG-antibodies. MR and STAG results are consistent..... 302
Figure 80. MTT optimization results. Analysis of the effect of transfection (EV or OV) and E2 (20 nM) in MCF-7 cells seeded with different cell numbers. 305

Figure 81. MTT optimization results. Analysis of the effect of transfection (EV or OV) and E2 (10 nM and 20 nM) in MCF-7 cells seeded with 5K and 8K cell numbers. Dunnett's multiple comparisons test was performed. (*: p<0.05, **: p<0.01, ***: p<0.001, ****:p<0.0001). 306
Figure 82. Comparison of cell viability according to Drospirenone concentrations on MCF-7 cell line. At 24 h incubation, cells were treated with Drospirenone (Drsp) prepared at 4 different concentrations from 0.01 μM to 10 μM. A) and B) show the results of repeating the same experimental setup in two 96-well plates with three biological replicates of each sample. 308
Figure 83. Comparison of cell viability according to Drospirenone concentrations on MCF-7 cell line. After 24 hours of incubation, cells were treated with Drospirenone prepared at 4 different concentrations from 12.5 nM to 100 nM. A) EV transfected cells; B) OV transfected cells. 309
Figure 84. Effect of Drsp, Aldo, E2 and combination treatments on cell viability levels of MCF7 cells transfected with empty vector (EV) and MR expressing vector (OV) for 24 hours. Experimental results are expressed as mean ± standard deviation (n=5 for all groups). Two-way Analysis of Variance and Sidak's test for intergroup comparisons were applied for statistical analysis (*: p<0.05, **: p<0.01, ***: p<0.001, ****:p<0.0001). 310
Figure 85. qRT-PCR results normalized to EV Aldo (D1:400 nM, D2:80 nM, Aldo:100 nM and E2: 10 nM). mRNA expression of A) PGR B) MR C) ZBTB16 genes. 313
Figure 86. qRT-PCR results in MCF-7 cells using cell cycle genes normalized to EV Aldo (D1: 400 nM, D2: 80 nM, Aldo: 100 nM and E2: 10 nM). Expression of A) CCND1, B) CCNE2, C) RRM2, D) CHK1, E) BAX and F) BCL2 genes. 314
Figure 87. qRT-PCR results with RNAs from T47D cells (*: p<0.05, **: p<0.01, ***: p<0.001, ****:p<0.0001). One-way ANOVA results for genes are as follows: F(2,3)=2.192 p=0.259 for NEDD4L; F(2,3)=1.211 p=0.412 for HSD11B2; F(2,3)=2.847 p=0.203 for TFF1; F(2,3)=2.023 p=0.278 for CLDN1; F(2,3)=1.056, p=0.45 for CCND1 and F(2,3)=1.539 p=0.347 for GATA3. 316
Figure 88. Concentration information before library preparation (A) and reads of raw data (B). 323
Figure 89. Hallmark enrichment of essential genes for cluster 2, cluster 4 and cluster 11 when comparing CDK4 inhibitor versus Serum Starvation groups. 324

List of Tables

Table 1. Comparison of multiple features of the available shiny applications and CAP-RNAseq..... 40
Table 2. Hormonal effects of certain progestogens. 69
Table 3. List of packages used in the development of CAP-RNAseq. 112
Table 4. List of materials, kits and reagents utilized in cell culture experiments. 113
Table 5. Hormones used in the experiments. 114
Table 6. List of materials, kits and reagents for RNA experiments. 114
Table 7. Laboratory materials, kits and reagents for ChipSeq experiments 115
Table 8. List of primers used in the experiments and their sequences..... 116
Table 9. List of equipment used in the experiments. 117
Table 10. Reaction conditions for qRT-PCR experiments..... 123
Table 11. Solution and Buffers used in Western Blot..... 126
Table 12. Antibody list used in Western Blot. 127
Table 13. The number of differentially expressed genes in each comparison using an absolute log-fold change threshold of 0.5..... 132
Table 14. Essentiality information of clusters for Demo 1 dataset. SD: standard deviation; OR: Odds Ratio; FET: Fisher's Exact Test; p.val: p-value. 137
Table 15. CV, log2 average CPM and log2 range CPM values of selected genes. 140
Table 16. Differential gene expression analysis results comparing TrkB.T1 and Control Samples using DESeq2 method for the identified genes. 142
Table 17. The number of differentially expressed genes in each comparison using an absolute log-fold change threshold of 0.5..... 148
Table 18. Essentiality information of clusters for Demo 2 dataset. SD: standard deviation; OR: Odds Ratio; FET: Fisher's Exact Test; p.val: p-value. 150
Table 19. First 20 results of the enrichment analysis of upregulated genes in cluster 2 for the REACTOME and GO terms. 154
Table 20. RIN values of samples from T47D cells. 167
Table 21. Essentiality information of clusters for T47D dataset. SD: standard deviation; OR: Odds Ratio; FET: Fisher's Exact Test; p.val: p-value. 171
Table 22. The result of DESeq2 analysis of genes in cluster 9. 175
Table 23. First 10 results of the enrichment analysis of the genes in cluster 4 for the KEGG, REACTOME, and GO terms. 178
Table 24. DESeq2 result of two genes selected from cluster 4. 179
Table 25. REACTOME and GO enrichment results of essential genes in cluster 3.... 189
Table 26. GO enrichment result of cluster 8. 191
Table 27. GO and REACTOME enrichment result of cluster 1. 193
Table 28. DESeq2 results of selected genes from cluster 10 and 7 in comparison of D_E2 versus Control samples. 196
Table 29. RIN values of samples from MCF-7 cells. 199
Table 30. Essentiality information of clusters for MCF-7 dataset. SD: standard deviation; OR: Odds Ratio; FET: Fisher's Exact Test; p.val: p-value. 202
Table 31. LogFC and p-values of Drsp and Drsp_E2 samples of MCF-7 and T47D RNA-seq data using compared to control. 223

CHAPTER 1

INTRODUCTION

1.1 Deciphering Transcriptomes with RNA-Sequencing

Transcriptomics technologies are known as the methods for studying the tissue specific transcriptome of the organism, i.e., whole RNA transcripts (Adams et al., 1991; Velculescu et al., 1997). An organism's transcriptome is encoded in its genome and transcriptionally expressed by multiple cell- or tissue-specific regulatory mechanisms. In this case, mRNA acts as a temporary bridge molecule in the information network leading to proteome while noncoding RNAs carry out a variety of other roles. Hence, in a single cell, a transcriptome represents all the transcripts present at a given moment in time (Lowe et al., 2017).

In the late 1990s, researchers began investigating the transcriptome as a whole for the first time, and since then transcriptomics is widely accepted due to the technological advancements from the time the first attempts were made (Adams et al., 1991; Seixas et al., 2023). The area of transcriptomics has been revolutionized by a number of technical advancements. Microarrays, which measure a set of pre-determined sequences, and RNA-seq, which uses high-throughput sequencing to read pieces from transcripts, are the two main modern methods in the area (Bainbridge et al., 2006; Schena et al., 1995).

1.1.1 RNA-seq Basics

A new age in molecular biology has been brought about by RNA sequencing, or RNA-seq (Z. Wang et al., 2009). With unprecedented precision and depth, researchers can now study the intricacies of transcriptomics and gene expression through this powerful technology. Depending on the sequencing technology used, the length of the nucleotide sequences produced in RNA-seq can vary from 30 bp to more than 10,000 bp. Typically, they are about 100 bp long. These reads are obtained from the sequencing facilities as text files in the "FASTQ" or "fastq" format, which is a common file type developed for transferring sequencing read data that includes the sequence and a corresponding per-base quality score (Pearson & Lipman, 1988). This format is based on an older format known as FASTA, which is a text-based format designed for the representation of protein or nucleotide sequences (Pearson & Lipman, 1988). FASTQ format has been designed to deal with the output of base quality measurements from sequencers. The quality scores and sequences are provided as individual ASCII characters in the FASTQ format (Cock et al., 2010). Each sequence in the format consists of four lines, which are layered on top of one another in text files produced by processes that involve sequencing. The four lines will each stand for a read. Through the alignment of reads to a reference genome or to each other (de novo assembly), RNA-seq uses deep sampling of the transcriptome with several short transcripts to enable computational reconstitution of the original RNA transcript (Z. Wang et al., 2009).

The process for RNA-seq data analysis can be divided into four phases: quality control, alignment of reads, quantification, and the differential expression (Van Verk et al., 2013). The great number of widely available RNAseq algorithms are performed via a command-line interface, which may be used in either a Unix or R programming including

Bioconductor (Huber et al., 2015). The first phase, quality control is applied on the raw fastq files using tool such as FastQC (Babraham Institute, n.d.) and FaQCs (Lo & Chain, 2014) to assess the quality of the data. This assessment includes the parameters such as high unexpected read duplication rates, overrepresentation of specific short sequence motifs (k-mers), high base call quality scores, and guanine-cytosine content that fits the predicted distribution (Conesa et al., 2016). Any abnormalities found during the library preparation or the sequencing process, found can be trimmed away or tagged for extra attention at the later steps (Van Verk et al., 2013).

Following the quality control, sequencing reads are aligned to reference genome/transcriptome or to one another if no reference is available (de novo alignment). Tools like BWA (H. Li & Durbin, 2009a) and Bowtie (Langmead et al., 2009) are frequently employed for transcriptome alignments, while TopHat (Trapnell et al., 2009) is a favored splice aligner for matching the reads to an annotated reference genome. Additionally, STAR (Spliced Transcripts Alignment to a Reference) can be used to detect splice-junctions using a customized method for RNA-seq reads that use a seed extension query that relies on compressed suffix arrays (Dobin et al., 2013). HISAT2 (Hierarchical Indexing for Spliced Alignment of Transcripts 2), another advanced splice-aware aligner that can align DNA and RNA sequences, employs a graph-based alignment method called the graph Ferragina Manzini index (D. Kim et al., 2015, 2019).

After alignment, the number of reads corresponding to each gene is counted to convert the data into a quantitative measure of gene expression. The result of this operation will be a table with the raw number of reads for each feature. HTSeq package is one tool that can be used for the exon and gene counts to be calculated (Anders et al., 2015). Transcript

level quantitation is more complex and involves probabilistic techniques, such as the use of Cufflinks software, to estimate the abundance of transcript isoforms from short read data (Trapnell et al., 2010).

When quantitative counts of each transcript or genes are obtained, it is time to identify genes that are differentially expressed under various circumstances. However, normalization of the count data is necessary before comparing expression levels between samples with accuracy. When normalizing count data, two significant forms of sequencing bias must be taken into account: between-sample bias, mostly due to variations in sequencing depth caused by technical effects (M. D. Robinson & Oshlack, 2010), and within-sample bias, primarily due to changes in GC content and transcript length (Oshlack & Wakefield, 2009; Risso et al., 2011).

RPKM/FPKM is one of the mostly used normalization techniques; it does both within- and between-sample normalization by adjusting raw counts to total gene length and the number of reads assigned within a given sample (Mortazavi et al., 2008; Trapnell et al., 2010). RPKM/FPKM measurements are helpful in examining variations in amount of alternative splicing variants among samples. Many other normalization techniques have been developed which rely on the variables such as distribution of read counts or library size. CPM (counts per million) only normalizes for sequencing depth (depth-normalized counts) by library size (Koch et al., 2018). CPM values are obtained by dividing the number of mapped reads by a per million scaling factor of the total number of mapped reads. Another method, "voom" (variance modeling at the observation level), is used to estimate the trend of mean-variance of log-transformed read counts and to anticipate the variance of the log2CPM values using the connection between mean and variance (Law

et al., 2014). Moreover, The DESeq normalization algorithm can be used to calculate the ratio of each read value of a gene to a geometric mean of all count values for that gene over all samples (Anders & Huber, 2010). There are also other normalization methods including upper quantile (Bullard et al., 2010), Median normalization (Dillies et al., 2013), trimmed mean of M-values (TMM) (M. D. Robinson & Oshlack, 2010), CuffDiff (Trapnell et al., 2013) and PoissonSeq (J. Li et al., 2012) which are explained in more detail in a review by Evans et al. (Evans et al., 2018).

Numerous statistical techniques have been developed to identify genes that show differential expression under various situations. Among them, edgeR (M. D. Robinson et al., 2010), limma (Ritchie et al., 2015) and DESEq2 (Love et al., 2014) stand out as commonly used methods. edgeR and DESEq2 offer their own internal normalization techniques and use the negative binomial distribution as a reference, while limma works by estimating the mean-variance relationship in log-scale read counts. Other than these, there are also methods that utilize Bayesian approaches such as baySeq (Hardcastle & Kelly, 2010) and EBSeq (Leng et al., 2013) and non-parametric approaches such as SAMseq (J. Li & Tibshirani, 2013) and NOISeq (Tarazona et al., 2011). Details about these techniques have been explained in the thesis of Ronaldo Leka (Ronaldo Leka, M.Sc. Thesis, 2022, Bilkent University). Figure 1 displays the flow chart for the analysis of RNA sequencing data.

Other than the bulk RNA-seq data which gives a complete view of gene expression over a whole cell population, there are new developments in the RNA-seq such as single-cell RNA-seq (scRNA-seq), which facilitates the identification of cell types, states, and heterogeneity within a population by capturing gene expression at the individual cell level

(Hashimshony et al., 2012; X. Li & Wang, 2021), and Single-nucleus RNA-seq (snRNA-seq), which evaluate gene expression from single nuclei as opposed to entire cells (Grindberg et al., 2013). While these phases of RNA-seq analysis mentioned above are applied to bulk RNA-seq data, similar outputs are obtained from the analysis of scRNA-seq and snRNA-seq, but they differ in the analysis methods (N. Kim et al., 2023; Quatredeniers et al., 2023). Moreover, with the amount of RNA-seq data increasing, there has never been a more pressing need for trustworthy analysis and interpretation tools.

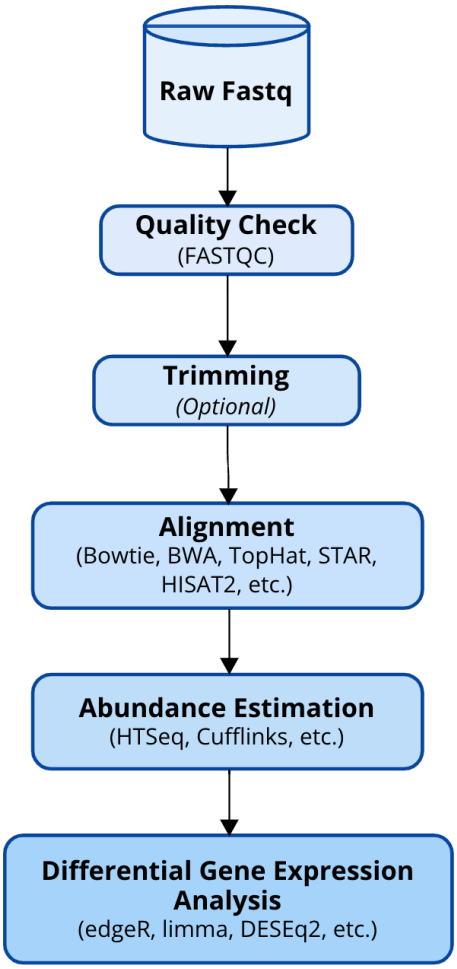


Figure 1. Flow chart of RNA-seq analysis.

1.1.2 Navigating the RNA-seq Landscape: Challenges and Opportunities

RNA-seq leverages recently discovered deep sequencing techniques. In general, a population of RNA (total or fractioned, such as poly(A)+) is used to generate a library of cDNA fragments with adapters attached at both ends. Each fragment is then sequenced in a high-throughput manner, either with or without amplification, to generate short sequences from one end (single-end sequencing) or both ends (pair-end sequencing). The reads vary in length from 30 to 400 bp, depending on the DNA sequencing method used (Z. Wang et al., 2009).

RNA-seq is not restricted to finding transcripts which match to identified genomic sequences, contrary to hybridization-based methods. It is broad and adaptable in that it may be used with non-model species whose genomic sequences are unknown (Vera et al., 2008). With single-base precision, RNA-seq is able to identify the precise positions of where transcription starts or ends. This degree of accuracy is useful for describing transcripts, particularly those with alternative splicing events. Moreover, short reads (e.g., 30-bp) from the studies can reveal the connections between individual exons within a transcript, while complex transcript architectures can be studied by longer reads or paired-end short reads that indicate connection across several exons (Z. Wang et al., 2009). Sequence changes, such as single nucleotide polymorphisms (SNPs), inside transcribed places can also be found by RNA-seq. This characteristic of RNA-seq is crucial to comprehend genetic diversity and differences in gene expression (Cloonan et al., 2008). The transcripts can be detected over a wide dynamic range of expression levels. In a study of *Saccharomyces cerevisiae* analyzing 16 million mapped reads, the predicted range was over 9,000-fold (Nagalakshmi et al., 2008). Another strength of RNA-seq is that the accuracy of RNA-seq in measuring gene expression levels can be confirmed using

quantitative PCR (qPCR) as well as spike-in RNA controls at predetermined quantities (Mortazavi et al., 2008; Nagalakshmi et al., 2008). Moreover, RNA-seq data shows a high degree of reproducibility for both biological and technical replicates (Cloonan et al., 2008; Nagalakshmi et al., 2008). While RNA-seq has several advantages, there are still some challenges that must be overcome. One of them is efficient data management, which involves processing, retrieving, and storing massive databases generated by even with the RNA-seq.

RNA-seq pipeline can be described shortly as the following: After obtaining high quality reads, quality control is required in data processing that involves identifying and removing low-quality sequence reads. Then, the short reads of RNA-seq need to be mapped to reference genome. However, reads which are exon-spanning junctions or include poly(A) ends cause some difficulties. The existence of trans- and alternative splicing in complex transcriptomes complicates the mapping of reads crossing exon junctions. This work can be made easier by creating a junction library with known and expected junction sequences (Mortazavi et al., 2008; Wilhelm et al., 2008). Creating computational techniques to detect new splicing events is still a problem, particularly when they happen across distant regions or between exons from distinct genes.

Alignment becomes complicated since a considerable fraction of sequence reads correspond to multiple locations in the genomic region. Assigning these multi-matched reads proportionately according to the number of reads corresponding to nearby distinct sequences can be one solution (Cloonan et al., 2008; Mortazavi et al., 2008). This technique works well for repeating sequences with few copies (Mortazavi et al., 2008). However, using paired-end sequencing methodologies or getting longer sequence reads

might help alleviate the multi-matching issue for high-copy, repetitive sections, or sequences with extended periods of repetition (X. Liao et al., 2021). Moreover, polymorphisms and sequencing mistakes can be problematic for genome mapping (Imelfort et al., 2009). Although most mapping techniques can handle single-base changes, better annotation of the reference genome for polymorphisms and more sequence coverage is needed to address bigger discrepancies. In conclusion, tackling these informatics issues is critical to precise and thorough RNA-seq data interpretation, especially for complex transcriptomes and genomes containing repeated sequences and polymorphisms. Improving RNA-seq analysis requires focusing on two main areas: creating novel computational methods and utilizing newly available sequencing technology.

1.1.3 The Significance of Clustering in RNA-seq Analysis

RNA-seq analysis is a potent approach that enables us to thoroughly examine the genetic activity of the organism and comprehend gene expression. These techniques frequently generate gene expression data as large, complex datasets. At this point, the importance of clustering analysis becomes apparent since many genes are transcribed at the same time or in the opposite direction during a particular process, e.g., cell cycle. After obtaining count data from RNA-seq, a variety of downstream analyses are used to investigate gene expression. Since RNA-seq data comprises a gene expression matrix with thousands of genes, it can be complicated and challenging to analyze. Cluster analysis provides a way of visualizing the relationships between data by partitioning a data matrix into groups with similar characteristics (Alkarkhi & Alqaraghuli, 2019; Fletcher et al., 1993). The importance of clustering in RNA-Seq analysis is illustrated by the following examples.

Cluster analysis, an unsupervised method for identifying groups with comparable patterns, is a widely used technique for identifying novel subtypes of diseases such as cancer (Sotiriou et al., 2003; Y.-H. Zhang et al., 2017). Tissues or cells with similar gene expression patterns might belong to the same subtype (Bertucci et al., 2005; Ren et al., 2016). These subtypes may be found via clustering expression profiles, which also aids in a deeper understanding of the physiological process or disease of interest.

Gene expression patterns that are similar between groups of genes can be found using cluster analysis. This facilitates our understanding of the genes that are active under specific conditions (e.g., treatment or disease). Clustering in RNA-seq data is an exploratory approach for organizing and visualizing the underlying connections among groups of genes and helps the selection of specific genes for additional analysis. Moreover, clustering has been widely used to classify and categorize Differentially Expressed Genes (DEGs), which can be defined as genes that exhibit statistically significant differences in their expression between the stated groups, e.g., control versus experimental condition (Fu et al., 2019; P. Li et al., 2010; Si et al., 2014). This strategy makes it easier to group genes with similar expression profiles in a systematic manner, which helps with categorization and offers a formal framework for understanding the dynamics and patterns of DEGs. In instances where complex expression patterns may emerge, such as time-course and multi-group RNA-seq data, gene clustering has proven to be an efficient tool (Osabe et al., 2021). It is interesting, nevertheless, that this approach is not as commonly used when comparing two groups, which is the bare minimum for comparative analysis.

Visualizing multi-dimensional RNA-seq data can be challenging (X. Yu et al., 2017). By clustering samples with similar gene expression patterns, the data becomes easier to visualize through heatmaps or line plots. For example, clustering was also used to detect differences between patients responding to treatment and those failing to do so improving the development of personalized medical interventions (Ortega et al., 2014).

To sum up, clustering in RNA-seq analysis is a crucial technique for improving the interpretability and visualization of significance of gene expression data. Numerous domains, including genetic studies (Vidman et al., 2019), drug development (Ye et al., 2018), disease treatment (Sotiriou et al., 2003) and diagnosis (Shigemizu et al., 2020) in biomedical research, have greatly benefited from this approach.

1.1.4 Methods of Clustering Gene Expression Data

In terms of grouping biological samples, cluster analysis approaches have shown to be helpful in understanding gene expression data by revealing previously undiscovered correlations among genes and distinct disease subtypes. Gene expression studies can employ a variety of clustering techniques. Hierarchical clustering is one of the most used methods. By creating groupings of the most closely associated and with comparable patterns of expression with each other, hierarchical clustering is a straightforward yet effective technique for analyzing gene expression data (Altman & Krzywinski, 2017; H. Y. Chang et al., 2013). Genes that have strong correlations or the least distances based on their expression matrices are grouped together repeatedly to achieve this. Consequently, a dendrogram is produced and its branches show how close the genes are to one another; the shorter the branch, the more similar (or less distant) the gene expression pattern is.

There are two types of hierarchical clustering technique: agglomerative and divisive. The agglomerative technique, also known as the bottom-up approach, begins with individual data points and treats each one as a separate cluster before beginning to connect them into clusters (Murtagh & Legendre, 2014). The top-down, or divisive, technique divides the data points into several sub-clusters after first treating each as a member of a single cluster (Roux, 2015) while the most popular method is the former. Another reason for the popularity of hierarchical clustering is its ability to illustrate the general commonalities between expression patterns. Biologists can rapidly evaluate the number and magnitude of expression patterns within a dataset.

Unlike hierarchical clustering, k-means clustering which is another popular clustering algorithm, requires choosing the number of clusters first (S. S. Khan & Ahmad, 2004). Each data point is assigned to the nearest centroid iteratively until no better centroid is available. This is done by taking the average of the data points mapped to each cluster into account at each iteration. The algorithm carries out this procedure until convergence, at which point a final clustering solution is produced. K-means, a partitional clustering technique, divides data observations into nonoverlapping groups (MacQueen, 1967). Each group (or cluster) must include at least one observation, and a data point cannot belong to more than one group. The number of clusters, also known as variable k, must be specified beforehand to use partitional clustering methods (Huh, 2002). However, k-means has many drawbacks, including its sensitivity to noise and outliers, bias towards clusters with a spherical form, and its requirement for previous knowledge of the number of clusters (Gao et al., 2023). Partitional clustering methods include k-means and k-medoids, which are non-deterministic, meaning that two successive runs with the same

input might provide different results, which is another distinction between partitional clustering and hierarchical clustering (Kutbay, 2018).

Another clustering methodology is model-based clustering that depends on a model to sort out the number of clusters. In the study of Si et al., several model-based clustering techniques were described in detail for RNA-seq data (Si et al., 2014). These techniques rely on the data being produced by mixing probability distributions: the Poisson distribution in the case of using only technical replicates and the Negative Binomial distribution in the case of using biological replicates. First, they suggested a model-based clustering technique for RNA-seq gene expression profile clustering that uses the expectation-maximization algorithm (MB-EM). The second approach is an initialization technique for cluster centers; the idea is to choose genes according to the distance between each of the genes and each of the selected centers, and then progressively add cluster centers by selecting one cluster center at random. They also introduce stochastic version of the expectation-maximization algorithm which adds randomness to the EM algorithm in order to avoid convergence to local optima.

While k-means and hierarchical clustering are popular classical algorithms, there are also other similar clustering techniques that can be used instead, such as e.g. self-organizing map (SOM) (Kohonen, 1982), Partitioning Around Medoids (PAM) (Kaufman & Rousseeuw, 1990) and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) (Ester et al., 1996).

Self-organizing maps (SOMs) (Kohonen, 1990) are a model-based clustering technique based on neural networks that are often used in gene clustering. High degrees of

interconnectivity are frequently seen in gene expression data, and one gene may show robust relationships with two or more other clusters. SOMs are very useful for capturing these complex linkages that exist throughout the data.

Another partitional clustering approach called PAM (Partitioning Around Medoids) cluster items according to their differences. It works well with objects that are measured using interval-scaled variables or using dissimilarity matrices as the input data. PAM identifies the centrally located item, known as the medoid, and creates the initial clustering around it. In comparison to K-means, PAM is more robust since it is minimizing the sum of dissimilarities, rather than the sum of squares of the Euclidean distances (Kaufman & Rousseeuw, 1990)

Density-based clustering method DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is resistant against noise and can find clusters of any shape (Ester et al., 1996). The way it operates is by first establishing a neighborhood surrounding every data point, which is subsequently expanded to include any point that can be reached by density. The algorithm does not have a pre-specification of the number of clusters and can handle datasets with varying densities. DBSCAN is particularly useful for clustering gene expression data and has been shown to outperform other clustering techniques such as hierarchical and K-means clustering.

In the literature, there are many studies that compare these clustering techniques. For example, after comparing four different clustering techniques and 12 distance measurements, Jaskowiak et al. came to the conclusion that hierarchical clustering with average linkage and k-medoids were often better than hierarchical clustering with

complete or single linkage (Jaskowiak et al., 2018). In another study where seven clustering techniques were examined on 35 gene expression datasets by de Souto and colleagues, a parametric approach assuming a mixture of multivariate Gaussians produced the best results (de Souto et al., 2008). While Freyhult et al. showed that k-means and hierarchical clustering with Ward's linkage outperformed PAM and SOM, they also identified pre-processing processes as potentially having a significant impact on the outcome (Freyhult et al., 2010).

1.1.5 Gene Selection Methods and Their Impact on Clustering Performance

Cancer and other diseases have a variety of origins, and it is plausible that the disorders that cause cancer today might be further subdivided into a number of subtypes, each of which may require a distinct course of treatment. Novel cancer subtypes have been successfully identified through the application of cluster analysis to gene expression data from cancer patients (Bertucci et al., 2005; Eisen et al., 1998; Fujikado et al., 2006; Lapointe et al., 2004; Ren et al., 2016; Sotiriou et al., 2003). However, finding novel subtypes is difficult because the majority of genes express themselves independently of the disease subtype. Ideally, only informative genes should be included in a cluster analysis intended to identify novel disease subtypes. Hence, it makes sense to apply a gene/feature selection process before performing a cluster analysis. If the subtypes (i.e., labels) that the patients had were known, this task would be reasonably simple; however, in unsupervised classification, the labels are not known, making it difficult to select genes. The removal "non-informative" genes prior to clustering is thought to improve performance in the case of cluster analysis. This is applied in the studies by excluding the lowly expressed genes, which are assumed as less informative (Källberg et al., 2021).

Next, the data could be subjected to a variance-stabilizing transformation (VST) using the DESeq2 package (Love et al., 2014). Based on a model that takes into account the variance-mean connection suggested by a negative binomial distribution for the counts of gene expression, the VST technique is implemented in DESeq2. The purpose of this modification is to guarantee that the data have roughly homoscedastic qualities, which are advantageous for clustering analysis.

Statistical techniques like t-tests, one-way ANOVA or Wilcoxon rank sum tests cannot be used to detect differentially expressed genes when the labels are unknown. Instead, other features of the data must be taken into account. However, when the labels are known, and for two group comparisons, the feature selection can be accomplished by running a two-sample t-test for each gene to find a set of informative genes, then choosing the genes with the smallest p values (Önskog et al., 2011). The feature selection techniques can be evaluated under for categories: selection of genes that a) have bi- or multimodal profiles, b) are highly variable, c) highly expressed or d) highly correlated. Genes can be sorted according to a bimodality measure in the context of two potential tumor subtypes; only the highest scoring genes with the highest bimodality measures are kept for additional examination. Numerous bimodality-based gene selection strategies have been put forth (Moody et al., 2019), such as the bimodality coefficient (BC) (SAS Institute Inc, 1990), the bimodality index (BI) (J. Wang et al., 2009) and the variance reduction score (VRS) (Bezdek, 1981; Hellwig et al., 2010) in its different forms. A more comprehensive strategy is to find genes that appear to have multimodal distributions; dip-test (DIP) is one technique that helps with this (Hartigan & Hartigan, 1985).

The variation in gene expression levels among samples is the subject of another class of gene selection techniques. The idea is that genes with high variability may show notable differences associated with different disease subtypes (Wolf et al., 2023). To find subsets in high-dimensional genomic data, for instance, clustering is sometimes performed on a subset of characteristics selected on the basis of their standard deviation (SD) across the samples, which serves as a score indicating a gene's level of informativeness for the underlying subgroups (Bentink et al., 2012; S. Kim et al., 2020; Shen et al., 2020). For gene selection, however, several scoring techniques beyond SD are taken into account. Among these options are entropy-based measurements which evaluate genes by computing the difference between the first and third quartiles of their expression levels and variability-based ratings such as the interquartile range (IQR) which measures gene expression variability using a statistical concept called Rényi entropy (X. Liu et al., 2005; Seal et al., 2016).

Choosing highly expressed genes is another strategy, which is predicated on the idea that informative genes are likely to become relatively highly expressed. When noise levels relative to mean expression values are high, it might be challenging to distinguish between disease subtypes, but among highly expressed genes, it is easier to identify differentially expressed genes. The third quartile (Q3) and mean value (M) of the approaches are included in this category (Källberg et al., 2021).

Another category includes the genes that are highly correlated. Genes that have strong correlations with one another might belong to the same functional networks and so be promising biomarkers. In clustering issues involving high-dimensional microarray data,

co-expression between informative genes was utilized for variable selection (Wang et al., 2014).

In the study, Källberg et al. investigate the above gene selection techniques to understand their influence on clustering performance in different data (Källberg et al., 2021). Out of the three techniques that depend on sample variability, ENT and IQR outperformed the widely used SD approach in most cases. Among all the approaches taken into consideration, the gene correlation techniques yielded the lowest results. DIP and BI were the modality-based approaches with the best overall performance.

1.1.6 Annotation of Clusters

Following the clustering of genes, the annotation of these clusters is a crucial step in understanding their biological significance. Through enrichment analyses, the clusters are examined to identify overrepresented biological terms or pathways, providing insights into the potential functions and processes linked to the genes within each cluster. Integration of molecular terms further enhances the annotation process by linking the clustered genes to specific molecular functions, biological processes and cellular components. There are multiple analytical methods for the annotation of genes in clusters such as GSEA (Gene Set Enrichment Analysis) (Subramanian et al., 2005), ORA (Over-representation analysis) (Goeman & Bühlmann, 2007) and QuSAGE (Quantitative Set Analysis of Gene Expression) (Yaari et al., 2013). Usually, the annotation terms acquired from the databases or gene set collections such as REACTOME (Croft et al., 2011), KEGG (Ogata et al., 1999), Gene Ontology (Ashburner et al., 2000), Disease Ontology (Baron et al., 2024) and CORUM (Tsitsiridis et al., 2023). Additionally, the exploration

of protein-protein interactions within the clusters elucidates the potential relationships and functional associations among the gene products, shedding light on their collaborative roles within cellular pathways (Rao et al., 2014). By employing these annotation strategies, the biological relevance of the gene clusters is unveiled, facilitating a deeper comprehension of their implications in various cellular processes and disease mechanisms.

Furthermore, the annotation of clusters displaying mirror-like expression patterns turns out to be beneficial and have been used in the studies (Baldessari et al., 2005; Lewis et al., 2013; Yaslianifard et al., 2023). This approach provides an improved understanding of the opposing expression patterns within the dataset, making it easier to identify gene pairs or groups whose expression profiles show inverse associations. Improving understanding of these opposite expression patterns can help identify important molecular mediators and pathways involved in disease development. This kind of annotations have the potential to help us better understand biological processes and pathways in which gene sets are expressed inversely to each other.

1.1.7 DepMap and its Potential as an Annotation Tool

Cancer Dependency Map (DepMap), joint initiative of the Broad Institute and the Wellcome Sanger Institute, is a comprehensive resource that plays a critical role in unraveling the genetic basis of cancer and aims to map genetic dependencies in a large number of cancer cell lines (Behan et al., 2019; Tsherniak et al., 2017). DepMap integrates data from in-depth genome profiles of cancer cell lines and functionally characterizes cancer cell dependencies, creating a comprehensive map of cancer's

intracellular genetic dependencies and vulnerabilities. This dynamic platform offers insights into the susceptibilities and dependencies of cancer cells on specific genes, thus facilitating the identification of potential therapeutic targets. Through systematic exploration of the genetic makeup of various cancer types, DepMap provides significant information on the indispensability of genes, patterns of expression, and molecular interactions.

Using loss of function genetic screens based on RNA interference (RNAi), it was sought to uncover genes essential for the survival and proliferation of 501 cancer cell lines in an extensive investigation (Tsherniak et al., 2017). This effort opens the door for the discovery of possible cancer targets by creating a strong cancer dependency map and introducing a novel method for identifying characteristics indicative of gene dependencies. However, there was growing doubt over the effectiveness of RNAi-based screening due to the recognition of shRNAs' off-target effects. Taking into account the issues raised by the off-target effects of small hairpin RNAs (shRNAs), a different strategy makes use of additional loss-of-function techniques. Specifically, the development of CRISPR-Cas9 technology offers a viable alternative to RNA interference (RNAi) for determining which genes are essential. To understand the cancer dependencies, hundreds of cancer cell lines across 30 cancer types were screened for CRISPR-Cas9 at the genomic level (Behan et al., 2019). By using CRISPR, they silenced each gene in turn and assessed the effects of these gene knockouts on cancer cell survival from 30 different cancers. Thus, by combining RNAi and CRISPR-Cas9 technologies, two powerful genetic screening methods were shown to act in concert to create a complete framework for cancer dependency mapping.

The DepMap portal, available at https://depmap.org/portal, provides detailed gene information obtained from these CRISPR and shRNA screens. This contains specifics like (1) the gene-dependent cell lines and lineages, (2) co-dependent genes, or those that have a significant positive or negative correlation with the gene in terms of growth effects, and (3) transcript abundances, copy numbers, and gene-related mutations. The information provided by DepMap serves as a foundation for various analyses. For example, an online application, ShinyDepMap, has been created to help researchers quickly determine whether a given gene is essential or selective in a variety of cell lines (Shimada et al., 2021). Through the integration of data from both CRISPR and shRNA screens, the tool generates reliable measurements that demonstrate how gene deletion affects cell viability. Among the developed metrics are **selectivity**, which shows how a gene's essentiality varies across various cell lines, and **efficacy**, which shows how much the absence of a gene inhibits cell growth in sensitive lines.

Another example of using DepMap's data in another tool is Chronos, a program designed to improve the understanding of CRISPR loss-of-function screens (Dempster et al., 2021). Although these screens are effective in investigating biological processes, they often contain biases and artifacts that may distort the outcomes. To overcome this challenge, Chronos employs a transparent model of cell growth kinetics following CRISPR gene deletion. Another example is NetControl4BioMed, a web based tool designed for generating or uploading directed protein–protein interaction networks and conducting structural network controllability analyses (Popescu et al., 2021). It integrated several protein data from prominent sources such as HGNC, Ensemble and UniProt, and notably, it enriches its functionality by incorporating DepMap data, specifically the 1526 sets of mutated genes associated with various cancer cell lines.

DepMap has also been included in an R Shiny application called DepLink (Nayak et al., 2023). Due to limitations in systematic analyses across all genes or drugs, DepLink aims to fill these gaps by facilitating the identification of potential surrogate drugs for gene knockouts and the determination of genes targeted by specific drugs. The application integrates large-scale genetic and drug screen data, molecular signatures, and cheminformatics through four interconnected modules to find out the drugs which are potential surrogate for knocking out a gene and the genes which are potential targets of a drug.

DEPMAP has shown promise in improving our understanding of tumor dependencies and in converting gene essentiality into therapeutic targets. However, dependency maps may not accurately reproduce the genetic factors seen in patient populations and have limitations when it comes to simulating the complexity of the complete tumor microenvironment (TME) (Flister & Bergom, 2018). Furthermore, not all elements of disease outcomes and patient survival may be captured by results from genetic perturbation screens (Tate et al., 2019). Although a great deal of genetic and molecular information on the cancer genome has been made available by the Cancer Genome Atlas (TCGA), many potential oncogenic drivers are still unknown in terms of their functional implications and clinical significance. In order to address these issues, $TCGA_{DEPMAP}$, the first hybrid map of translational tumor dependencies, is created by converting gene essentiality data from DEPMAP into TCGA patient data by machine learning (Shi et al., 2022). $TCGA_{DEPMAP}$ provides patient-relevant translatability while utilizing DEPMAP's experimental strengths. The map shows its robustness as a translational dependency map

by capturing known cancer lineage dependencies, oncogenes, and synthetic lethalities with significant implications for treatment response and clinical outcomes.

1.1.8 Genomic Exploration of Essential Genes

One significant field of research in genomics is identifying and characterizing essential genes. These genes are vital for an organism to survive and function correctly, playing pivotal roles in fundamental biological processes (Itaya, 1995; Kobayashi et al., 2003). High-throughput screens in cellular and animal models have identified several of these genes. Investigating them at the cellular level can be achieved through techniques such as RNA interference, gene trapping, or gene editing such as CRISPR-Cas9 (Lieben, 2016). The function of essential genes is considered to be fundamental to life and is therefore likely to be common to all cells. They have the crucial roles in various biological processes, offering insights into the molecular basis of diseases and disorders. Understanding their properties and functions is important in both biological and medical domains, as dysregulation of essential genes can have profound consequences. This information on essentiality is extensively applied in life science research, serving purposes such as identifying drug targets in areas like cancer therapy (Jeon et al., 2014) and insecticide development (Muthu Lakshmi Bavithra et al., 2023), as well as contributing to the design of minimal genomes in synthetic biology (B. H. Sung et al., 2016).

Essential genes could be identified by inactivating the genes in the cell. One method for identifying essential genes is the CRISPR/Cas9-based gene editing system (Cong et al., 2013; H. Wang et al., 2013). The discovery of CRISPR/Cas9 (CRISPR-associated RNA-

guided endonuclease Cas9), a straightforward but effective tool for editing genomes, has been one of the greatest advances in biotechnology (Cong et al., 2013; Jinek et al., 2012). The genomes of whole organisms and cultured cells can be engineered by this technique. Researchers can determine which genes are necessary for the best possible proliferation and survival in a particular cell line by building a genome-wide single-guide RNA (sgRNA) library and running a proliferation-based screen (Koike-Yusa et al., 2014; T. Wang et al., 2014). A guide RNA (gRNA) can direct the endonuclease Cas9 to the specified sites inside complicated genomes. The investigation of gene functions at the systems level has been made possible by the straightforward and scalable nature of Cas9-mediated gene editing. Moreover, these screens have revealed the involvement of the essential genes in fundamental cellular maintenance processes such as DNA, RNA and protein synthesis (D.-U. Kim et al., 2010). The proteins encoded by these genes show significant connectivity in protein-protein interaction (PPI) and metabolic networks (Aromolaran et al., 2020; Plaimas et al., 2010). Also, identifying core essential genes across different model systems highlights their evolutionary conservation (Luo et al., 2015).

Genome-wide identification of essential genes in various human cell types has been investigated in many studies. Blomen et al. generated a synthetic lethality map in human cells by conducting extensive mutagenesis on haploid cells (Blomen et al., 2015). They have identified roughly 2000 genes that are critical for optimal fitness under cultured conditions. Hart and colleagues screened five cell lines using CRISPR-based methods to identify fitness genes, resulting in the discovery of 1580 human core fitness genes and contextual fitness genes, such as those that confer pathway-specific genetic vulnerability in cancer cells (Hart et al., 2015).

Wang and colleagues applied CRISPR-based methods to evaluate various cell lines, revealing distinctive gene dependencies in tumors (T. Wang et al., 2015). They computed the CRISPR score (CS) for each gene, which represents the fitness cost associated with gene inactivation. High CS score genes are thought to be essential for the survival and growth of cells. Using this method, a group of cell-essential genes that are enriched for elements of basic pathways and highly expressed genes have been identified.

Essential genes are expected to be subject to stronger purifying selection than non-essential genes and exhibit greater conservation across closely related species since they show greater sequence conservation and are less prone to stop-gain deleterious variants (T. Wang et al., 2015; Wilson et al., 1977). Additionally, essential genes are often expressed more and produce proteins involved in a greater number of protein-protein interactions (Hart et al., 2014; Ishihama et al., 2008; Jeong et al., 2001; T. Wang et al., 2015). This shows that essential genes are involved in extensive protein networks and have significant roles in cellular functions.

Genes that have paralogous copies in the genome of *S. cerevisiae* have a decreased degree of essentiality, most likely as a result of at least some functional overlap (Gu et al., 2003). The paralogous genes were also found as less likely to be essential in human genome, which suggests that paralogs can reduce the essentiality of particular genes by providing functional redundancy at the cellular level (T. Wang et al., 2015). On the other hand, essential genes are enriched for key biological processes, including DNA replication, RNA transcription, and mRNA translation, which are vital for cell survival and proliferation (Subramanian et al., 2005).

The most comprehensive project to date, Project Achilles, aims to discover essential genes by methodically screening 1000 cancer cell lines (Tsherniak et al., 2017). In a study comparing these Achilles Project-specific essential genes and Neuronal Essential Genes (NEGs), it was found that they significantly overlapped in terms of risk genes associated with neurodevelopmental, neurodegenerative, and psychiatric disorders, although they were enriched in distinct modules and displayed distinct expression patterns (W. Zhang et al., 2021). After birth, expression levels of NEGs remained comparatively high, demonstrating their significant roles in adult neural functioning. Conversely, Achilles Project-specific essential genes were highly expressed throughout development but rapidly decreased after birth. Taken together, this study suggests that essential genes from cell-based screens, and in particular in vivo screens of healthy CNS tissue, can reveal potential risk genes and pathogenic variants that underlie human brain diseases.

Essentiality screens can be obtained from Project Score, DepMap project and their integrated analysis (Behan et al., 2019; Dempster et al., 2019; Meyers et al., 2017; Pacini et al., 2020). These datasets are included into one tool highlighting the context-specific essentiality of genes for cell fitness, called CEN-tools (Context-specific Essentiality Network-tools) (Sharma et al., 2020). CEN-tool facilitates the exploration of large-scale genetic screens that are publicly available in order to find correlations between specific contexts and gene essentiality. Furthermore, it creates dependency networks (CENs) that show the relationships between genes and specific context. These CENs offer a methodical technique to find genetic dependencies and uncover essential cell processes that are context-dependent, especially in cancer cells.

1.1.9 Correlation Between the mRNA and Protein Levels in the Cell

Deciphering the complex mechanisms of gene expression requires an understanding of the association between mRNA and protein levels in the cell. However, the genome-wide correlation between mRNA and protein expression levels has been found to be quite low within a number of studies (De Sousa Abreu et al., 2009; Gygi et al., 1999; Vogel & Marcotte, 2012). These genome-wide analyses determine the correlation coefficient between the expression of mRNA and the corresponding expression of proteins for a wide variety of genes (Anderson & Seilhamer, 1997; Ideker et al., 2001; Ørntoft et al., 2002; Washburn et al., 2003). According to some studies, some classes of genes have higher correlations than others, because they are generally expected to be in tight synchronization. Examples of these include time-dependent cell cycle genes (Greenbaum et al., 2003; Ly et al., 2014), secreted proteins (De Godoy et al., 2008; Kawamoto et al., 1996) and bacterial genes optimized for rapid translation (Jayapal et al., 2008). The correlation of individual genes between samples or experimental circumstances has only been examined in a small number of research (G. Chen, Gharib, Huang, Taylor, et al., 2002; G. Chen, Gharib, Huang, Thomas, et al., 2002; M. L. Fournier et al., 2010; Ideker et al., 2001; Jayapal et al., 2008; Shankavaram et al., 2007). Some of these studies simply show the variations in the correlation coefficients to illustrate the fact that the expression of mRNA and protein can occasionally be discordant (G. Chen, Gharib, Huang, Taylor, et al., 2002; Shankavaram et al., 2007). Some of the other studies uncover proteins that are differently expressed and look at the related mRNA expression, usually detecting 0–50% of the mRNA levels (G. Chen, Gharib, Huang, Thomas, et al., 2002; Ideker et al., 2001), while others identify groups of concordant and discordant mRNA-protein pairs by performing clustering or similar analyses on expression profiles across sample (M. L. Fournier et al., 2010; Jayapal et al., 2008; Tian et al., 2004). Compared to non-

differentially expressed mRNA profiles, it was also shown that differentially expressed mRNA profiles showed a substantial trend toward stronger correlations with their protein products (Koussounadis et al., 2015). Some of the non-differentially expressed mRNA profiles had strong relationships with their protein product for non-experimental reasons, and some of the low and even negative associations were also found within the profiles of the differentially expressed mRNAs.

Furthermore, studies have indicated that the distribution of protein and mRNA levels is not normally distributed (Futcher et al., 1999; Ghaemmaghami et al., 2003). Even though the Spearman rank coefficient, which does not imply a normal distribution (Caruso & Cliff, 1997), is a better fit for the calculation of the correlation than the Pearson correlation coefficient, which measures a linear correlation (Lee Rodgers & Nicewander, 1988), they both produced similar results and are used equally in the literature for the correlations between mRNA-protein levels (Maier et al., 2009).

1.1.10 Prioritization of Genes

In bioinformatics and computational biology, gene prioritization is an essential component that seeks to identify genes most likely to be associated with particular phenotypes or diseases. To prioritize the candidate genes, a number of methods and tools have been created, utilizing a variety of computational approaches and data sources (Adie et al., 2005; Freudenberg & Propping, 2002; Thornblad et al., 2007; Turner et al., 2003).

Finding the genes linked to both wellness and disease is a recurring problem. Therefore, Aerts et al. presented an interactive program and a bioinformatics technique, called

ENDEAVOUR, that ranks potential genes associated with diseases or biological processes according to how similar they are to known genes implicated in these phenomena (Aerts et al., 2006; Tranchevent et al., 2008). This strategy uses a training set of genes that are known to be linked to particular biological processes. It infers models from several genomic data sources, ranks candidate genes based on the characteristics of known genes, and combines many rankings into a global rating.

When examining the genetic underpinnings of disorders discovered by complex-trait linkage studies, the regions of interest, which frequently encompass hundreds of genes, often span enormous genomic areas. There are two types of candidate gene prioritization techniques: functional annotation-based (Tranchevent et al., 2016) and sequence feature-based (Juncker et al., 2009). Due to incomplete annotations for some genes, functional annotation approaches may be biased. Sequence-based methods prioritize genes based on overall potential involvement in disease generally rather than the particular disease of concern (Jalilvand et al., 2019). SUSPECTS presents a unified methodology that blends the accuracy of annotation-based techniques with the recall of sequence-based approaches (Adie et al., 2006). Using the idea that genes associated with specific diseases have comparable annotations, suggesting shared biological pathways, this novel method automatically selects candidate genes from broad genomic areas for complicated or oligogenic disorders.

High-throughput genome-wide studies are useful for the characterization and classification, however, they are not always effective in determining or ranking the genes that cause a specific disease. In order to close this gap, Chen et al. presents a unique method for prioritizing disease-causing genes by using mouse phenotypic data (J. Chen

et al., 2007). The ToppGene platform is presented and shown to outperform existing candidate gene prioritizing techniques, SUSPECTS and ENDEAVOUR, using a variety of data integration techniques (J. Chen et al., 2007). The integration of phenotypic data from mouse orthologs of human genes is emphasized as a noteworthy advancement in the evaluation and ranking of candidate genes for human disorders, offering a greater understanding of the genetic base of diseases.

Chen et al. have introduced the ToppGene Suite, which encompasses the enrichment analysis of the gene list and the prioritization of the candidate genes (J. Chen et al., 2009). This tool provides valuable insights into disease-related proteins, genetics, and candidate genes. It performs the prioritization according to the functional resemblance to training gene list using a fuzzy-based similarity measurement to calculate the similarities among any two genes based on the topological features and semantic annotations in protein–protein interaction network.

In another study, Yankina et al. utilized the ANDSystem, a computer system for gene network reconstruction developed by Ivanisenko et al., to prioritize genes associated with major depressive disorder (MDD) (Ivanisenko et al., 2015; Yankina et al., 2018). ANDSystem is an automated information mining system for scientific papers and databases concerning the interactions of proteins, genes, metabolites, microRNAs, illnesses, biological processes, phenotypic characteristics, etc. (Ivanisenko et al., 2015). Yankina et al. investigated how apoptosis functioned in the associative gene network of MDD, using prioritizing techniques to identify potential genes for additional study (Yankina et al., 2018). They used a hybrid strategy consisting of previously established standard methodologies together with the ANDSystem to assess the prioritizing methods

of the extrinsic apoptotic signaling pathway genes. The importance of prioritization in determining the genes involved in complex phenotypic characteristics, such as asthma and hypertension was also emphasized by using well-known techniques like ToppGene and Endeavor as well as a cross-talk centrality criteria that was obtained from associative gene networks that were examined using ANDSystem (Saik et al., 2018).

There are several studies that have proposed novel gene prioritization methods, different than the others. Wang et al. presented probabilistic knowledge model for gene prioritization, which is an enhanced strategy for link analysis that prioritizes candidate genes linked to certain diseases (S. Wang et al., 2009). This approach is built using more reasonable distributional assumptions. Utilizing a collection of known seed genes linked to the disease, the method entails developing a protein-protein interaction network from a curated database and utilizing the enhanced link analysis method to uncover novel genes likely connected with the condition. Martinez et al. (2014) introduced ProphNet, a generic prioritization algorithm that integrates and propagates information across any number of heterogeneous data networks (Martínez et al., 2014). ProphNet is generic and enables the priority of biological entities of any kind in relation to entities of a different kind, providing customizable flexibility for a range of prioritization tasks (e.g., disease-gene, domain-disease, drug-disease). Its capacity to combine more than two network types is a crucial component that helps it get around a drawback in current network-based prioritizing techniques.

Another novel gene prioritization method is called MGOGP (Module and Gene Ontology-based Gene Prioritization), which considers gene module information rarely utilized in cancer-related prioritization (Su et al., 2018). Unlike other approaches,

MGOGP takes into account data from both individual genes and the modules that have associations to them. It uses known cancer-related genes as well as fuzzy measure values based on Gene Ontology (GO) as heuristics for gene ranking. Moreover, Li et al. demonstrated the robustness of the PEP method, which stands for Prioritizing genes using Expression information in PPI (Protein–protein interaction) networks with disease risks transmitted across genes, in prioritizing candidate genes for chronic obstructive pulmonary disease (COPD) (W. Li et al., 2021). PEP is based on network propagation, which describes the transmission of disease risks between genes influencing each other. This technique uses expression data to create a weighted COPD PPI network. Then, in descending order, disease candidate genes are ranked according to the related disease risk scores. When using expression data from sequencing or microarray data, the technique works well and is reliable for ranking potential genes for COPD.

In addition to these approaches, there are integrative approaches that have been proposed for gene prioritization. Zhai et al. first developed RafSee, which combines plant sequencing, evolutionary, and epigenetic information using a random forest method (Zhai et al., 2016). Then, an integrative strategy called RAP (Rank Aggregation-based data fusion for gene Prioritization) was developed to improve prediction accuracy by fusing a network-based gene prioritization technique with the ranks from RafSee. Another study presented an integrative network-driven pipeline for prioritizing genes related to Alzheimer's disease, highlighting the importance of integrating heterogeneous data in disease gene prioritization (Browne et al., 2014). The pipeline incorporates many heterogeneous data sources, such as gene expression, betweenness measurements, protein-protein interaction networks, and similarities based on ontology. Tissue-specific gene expression data are also included in the evaluation procedure.

Additionally, various computational techniques have been used to help prioritize genes. Kanduri and Järvelä have presented GenRank, an Bioconductor package of R programming for candidate gene prioritization, which fills a gap in the methods of the prioritization of candidate genes within the project of Bioconductor (Kanduri & Järvelä, 2016). It offers methodologies which rely on converging evidence from multiple independent levels of evidence and includes three methods for the integration of gene-level data from multiple levels of evidence (Rank Product method, Convergent Evidence method and Combining p-values). To overcome the difficulties in cancer-associated gene prioritizing, Wu et al. (2012) developed Networked Gene Prioritizer (NGP), a technique that combines gene expression and protein-protein interaction networks to prioritize the genes that are associated with cancer (C. Wu et al., 2012). It assumes that differential expression of interacting genes is caused by cancer-associated genes in comparing samples.

In summary, for the identification of candidate genes linked to the diseases and phenotypes, gene prioritization plays a central role. The wide variety of methods and tools that have been developed for gene prioritization make use of computational algorithms, integrative approaches, and diverse data sources to provide valuable insights into the genetic basis of disease and biological function. However, there is no online RNA-seq analysis tool yet that make use of essential gene scores or the level of congruence between mRNA and protein levels of genes in gene/cluster prioritization.

1.1.11 Online RNAseq Analysis Tools

Large volumes of data have been produced as a result of the rise in RNAseq investigations, and it now became crucial to analyze and interpret these data properly. Nevertheless, such an analysis needs bioinformatics expertise at the comment line unless the platform is available as a vignette or online since many molecular biology researchers may not have adequate knowledge in programming. Because of this, researchers have been developing web-based tools of which even someone with minimal or no background in bioinformatics may use. Due to these user-friendly online tools, researchers can now perform advanced analyses and apply established pipelines to their own datasets without the need to learn how to code.

It is important that these applications are simple to use and have repeatable outcomes. RNA-seq data analysis may be achieved using a variety of online tools that are accessible in the literature. These tools differ in terms of their graphical user interface, functionality, and options for the user. Table 1 provides an overview of the features shared by these applications that are available in the literature and several are discussed below in further detail.

1.1.11.1 iDEP

iDEP was developed as a Shiny application to conduct differential gene expression analysis of RNA-seq data along with clustering and pathway analysis (Ge et al., 2018). Several versions of iDEP have been created, and each has included improvements to the previous version. As a result, the most recent version of iDEP will be referred to in this context.

In iDEP, users must submit an RNA-seq dataset containing raw read count values, FPKMs values or fold-changes, and, optionally, a metadata file in order to utilize this program. It allows the filtering of lowly expressed genes and performs DGE by DESeq2 or limma methods. After DGE analysis, up/down-regulated genes can be used to perform pathway analysis using several options such as GO terms, TD target genes, KEGG pathways, etc. In the last version of iDEP, these genes can also be visualized on the genomes. For the clustering, it performs k-means and hierarchical clustering along with the PCA analysis, using data transformed as vst, rlog or logCPM. For the clustering, firstly, genes are sorted by their standard deviation among samples in transformed data and the user can select the number of genes from this data to cluster. After clustering, enrichment analysis using gene sets such as KEGG, MSigDB, PANTHER, etc. can be run for each cluster and results are visualized as a table, tree or network. Co-expression networks can be also identified using WGCNA (Langfelder & Horvath, 2008) and results are shown as network, heatmap or module plots. iDEP also allows user to download and use datasets generated by ARCHS4 (Lachmann et al., 2018) and DEE2 (Ziemann et al., 2019), and can be used for different species such as human, mouse, zebrafish and etc.

1.1.11.2 GENAVi

Another Shiny-based online application for the analysis of gene expression is GENAVi (Gene Expression Normalization Analysis and Visualization) (Reyes et al., 2019). Users of this application can submit their own dataset as count data or browse a panel of 20 cell lines that are often utilized in breast and ovarian cancer research. It allows the user to select the genes by several options: entering the name of genes, uploading a gene list, selecting 1000 genes having the highest variance or selecting the rows from the table. In

the visualization part of GENAVi, expression heatmaps for the selected genes; PCA plot for the top 1000 variable genes, all or selected genes; and correlation heatmap of all samples and genes can be displayed. GENAVi requires uploading metadata to perform DGE analysis using the DESeq2 package of R. Then, it performs a volcano plot with several parameters such as fold change and p value thresholds. For the enrichment analysis, it offers gene set enrichment analysis (GSEA), Over Representation Analysis (ORA), WikiPathways analysis, MSigDB analysis, Gene Ontology Analysis and KEGG Analysis. Results are visualized as a table, dot plot or enrichment map.

1.1.11.3 DEBrowser

DEBrowser is also another tool that utilizes Shiny for the RNA-seq datasets, and it is also developed as an R Bioconductor package (Kucukural et al., 2019). It requires uploading count data and metadata files. It performs first filtering of non and lowly-expressed genes and then normalization of the data using different methods such as TMM, MRE, RLE and quartile normalization. It also has the unique feature of batch effect correction using the Harman or SVA/Combat methods (Leek, 2014; Oytam et al., 2016). DEBrowser also includes k-means clustering of differential regions to enhance the differential analysis process, which can be used in gene ontology enrichment analysis.

1.1.11.4 RNfuzzy

RNfuzzy is another online application developed to analyze RNAseq data and performs fuzzy clustering (Habermann & Haering, 2021). It accepts raw or normalized read counts. It has features similar to the others, such as filtering out genes expressed at low levels,

data normalization, hierarchical clustering and PCA analysis. The unique features of RNfuzzy are the Mfuzz clustering for larger experimental setups, which performs clustering of genes based on their expression patterns over time (L. Kumar & Futschik, 2007), and the enrichment of clusters, which is not available in many applications. Additionally, it performs orthology assignments and ID conversions, which are helpful when studying human disorders on mice.

1.1.11.5 BEAVR

Another tool developed to automatically analyze and explore RNAseq datasets is BEAVR (Perampalam & Dick, 2020), which is only available as a Docker container. It requires raw count data and metadata to be uploaded and performs DGE analysis by DESeq2. Its unique feature is to create a Pathway Enrichment Map which results in an interconnected map of pathways. The other features are similar to other tools and given in Table 1.

1.1.11.6 WebMeV

A cloud-based application, WebMeV (Multiple Experiment Viewer), is designed to analyze, visualize, and stratify large genomic data, in particular the RNA-Seq and microarray data (Y. E. Wang et al., 2017). While it accepts raw count data, it also allows the download of public datasets from TCGA, TARGET, or GTEx. Clustering can be performed as a hierarchical clustering or using PCA or k-means clustering. The unique features of WebMeV are batch correction, regulatory network analysis using LIONESS (Kuijjer et al., 2019), PANDA (Glass et al., 2013) or PUMA (Kuijjer et al., 2020).

WebMeV also offers single-cell RNA-seq data analysis and IGV visualization, which is not found in the previously mentioned tools that perform bulk RNA-seq analysis.

1.1.11.7 3D RNA-seq

3D RNA-seq is another RNA-seq analysis tool that provides differential expression and alternative splicing data analysis through a web-based R-Shiny application (Guo et al., 2021). Different from others, it requires 3 input files: metadata, a folder containing transcript quantification outputs and a transcript-gene mapping file. The unique features of 3D RNA-seq are the ability to perform analyses of differential alternative splicing, differential transcript usage, time-series trend and isoform switching for pairwise and time-series datasets. Other features of 3D RNA-seq are shown in Table 1.

1.1.11.8 Omics Playground

Omics Playground is developed to explore big omics data, that include transcriptomics and proteomics datasets (Akhmedov et al., 2020). It requires a count/expression file and a sample meta-data file. It offers many different tools to analyze, visualize and interpret the data, such as low count filtering, PCA/MDS, tSNE, DGE analysis, correlation plots, scatter plots, heatmap, enrichment analysis and interaction network visualization. The unique features are biomarker selection, survival analysis, single-cell profiling, and drug connectivity analysis. Omics Playground has been greatly enhanced in terms of enrichment analysis; it allows user to assess their gene signature by calculating an enrichment score. The platform also provides specialized functional analysis tools to understand better the biological pathways, including GO, KEGG, and drug connectivity

mapping. Additionally, it offers WordCloud analysis, also known as 'keyword enrichment' analysis, which evaluates the enrichment of keywords for the contrasts. This analysis method identifies words frequently appearing in the top-ranked gene sets, providing an unbiased description of the contrast.

1.1.11.9 FungiExpresZ

FungiExpresZ is web-based tool designed to analyze and visualize fungal gene expression data (Parsania et al., 2023). While it is specific to fungal species, it is one of the tools that performs k-means clustering. It also allows users to upload their data or select public datasets from NCBI-SRA (Leinonen et al., 2011) FungiExpresZ is enhanced by providing 12 different visualizations, including pairwise scatter plot, PCA plot, correlation heat box, box plot, density plot etc., and performs GO enrichment and its visualizations. One of its key features is its ability to do a functional annotation for a user-selected cluster.

Table 1. Comparison of multiple features of the available shiny applications and CAP-RNAseq

Apps / Features	CAP-RNAseq	iDEP	GENAVi	DEBrowser	RNfuzzy	BEAVR	WebMeV	Omics Playground	FungiExpresZ	3D RNA-seq
Availability										
Online server	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Docker container	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
as R package	No	Yes	No	Yes	Yes	No	No	No	Yes	No
Core functionality										
Upload count file	Yes (raw counts)	Yes (raw counts, FPKMs values or FC)	Yes (raw count)	Yes (raw count)	Yes (raw or normalized read counts)	Yes (raw count)	Yes (raw count)	Yes (raw count/ expression file)	Yes (raw or normalized gene expression counts)	Yes (Transcript quantification file)
Upload metadata	Yes	Optional	Yes	Yes	No	Yes	Optional	Yes	Optional	Yes
Import from online data repositories	No	Yes	No	No	No	No	Yes	No	No	No
Data Processing										
Data filtering step	Yes	Yes	No	Yes	Yes	Yes	No	Yes	No	Yes
vst+ANOVA to filter the gene	Yes	No	No	No	No	No	No	No	No	No
Data normalization	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Batch effect correction	No	No	No	Yes	No	No	Yes	Yes	No	Yes
Clustering										
Hierarchical clustering	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
k-means clustering	Yes	Yes	No	Yes	No	No	Yes	No	Yes	No
PCA	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Apps / Features	CAP-RNAseq	iDEP	GENAVi	DEBrowser	RNfuzzy	BEAVR	WebMeV	Omics Playground	FungiExpresZ	3D RNA-seq
Differential gene expression analysis										
Limma/voom	Yes	Yes	No	Yes	No	No	Yes	Yes	No	Yes
DESeq2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
edgeR	No	No	No	Yes	Yes	No	Yes	Yes	No	No
Gene set enrichment analysis										
Enrichment analysis of DEG	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Enrichment analysis of clusters	Yes	Yes	No	No	Yes	No	No	Yes	Yes	No
Databases	MSigDB	GO, KEGG, TransmiR, MSigDB, TF target genes, miRNA target genes, WikiPathways, etc.	GSEA, ORA, WikiPathways, MSigDB, GO, Disease Ontology and KEGG	GO, KEGG and disease ontology	GO, Wikipathways, Reactome, KEGG, Human Protein Atlas, CORUM data on protein complexes and TRANSFA	GSEA and ORA	Fast Gene Set Enrichment Analysis (fgsea), WGCNA, TopGO enrichment analysis, SNF (Similarity Network Fusion)	More than 42 published reference databases such as MSigDB, Wikipathways, and GO	GO	GO

Apps / Features	CAP-RNAseq	iDEP	GENAVi	DEBrowser	RNfuzzy	BEAVR	WebMeV	Omics Playground	FungiExpresZ	3D RNA-seq
Visualizations										
Dissimilarity heatmap (mirror clusters)	Yes	No	No	No	No	No	No	No	No	No
Heatmap for sample correlation	Yes	No	Yes	No	No	Yes	No	No	No	No
Network visualization	Yes	Yes	Yes	No	No	Yes	Yes	No	No	No
DepMap essentiality plots	Yes	No	No	No	No	No	No	No	No	No
DepMap RNA~Protein correlations	Yes	No	No	No	No	No	No	No	No	No
HPA visualization	Yes	No	No	No	No	No	No	No	No	No
Protein expression levels	Yes	No	No	No	No	No	No	No	No	No
Volcano plot	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Scatter plot	No	Yes	No	Yes	No	No	No	Yes	Yes	No
MA plot	No	Yes	No	Yes	Yes	No	No	Yes	No	No

1.2 Senescence and Gene Essentiality

It has been more than 50 years since the cellular senescence was formally identified as a mechanism that limits the ability of healthy human cells to proliferate in culture (Hayflick, 1965). In the literature, cellular senescence may have contradictory roles during survival and aging (Rodier & Campisi, 2011). Senescence inhibits the ability of cells to become tumors by preventing proliferation and has significant implications for tissue repair, regeneration and embryonic development; yet populations of senescent cells can cause cancer development, age-related diseases, and inflammation (Herranz & Gil, 2018). Indeed, cellular senescence is a complex process, characterized by irreversible arrest of cell cycle, which is induced by several factors, such as telomere thinning, oncogene activation, oxidative stress, mitochondrial dysfunction, genotoxic stress and DNA damage (Anerillas et al., 2022a; Campisi & d'Adda Di Fagagna, 2007; Kuilman et al., 2010). Other than the irreversible growth arrest of cells, senescent cells also display several traits such as morphological alterations and metabolic reprogramming including modified lysosomal function and elevated levels of senescence-associated β-galactosidase (SA-β-Gal) activity that is a senescence biomarker (Dimri et al., 1995; Gorgoulis et al., 2019).

When cells experience senescence, whether due to DNA damage, mitochondrial dysfunction, telomer dysfunction or oncogenic mutations, they display significant alterations in their transcriptomes (Figure 2). One important result of cells going through senescence is the release of several substances, including cytokines and chemokines (Acosta et al., 2008; Campisi, 2005). Numerous studies, particularly those involving fibroblasts going through replicative senescence, have found this change in the secretome

of senescent cells (Krtolica et al., 2001). Strong inflammatory response in senescent fibroblast cells as observed in wound healing has been shown by microarray analysis, suggesting that senescent cells may play a role in chronic wound diseases (Shelton et al., 1999).

Each cell or tissue has a unique secretome of senescent cells, which depends on healthy or pathological circumstances (Zullo et al., 2015). Chemokines and other factors found in secretomes of senescent cells can lead to the accumulation of certain types of immune cells that cause senescent cells to develop the senescence-associated secretory phenotype (SASP), one of the most notable features of senescent cells, which ensures the viability and metabolic activity of the senescent cell (Birch & Gil, 2020; Cotarelo et al., 2016). SASP is defined by the active secretion of proinflammatory factors and enzymes that remodel the matrix, which is a phenotype that has the ability to profoundly affect physiological and disease processes as well as the cellular microenvironment. The secretion of pro-inflammatory factors can cause pro-oncogenic effects such as increased proliferative rate, migration and invasion of premalignant cells (Dilley et al., 2003; Krtolica et al., 2001).

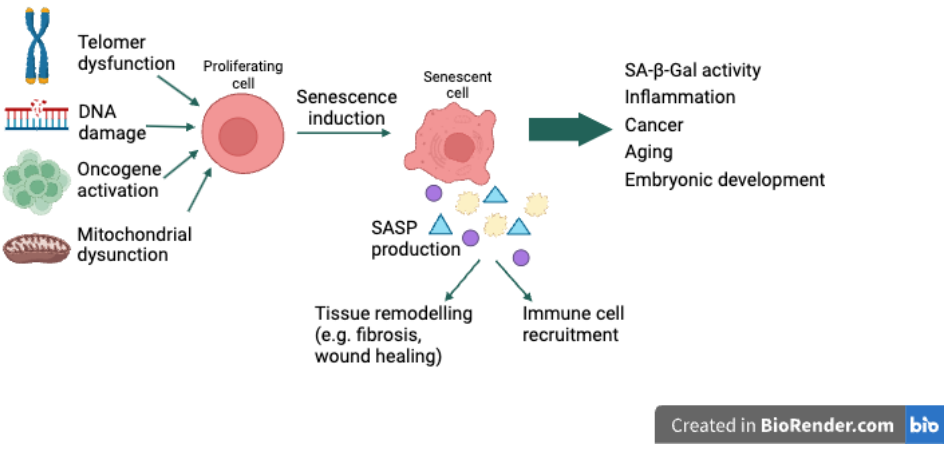


Figure 2. Mechanisms of cellular senescence and its consequences.

(This image was created by using BioRender.com).

Senescent cells when compared to normal proliferating cells, they appear to be more resistant to apoptosis in a cell specific manner. For instance, endothelial cells do not show resistance to ceramide-induced apoptosis, while senescent human fibroblasts do (Hampel et al., 2004). Senescent human fibroblasts are also resistant to oxidative stress-induced apoptosis (Sanders et al., 2013; Sasaki et al., 2001), cytotoxic drugs (Tepper et al., 2000), serum deprivation (E. Wang, 1995), UV radiation (Chaturvedi et al., 1999; Marcotte et al., 2004). The remarkable stability of senescent cells in culture can be partly attributed to their resistance to apoptosis. The observed increase in the number of senescent cells with age may also be explained by this feature of senescent cells.

Recent developments in the CRISPR-Cas9 system have completely changed large-scale gene function screens, providing more precise and reliable findings than traditional approaches. In a previous study, a large-scale pooled CRISPR-Cas9 knockout screen was carried out in primary human dermal fibroblasts to find the essential genes for senescence (X. Liu et al., 2019). It revealed different transcriptome patterns in bypass cells, which some has neutralized SASP expression profiles, and downregulated pathways linked to age-related neurodegenerative diseases.

Essential genes were also investigated in the context of the aging process. Essential genes, especially those related to growth and development, are highly conserved in evolution and have a key role in regulating life span (Curran & Ruvkun, 2007). The highly conserved essential genes across species are involved in a variety of cellular processes, such as gene regulation, metabolic regulation, genetic stability and stress tolerance (Jazwinski, 2000). Inactivation of most of them increase longevity to an extent comparable to the most potent known anti-aging regulators. It was reported that growth

arrest was induced by early gene inactivation of some essential genes in larval stages and compared to adult wild-type animals, some of these arrested species had substantially longer lifespans (Curran & Ruvkun, 2007). In a previous study hypothesizing that one unexplored source of pro-longevity genes might be the conserved essential genes. It was found that overexpression of conserved essential genes in budding yeast caused increased replicative lifespan (Oz et al., 2022). Moreover, a positive correlation was found between aging-related changes and signatures of long-lived animals, which are enriched for essential genes that have existed for a long time throughout evolutionary history (Tyshkovskiy et al., 2023). These genes were associated with the degradation of proteins and the pathway of PI3K-Akt signaling. However, the relationship between essential genes and cellular senescence, is still mostly unsolved due to limited research available. A lack of study in this field reveals a significant knowledge gap about the ways in which essential genes affect or modulate aging, and it suggests more investigation is required to clarify their possible function in this intricate biological process.

1.3 Essential Genes and Cancer in General

One of the main uses of CRISPR technology, which enables to discover the essential genes for survival of cells, is the examination of the effect of gene inactivation on the selective decrease of cancer cell viability (Behan et al., 2019; T. Wang et al., 2015). The distinct cancer vulnerabilities that depend on the genetic background have been discovered by the comprehensive CRISPR screenings on large panels of immortalized tumor cell lines (Dempster et al., 2019; Martinez-Lage et al., 2018; Tsherniak et al., 2017).

The analysis of essential genes in cancer cells can reveal the prognostic markers or therapeutic targets in the diseases. For example, in the analysis of the finding the genes essential for osteosarcoma cell, these genes were found to be enriched in ribosome pathway and specifically, Rps28 has been revealed as essential gene for osteosarcoma cell lines (C. Liang et al., 2024). Moreover, the proliferation essential genes (PLEG or PEG) were identified for the colon cancer through the DepMap database, and these upregulated PLEGs in the colon cancer were found to be enriched in cell cycle pathway, mismatch repair, DNA replication and ribosome biogenesis in eukaryotes (J. Liu et al., 2023). This model for proliferation essential genes was also predictive of the prognosis of colorectal cancer patients. After undergoing chemotherapy, patients in the PLEGs-High subtype had a worse prognosis than those in the PLEGs-Low group. Moreover, in the study of bladder cancer (BLCA), which investigate PEG (also called PLEG) signature as a prognostic biomarker for accurately determining how BLCA would react to chemotherapy, compared to patients with PEGs-low, individuals with high PEG signature scores (PEGs-high) showed reduced susceptibility to treatment and shorter overall survival, and POLE2 gene was found as the most significant PEG linked to the advancement of BLCA, which was up-regulated in tumor tissues' mRNA and protein levels (L. Yu et al., 2024).

Additionally, essential genes are also used in the prediction of drug response (Vazquez & Boehm, 2020). It was observed that addition of essential gene information to a method developed for the drug response prediction increased the accuracy of the model (P. Wu et al., 2023). In the context of evolutionary analysis, the essential genes in lung cancer subtypes revealed that genes that are highly expressed and essential for biological processes have a higher probability of being conserved over time, and these conserved

genes rarely exhibit an important alterations in gene expression (Freischel et al., 2022). Moreover, the role of essential genes was analyzed to comprehend the genetic and functional effects of adaptive Copy Number Variants (CNVs) in Saccharomyces cerevisiae (yeast) strains (Avecilla et al., 2023). It was found that CNVs drive higher mutation tolerance in amplified essential genes.

The association of essential genes with brain cancer and breast cancer is described in detail below.

1.3.1 Essential Genes in Brain Cancers and Brain Stem Cells

To create effective treatments and therapies, it is important to understand the essential genes in brain malignancies and brain stem cells. Finding these genes can provide light on the biological mechanisms behind tumor development, treatment resistance, and the risk of recurrence.

In the analysis of three gene regulatory networks which are human brain-specific, biologic significance of the critical genes was investigated (Z. Chen et al., 2022). It was revealed that these critical genes were found to be more abundant in essential genes, genes related to neuron disease and cancer than ordinary genes, suggesting that critical genes can be biomarkers in the brain-specific gene regulatory network. Pathway analysis of these critical genes revealed the fact that these genes are linked to brain tumor and enriched in pathways in cancer, transcriptional misregulation in cancer, etc.

Determining hub genes and understanding the molecular process behind Low-grade gliomas (LGG) would be made possible by the alterations in the gene expression profile resulting from the CRISPR/Cas9 screen of the DepMap database. The combined DepMap and TCGA survey revealed 145 oncogenes that are abundantly expressed and lethal to LGG and found that the primary players in functional regulation in LGG are genes linked to the Hippo signaling system (Mijiti et al., 2023). In another study using data for glioma-associated genes from CRISPR-screens of the DepMap, 10 cancer-essential ferroptosis-related genes (CHMP6, LONP1, RRM2 HSPA5, MTOR, RPL8, NFS1, ISCU, EIF2S1 and AURKA) have been identified that are closely associated with ferroptosis which is iron-dependent kind of programmed cell death (Dixon et al., 2012), and are required for glioma cell proliferation and survival (Yun et al., 2022).

Recent glioma studies employing the CRISPR/Cas9 technique have allowed researchers to identify a number of factors, including USP8 and SOCS3, that are essential for the development and proliferation of glioma cells (MacLeod et al., 2019). In another study using same technique to understand the signaling pathways underlying glioblastoma invasion, 33 genes essential for invasion were identified, among which mitogen-activated protein kinase 4 (MAP4K4) was found to be the strongest regulator of invasion, and inhibition of it prevented tumor cell migration (Prolo et al., 2019). CRISPR-Cas9 knockout screen have also been used to identify more than 100 novel genes essential for the growth of cancer stem cells (CSC) arising from pediatric high-grade glioma (HGG) with the H3K27M mutation (Wenger et al., 2023). Loss of these genes caused the significant decrease in CSC growth and an alteration in stem cell and differentiation markers.

Furthermore, genome-wide loss-of-function screens on human pluripotent stem cells (hPSCs) revealed that the essential genes enriched in hPSCs were related to DNA repair and cell cycle pathways (Yilmaz et al., 2018). In another study, it was shown that essential genes in hPSCs exhibit significant differences in comparison to different human model cell lines, and they are extremely context-specific in hPSC based on the type of growth substrate used (Mair et al., 2019). In the study focusing on the identification of the essential genes required for hPSCs to differentiate into the three embryonic germ layers, many of them were found to be linked to the plasma membrane, emphasizing the signaling pathways required for the differentiation of each lineage (Yilmaz et al., 2020).

Essential genes are strongly expressed in mouse embryonic stem cells (mESCs), tend to lack paralog genes, and are enriched for fundamental biological processes such as cell cycle, DNA metabolism and RNA processing (Shohat & Shifman, 2019). Essential genes specific to mESCs contribute to processes related to their pluripotent status. Given that ESCs are used as a model to investigate various human diseases, researchers have found that the human traits most prominently linked to the genes required for ESCs are neurodevelopmental.

1.3.2 Essential Genes in Breast Cancer

CRISPR-Cas9 screens were also used for the identification of essential genes in breast cancer. In a study using DepMap and TCGA dataset, enrichment analysis revealed that 86 oncogenes that were overexpressed and had a fatal effect on breast cancer were enriched in cell cycle pathways (Sun et al., 2021). In particular, the MCM2 and MCM4

genes, which are involved in initiating DNA replication (Forsburg, 2004), have been identified as essential for breast cancer survival.

To enhance the understanding of molecular processes in breast cancer, researchers have developed a strategy, called OncoOmics, to analyze a range of genomic data to reveal essential genes in breast cancer (López-Cortés et al., 2020). In this analysis, 140 essential genes for breast cancer were identified, with the most common genomic alteration being mRNA up-regulation. The molecular subtypes which are Her2-enriched and basal-like were linked to the most affected signaling pathways. In another study developed an epigenome-wide CRISPR-Cas9 knockout library (EPIKOL) targeting the variety of epigenetic regulators and their cofactors, 15 essential genes were identified in TNBC cell lines, which contains many known cancer cell fitness regulators (Yedier-Bayram et al., 2022). Notably, SS18L2 was emerged as one of them, which involves in cell cycle progression. The knockout of SS18L2 caused the decreased expression of a specific group of genes related to cell cycle.

CRISPR/Cas9 screen in breast cancer for essential genes regulating the development and metastasis of disease prioritized ZNF319, a gene whose higher expression in nearly all subtypes of breast cancer is associated with a better clinical outcome (L. Wang et al., 2022). The transcriptome study of breast cancer cells that have been silenced for ZNF319 reveals that ZNF319 is engaged in a variety of biological processes and signaling pathways, including those related to the cell cycle and proliferation. PEGs were also identified in patients with triple-positive breast cancer (TPBC) (X. Li et al., 2024). Of the 900 TPBC-PEGs, 437 exhibited a considerable variation in TPBC expression compared to normal tissues. Compared to patients with low PEG signature scores, those with high

values had a shorter overall survival and less sensitivity to neoadjuvant chemotherapy (NACT).

It has been extensively shown that several Nuclear Receptors (NRs), including the androgen and estrogen receptors, are functionally engaged in cancer and can be used as therapeutic targets and instructive biomarkers in oncology since they play important roles in many physiological processes, including immunity, development, metabolism and homeostasis (Gronemeyer et al., 2004; Zhao et al., 2019). Compared to normal tissue, the NR expressions were decreased in tumors and significantly cancer-type specifically, and 10 NRs as highly selective essential genes for cancer cell growth were discovered by the analysis of large-scale CRISPR and RNAi screens (Jiang et al., 2022). Despite these results, there is still not enough information in the literature on the specific relationships between the essential genes found using CRISPR/Cas9 screening and the steroid hormones such as estrogen or progesterone.

1.4 Brain Cancers

As a result of abnormal growth of brain cells, brain cancers can develop in various parts of the brain, including the cerebellum, cerebrum, and brain stem. Based on the site of origin, more than 120 types of brain tumors have been classified (Mohammed et al., 2019). There are two main categories of brain cancers: primary (originating in the brain) such as gliomas and secondary (spreading from other places in the body to the brain) that is indicative of metastasis. The secondary tumors are more prevalent with a five to ten times higher frequency than primary tumors (Gavrilovic & Posner, 2005). This could be attributed to the growth rates of brain tissues, even the glia, with respect to other tissues

that can become tumorigenic. The most common tumors to spread to the brain are lymphoma, melanoma, leukemia, lung, kidney, and breast malignancies (Pekmezci & Perry, 2013). While cancer is the primary cause of death worldwide, responsible for nearly 10 million fatalities in 2020, the 10th largest cause of death is brain cancer and associated central nervous system (CNS) malignancies (S.-S. Kim et al., 2015; World Health Organization: Geneva, Switzerland, 2022).

Data from the World Health Organization (WHO) indicate that there were 308,102 cases of brain and CNS cancers worldwide in 2020, with 251,329 deaths in both sexes and all age groups (H. Sung et al., 2021). The most common malignant primary tumors of the brain are gliomas (Ceccarelli et al., 2016; Louis et al., 2016). They account for around 80% of malignant tumors and 30% of primary brain tumors (Mukherjee et al., 2023). The most prevalent kind of primary brain tumor is glioblastoma and other gliomas, which have an incidence of 5 in 100,000 in the US population (Kleihues et al., 2000; Legler et al., 1999). Gliomas may arise from any of the several glial cell morphologies, including astrocytes, oligodendrocytes, microglia, and ependymal cells. Gliomas are traditionally classified based on their histological characteristics, which are microscopic resemblance to the neuroglial cell types in the healthy brain (such as astrocytes, oligodendrocytes, etc.) and malignancy degree (Grades I–IV) according to how aggressive the malignancy is (Louis et al., 2016; Parodi et al., 2019). WHO classified glioblastoma as a grade IV glioma, which is the most aggressive and proliferative one (Louis et al., 2007). Adult cases of gliomas include infiltrative astrocytomas of different grades, including diffuse astrocytoma (WHO grade II), anaplastic astrocytoma (WHO grade III), glioblastoma (WHO grade IV), as well as oligodendrogliomas, and the controversial mixed oligoastrocytoma syndrome (Louis et al., 2007). However, Low-Grade Gliomas (LGG)

correspond to grades I and II since they are slower growing and lack aggressive characteristics (Alghamri et al., 2021).

1.4.1 Glioblastoma Multiforme (GBM)

One of the deadliest brain tumors in the world, glioblastoma multiforme (GBM), results in deaths of numerous people each year. It is a very fast-growing form of cancer that spreads quickly throughout the brain, which makes it difficult to treat (Stensjøen et al., 2015). In addition to being a very invasive and recurrent disease, GBM has a poor clinical prognosis and a high death rate with 12-15 months of median survival rate and less than 5% of patients surviving for more than five years (Campos et al., 2009; D. Wang et al., 2019). GBM exhibits complex heterogeneity which can be categorized into two areas: the tumor inner core and the tumor outer edge (infiltrating) (Randall et al., 2020). These areas have unique conditions that support the growth of GBM and are connected to pro-malignant processes such enhanced invasiveness and phenotypic alterations. For example, it was shown that the edge areas display the upregulation of genes that are associated with fatty acid metabolism, and an elevated concentration of acylcarnitine which has pro-inflammatory properties (Garcia et al., 2021; Rutkowsky et al., 2014). Furthermore, the edge area exhibits increased glucose metabolism, while core area has a higher degree of metabolic flexibility and dependency on the metabolism of amino acids (M. E. Baxter et al., 2023; Randall et al., 2020). Although it is clear that development of GBM can be influenced by metabolic alterations, little is known about the specific metabolites and metabolic pathways that underpin these alterations. One potential benefit of metabolomics is its ability to identify functional changes in cells that are complementary to tumor's transcriptional and proteomic patterns (Griffin & Kauppinen,

2007). The most valuable metabolites to distinguish between different kinds of tumor cells include glutamate (Glu), alanine (Ala), tau, and total creatine, as well as metabolites containing choline, and these typical brain metabolites were detected in low amounts in glioma cells (Florian et al., 1995).

Because of the molecular complexity of GBM, most patients experience recurrence even with aggressive therapy regimens. As a result, several possible biomarkers for diagnosis, prognosis, and prediction have been explored. One of them is the mutations observed in the IDH genes. Since the mutations in isocitrate dehydrogenase (IDH-1) and isocitrate dehydrogenase 2 (IDH-2) are found in glioma patients, they have been identified as molecular signature for glioblastoma and have been proposed as drivers of early-stage gliomas, including oligodendroglioma and pilocytic astrocytoma, (Han et al., 2020; Nobusawa et al., 2009). Another well-established biomarker of glioblastoma is the high expression of O-6-methylguanine-DNA methyltransferase (MGMT), a protein that acts as a DNA repair enzyme in the event of DNA damage (Kitange et al., 2009). Promoter methylation in MGMT is frequently identified in secondary GBMs. The abnormalities linked to disrupted methylation of the MGMT promoter led to the transcriptionally silencing of the MGMT gene. Consequently, there is a significant reduction in MGMT expression, impairing DNA repair mechanisms (Szopa et al., 2017). Furthermore, mutations or amplifications of Receptor tyrosine kinases (RTKs) such as basic fibroblast growth factor receptor 1 (FGFR-1), Platelet-Derived Growth Factor Receptor Alpha (PDGFRA), Epidermal Growth Factor Receptor (EGFR), and insulin-like growth factor receptor (IGFR-1) are commonly found in over 80% of primary Glioblastoma Multiforme (GBM) (Szopa et al., 2017). These genetic alterations play a crucial role in the pathogenesis of GBM by driving uncontrolled cell growth and proliferation. For instance,

EGFR mutations can lead to constitutive activation of cell signaling pathways, promoting tumor progression (Wee & Wang, 2017). Similarly, aberrations in PDGFRA can contribute to enhanced cell survival and invasion capabilities in GBM cells (Verhaak et al., 2010). Overall, the high prevalence of RTK mutations underscores their significance as potential therapeutic targets in the primary GBM treatments.

1.4.2 Oncogenic Signaling

The term "oncogenic signaling" describes the aberrant signaling pathway activation in GBM which supports tumor growth and survival (Nakada et al., 2011). These signaling pathways are essential to the initial appearance and progress of the disease. One of the primary oncogenic signaling pathways in GBM is the phosphoinositide 3-kinase (PI3K) signaling pathway (Belda-Iniesta et al., 2008; Gallia et al., 2006). After the activation of the epidermal growth factor receptor (EGFR), PI3K moves to the cellular membrane where it is phosphorylated to 3-phosphate from phosphatidylinositol-4,5-biphosphate (Marie & Shinjo, 2011). This phosphorylation activates target molecules like AKT and mTOR. This PI3K/AKT/mTOR pathway, which prevents apoptosis and promotes cell survival, is typically dysregulated in GBM (Brennan et al., 2013). In fact, the majority of GBM patients have active PI3K/Akt/mTOR signaling (Network, 2013; Touat et al., 2017). This crucial pathway regulates several downstream effectors involved in a wide range of pro-tumorigenic activities, including motility, protein translation, cellular survival, and proliferation. Amplification or mutation of EGFR and loss of function of phosphatase and tensin homolog (PTEN) are two of the hallmarks of GBM pathogenesis that impact the PI3K pathway (Rodon et al., 2013). Interestingly, mutations activated by either mTOR activating or Akt have not been found in GBM, which indicates that altered

functions of these kinases are connected to loss of function in PTEN and/or Receptor tyrosine kinases (RTKs) signaling abnormalities (Sunayama et al., 2010). PTEN functions as a tumor suppressor gene by encoding a protein phosphatase, which inhibits intracellular PIP3 levels and, in turn, the PI3K/Akt pathway. Thus, Akt activity is increased by the loss of PTEN function, which in turn causes mTOR activity to increase, promoting cell survival and proliferation (Garami et al., 2003; Inoki et al., 2002; Vivanco & Sawyers, 2002). PI3Ks are activated by RTKs either directly or by interacting with insulin receptor substrate proteins (IRS) (Kuger et al., 2013). This leads to PIP3 production, activating PDK1, PDK2, and Akt (Chalhoub et al., 2009; Engelman et al., 2006). EGFR is frequently activated in GBM, and its impaired internalization and recycling contribute to PI3K hyperactivation (Dai et al., 2018). Therefore, in glioma cells, this pathway is permanently activated and is a driver of cancer development.

1.4.3 Neurotrophin Receptor Splicing in GBM

Despite decades of research on oncogenic signaling, recent discoveries emphasize the growing significance of aberrant splicing in cancer (Dvinge et al., 2016; Oltean & Bates, 2014; Singh & Eyras, 2017). Tumor suppressors, oncogenes, DNA methylation, and environmental variables have been identified as important drivers of oncogenic signaling; yet there is still a possibility that these factors might influence the splicing processes at the center of oncogenesis. The splicing mechanism is frequently highjacked in cancer to generate abnormal proteins that play a role in the development and growth of tumors (Bonomi et al., 2013). However, despite their oncogenic characteristics, there is not much known about the clinical significance of specific splice variants and the methods in which they might disrupt cellular processes to encourage cancer.

While the neurotrophic tyrosine receptor kinase 2 (NTRK2) gene is involved in the development and growth of gliomas and its unique fusions have been linked to several glioma subtypes (Gatalica et al., 2019; D. T. W. Jones et al., 2013; López et al., 2019; Ni et al., 2017; G. Wu et al., 2014), not much is known about endogenous NTRK2 splicing in the human brain and its possible significance in the biology of brain tumors. Previous studies have showed that in the absence of growth factors (EGF and FGF), TrkB plays a role in the survival of brain tumor-initiating cells (Lawn et al., 2015), and NTRK2 was found to be crucial for the formation of gliomas in a mouse glioma model using a library that targets kinases (Ni et al., 2017).

Proteins called neurotrophins are released and regulate a variety of neurogenesis-related functions, including neuronal survival, proliferation, and differentiation (Chao, 2003; Reichardt, 2006). Four mammalian neurotrophins, including BDNF, NGF, neurotrophin 3 (NT3), 4 and neurotrophin 4 (NT4), have been shown to bind to two different types of receptors: the Trk (tropomyosin receptor kinase) receptors and the p75 neurotrophin receptor (p75NTR). When they bind to Trk receptors, neurotrophins can promote cell survival, proliferation, and differentiation. NTRK2 gene encodes the neurotrophin receptor TrkB, which is required for the maintenance and differentiation of certain neuron populations (Chao et al., 2006). Obesity, eating problems, several types of cancer, and neurological disorders have all been linked to TrkB dysregulation (Altar et al., 2009; Desmet & Peeper, 2006; Gupta et al., 2013; Yeo et al., 2004).

The TrkB gene is located on chromosome 9 and spans more than 350 kbp (Nakagawara et al., 1995). Out of its 24 exons, first five exhibit complicated splicing patterns and function as alternative transcription start-sites (Stoilov et al., 2002). While the full-length

TrkB receptor (TrkB.FL) is the full-length product of the NTRK2 gene, which influences the fate of neural stem cells in both normal brain tissue and gliomas, several splice variants are also produced. The C-terminally truncated receptors TrkB-T1 and TrkB-T-Shc are also encoded by the human TrkB gene using its exon 16 or exon 19, which exhibit translational stop-codons and alternative polyadenylation signals, respectively, other than the TrkB.FL form (Klein et al., 1990).

TrkB.T1 may function as a dominant-negative inhibitor by preventing ligand-induced phosphorylation, since it lacks a kinase domain (Brodeur et al., 2009). However, every species, including rats and humans, have retained its unique 11-amino acid C-terminal sequence. TrkB.T1 and its variations share the extracellular and transmembrane domains along with the first 12 intracellular amino acids (Fenner, 2012; Luberg et al., 2010). In the study investigating the TrkB.T1 effect in several gliomas, TrkB.FL expression is found similar in all samples, while TrkB.T1 is found as the most abundant isoform in LGG and GBM compared to normal brain tissues (Pattwell et al., 2020). In the other studies, TrkB.T1 has been also demonstrated to increase outgrowth of filopodia (Haapasalo et al., 1999), evoke Ca2+ signaling (Rose et al., 2003), regulate neuronal complexity (Carim-Todd et al., 2009; Hartmann et al., 2004), influence glial cell morphology via Rho GTPases (Ohira et al., 2005), and contribute to signal transduction and proliferation (G. T. Baxter et al., 1997; Ohira et al., 2005, 2006). It has been also discovered that by stimulating RhoA activation, TrkB.T1 can cause pancreatic cancer cells to metastasize to the liver (Z. Li et al., 2009). These findings suggest that TrkB.T1 may play distinct and significant roles in the development of the brain, both normally and abnormally.

1.5 Breast Cancer and Steroid Hormones

In 95% of the world's countries, breast cancer, one of the most common malignancies worldwide, affecting over 2.3 million women annually, is the leading or secondary cause of death for female cancer patients. (World Health Organization, 2023). Globally, 685,000 women died of breast cancer in 2020 (WHO, 2021), and it was predicted that 43,170 women and 530 men will die of the disease in 2023 (Siegel et al., 2023). While only a small subset of breast cancer cases (5-10%) are considered hereditary, there are several risk factors that may make the disease more likely to occur, including age, family history, genetic abnormalities (including BRCA1 and BRCA2), hormonal impacts, and lifestyle decisions (Apostolou & Fostira, 2013; Centers for Disease Control and Prevention, n.d.). Because of the distinctive characteristics and development patterns, breast cancers are categorized at the molecular level. The molecular categorization of breast cancer implies the type of cancer cells that are present as well as the pattern and behavior that these cells exhibit. There are four main molecular subtypes, which are essentially determined by the participation of human epidermal growth factor receptor 2 (HER2), and hormone receptors (HRs) which are estrogen receptor (ER) and progesterone receptor (PR) (Cheang et al., 2015):

1. Luminal A (HR+/HER2-): grows slower and less aggressive than other subtypes. Example cell lines are MCF-7 and T47D (Alexandrova et al., 2019).
2. Luminal B (HR+/HER2+): higher grade than luminal A and results in poorer outcomes. Example cell lines are MDA-MB-361 and BT474.
3. Triple-negative (HR-/HER2-): usually invasive and begins in the breast ducts. Example cell lines are MDA-MB-231 and BT20.
4. HER2-positive: is generally ER- and PR-negative and more aggressive than other types. Example cell lines are HCC1954 and AU565.

According to the Annual Report to the Nation on the Status of Cancer, the distribution of female breast cancer subtypes in the United States reveals that about 73% of breast cancer diagnoses are of the luminal A subtype, followed by luminal B at 10%, triple negative at 12 percent, and HER2-positive at around 5% (Kohler et al., 2015).

Hormone replacement treatment (HRT) and the usage of oral contraceptives have been linked to an increased risk of breast cancer; however, this is still an up for debate. Many studies have reported that women who use combined oral contraceptives or HRT have a higher chance of developing breast cancer and that the risk increases with the length of time HRT is used (Chlebowski et al., 2009; Collaborative Group on Hormonal Factors in Breast Cancer, 1996, 1997; R. K. Ross, 2000). On the other hand, there are some studies suggesting that oral contraceptive use does not raise the risk of getting breast cancer (Marchbanks et al., 2002) and reduces the risk of developing other cancers such as ovarian and endometrial (Michels et al., 2018).

There is also evidence from some studies that the use of estrogen and progestin combination, rather than estrogen alone, increases the risk (Beral et al., 2011; Beral & Million Women Study Collaborators, 2003). However, these studies used older types of progestins, and the use of Drospirenone (Drsp) which is a fourth-generation drug may have better results because it more closely mimics natural progesterone (Bird et al., 2011).

1.5.1 Steroid Hormone and Nuclear Receptor Signaling

Based on the kind of nuclear receptor they bind to, steroid hormones are classified into five kinds: estrogens, progestogens, mineralocorticoids, androgens, and glucocorticoids. In the present study, I mainly focus on the first three and their receptor signaling and crosstalk.

1.5.1.1 Steroid Hormones

Steroid hormones are essential in several physiological processes within the human body. These hormones, which include progesterone, aldosterone, and estrogen, are produced from cholesterol (Holst et al., 2004).

1.5.1.1.1 Estrogen

Estrogen is a steroid hormone that is involved in the development of female sexual characteristics and is linked to the female reproductive system. Estrogen can be called as estrone, estradiol, and estriol (Fuentes & Silveyra, 2019). Estradiol is the most widely used form of estrogen in hormone replacement therapy (HRT) for the treatment of menopausal symptoms. Females generate all forms of estrogens throughout their lives; however, the hormones, estriol and estretrol, are primarily present during pregnancy, while estrone is often present in larger concentrations following menopause (Samavat & Kurzer, 2015).

While it also has significant effects on males, estrogen is a crucial sex hormone mainly related to the growth and function of the female reproductive system (Hamilton et al.,

2017). It is a lipid-soluble steroid hormone that is mostly produced by the ovaries, testes, and adrenal cortex and has several essential physiological functions. Estradiol is the major circulating estrogen in humans and is mostly released by the corpora lutea and the granulosa cells of the ovarian follicles (Chauvin et al., 2022). During pregnancy, the placenta also produce estrogen to aid in preparing the mammary gland to make milk (Voogt, 1978). Based on where they come from, endogenous and exogenous are the two primary categories into which estrogens are divided. Endogenous estrogens are released by cells or glands within organisms (Flasch et al., 2022). On the other hand, synthetic estrogens, medications, and nutrition are sources of exogenous estrogens (Ohta et al., 2015).

Among the naturally occurring estrogen hormones estrone (E1), 17β-estradiol (E2), estriol (E3), and estetrol (E4) (R. S. Kumar & Goyal, 2021), the most powerful and common form is estradiol (E2). All four estrogens have similar chemical roles and structures and are made up of 18 carbon atoms (P. Chen et al., 2022). Because E2 is widely distributed and has physiological significance in many tissues and organ systems, the word "estrogen" is frequently used to refer to it (Fuentes & Silveyra, 2019). E2 has proliferative effects and regulates the development of secondary female sex characteristics, the menstrual cycle, and the production of the uterine lining (R. S. Kumar & Goyal, 2021). On the other hand, the main estrogen that the placenta secretes while a woman is pregnant is called E3. Because of this, E3 levels in males and non-pregnant women are negligible in comparison to their enormous rise during pregnancy, while E4 is a naturally occurring fetal estrogen that is just apparent during pregnancy (Gallez et al., 2021; R. S. Kumar & Goyal, 2021).

Estrogens are produced from cholesterol, which is common precursor of all steroid hormones (Fuentes & Silveyra, 2019). In short, it is made from androstenedione when oxygen and NADPH are present. The key enzyme responsible for this activity is aromatase (CYP19A1), which involves in the last step of the synthesis of E1 and E2. Adipose tissue, the adrenal cortex, the gonads (primarily the granulosa cells of the ovaries), and other tissues; the breast, placenta (W. L. Miller, 2017), and, to a lesser degree, the fetal liver in the case of E4 (Gallez et al., 2021), all undergo estrogen production.

Estrogens primarily stimulate the development and proliferation of specific cells inside the body, which are responsible for of developing the majority of a woman's secondary sexual traits (Fuentes & Silveyra, 2019). Estrogens are released into the circulation after their production. Then, they are delivered to the target organs after binding to a plasma protein (albumin) to produce their effects. The primary function of estrogens is to cause development and maturation of sex organs and other reproductive tissues. Estrogens trigger the growth of the breasts and the machinery that produces milk (Bermejo-Haro et al., 2023). They also promote bone development by preventing excessive bone demineralization (J. Liang & Shang, 2013), and the decrease in blood estrogen levels decrease after menopause results in osteoporosis (Tella & Gallagher, 2014).

1.5.1.1.2 Progesterone

Another essential steroid hormone for reproductive health is progesterone (P4) hormone which consists of 21 carbons and is mostly produced in testicle, adrenal glands and in the ovaries (Shchelkunova & Morozov, 2015). One of the important roles of progesterone is the maintenance of pregnancies by decreasing myometrial contractions needed for

implantation, preparing the endometrium, encourage gestation and prevent lactation while pregnant. It also affects the production of inflammatory mediators in the uterine cavity (Arab et al., 2019). Hence, the loss of progesterone causes decreased ability to fight immunologic threats and increased myometrial contractility, which leads to a higher risk of miscarriage and preterm delivery of the fetus. Progesterone also affects the menstrual cycle by promoting capillary formation and expansion, which elevates blood flow and vascularization (Stanczyk, 2007). In addition, progesterone increases the mucous lining of the cervix, creating an inhibitory barrier that prevents sperm from penetrating (Ola et al., 2003). Therefore, due to the characteristics of increased mucus secretion of progesterone, progestin-dependent contraceptives have a therapeutically meaningful use.

In addition, progesterone is crucial for a number of organs other than the reproductive system, including the central neurological system, the cardiovascular system, the mammary gland, and the bones (Taraborrelli, 2015). Research has also been done on progesterone's ability to prevent osteoporosis, benign prostatic hyperplasia, a number of neurological conditions and traumatic brain injury (Kaore et al., 2012).

Progesterone is also derived from cholesterol, which is cleaved by the mitochondrial cholesterol side-chain cleavage enzyme (cytochrome P450scc) into the P4 precursor, pregnenolone (Barros et al., 2015). 3β-hydroxysteroid dehydrogenase (3-β-HSD) subsequently catalyzes the conversion of pregnenolone to P4. P4 can interact with progesterone receptors (PR), a nuclear transcription factor that controls transcription by acting through certain P4 response elements located in the target gene's promoter region (Sitruk-Ware et al., 2021).

The word "progestagen" refers to steroid hormones, including P4, that have progestational actions and the name “progestin” is used to define synthetic progestogens (Stanczyk, 2002). There is only one natural progestogens, and it is defined as “progesterone”. There are two approaches to subclassify progestins: (1) generationally, or (2) according to their structural characteristics (M. Edwards & Can, 2024). Accordingly, the first, second, third, and fourth-generation agents make up the generational subclassification, which is based on the date the progestin group was released onto the market, independent of the derivation of its structure. The most recent preparations of fourth generation progestins have some antiandrogenic characteristics. Moreover, categorization based on structural features includes 3 groups: pregnanes, estranes, and gonanes. Pregnanes which are derivatives of progesterone, include compounds such as nomegestrol acetate and medroxyprogesterone acetate. In contrast, other two groups are derived from testosterone and have androgenic activity (Mishell, 2012). Estranges comprise norethynodrel, ethynodiol diacetate, norethindrone acetate and norethindrone, while gonanes includes gestodene, norgestimate, desogestrel and levonorgestrel.

Synthetic progestins, which are extensively used and sold under the brands such as Provera, PremPro, and Cycrin, have a variety of serious adverse effects that can occur, including dysphoria, fluid retention, enhanced androgenicity, changes in lipid levels, fatigue and hypercoagulant states (Goletiani et al., 2007). The side effects of natural progesterones are considered less severe, depending on way it is administered.

The fourth-generation progestin, drospirenone (Drsp), is particularly significant. It is derived from Spironolactone, which is an antagonist of the mineralocorticoid receptor, MR, and is also used in the treatment of hypertension and heart failure (Bird et al., 2011;

Patibandla et al., 2023). Drospirenone is a progestin that is currently being used in both combined oral contraceptives and menopausal hormonal therapy (HRT) (Rapkin & Winer, 2007). The discovery of Drsp has been motivated by the need to maximize the anti-aldosterone activity of spironolactone while limiting its endocrinological impacts. Drsp and spironolactone are structurally similar but have a difference in that Drsp does not have C7α acetylthio substitution which is present in spironolactone (Figure 3). Instead, two methylene groups are substituted in Drsp at the positions C6β-7β and C15β-16β (Paton, D.M., 2020).

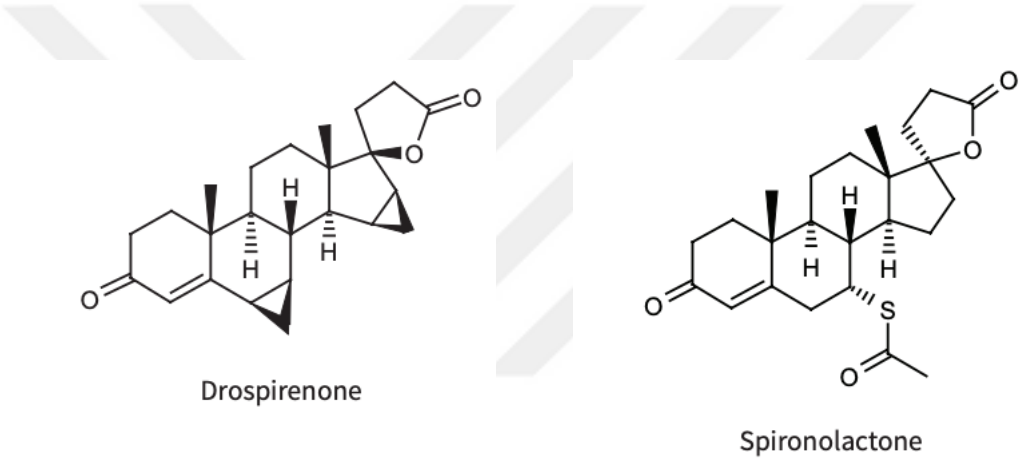


Figure 3. Structure of drospirenone and spironolactone.

(This image was reproducible without the need of copyright permission and required citation and it was obtained from Paton D. M. (2020). Drospirenone: a novel progestogen used as an oral contraceptive without an estrogen component. *Drugs of today (Barcelona, Spain : 1998),* 56(5), 321–328. https://doi.org/10.1358/dot.2020.56.5.3134788)

Similar to spironolactone, Drsp exhibits both an anti-mineralocorticoid and anti-androgenic activity (little or no androgenic activity), but the former activity is more potent (about 8 times higher than that of spironolactone), while Drsp has a five- to ten-times greater anti-androgenic activity than that of progesterone (Muhn et al., 1995b). The relative binding affinity of Drsp to the progesterone receptor is 19% when the affinity of

progesterone to its receptor is taken as 100%. While Drsp also shows a higher affinity to the mineralocorticoid receptor (MR), it exhibits a low binding affinity to the glucocorticoid receptor (GR), little or no binding to the androgen receptor and no binding to the estrogen receptor (Elstein & Furniss, 1995; Fotherby & Caldwell, 1994; Krattenmacher, 2000).

Compared to previous progestin generations which have more binding affinities to both the progesterone and androgen receptors, Drsp is more similar to natural progesterone (Louw-du Toit et al., 2017). The hormonal activities of some of the selected progestogens are shown in Table 2. Three distinct activities serve as a general classification of progestin potency; these actions are what enable progestins to be effective under various circumstances. Progestational activity is defined as the capacity to continue a pregnancy and transform the endometrium into the secretory phase (Sitruk-Ware, 2004). In contrast, anti-estrogenic activity is characterized as the capacity to suppress the ER and, as a result, reduce the endometrium's thickness when primed with estrogen, while anti-androgenic activity is defined as the capacity to prevent testosterone from attaching to AR, hence reducing the androgen impact and inhibiting 5α-reductase.

Table 2. Hormonal effects of certain progestogens.

Progestogen	Estrogenic	Anti-estrogenic	Androgenic	Anti-androgenic	Glucocorticoid	Anti-mineralocorticoid
Progesterone	−	+	−	(+)	+	+
Cyproterone acetate	−	+	−	+	+	−
Dydrogesterone	−	+	−	−	?	(+)
Norethisterone	+	+	+	−	−	−
Levonorgestrel	−	+	+	−	−	−
Medroxyprogesterone acetate	−	+	(+)	−	+	−
Medrogestone	−	+	−	−	?	−
Gestodene	−	+	+	−	(+)	+
Drospirenone	−	+	−	+	(+)	+

+, effective; (+) weakly effective; −, ineffective; ?, unknown.

Data were obtained from different sources (Elstein & Furniss, 1995; Fotherby & Caldwell, 1994; Krattenmacher, 2000; Muhn et al., 1995a; Paton, D.M., 2020; Wiegratz & Kuhl, 2004).

1.5.1.1.3 Aldosterone

Aldosterone is an endogenous mineralocorticoid hormone produced from cholesterol within the zona glomerulosa of the adrenal cortex and controls sodium homeostasis due to its mineralocorticoid activity (Scott et al., 2023). Its main function is to maintain water and electrolyte balance by regulating transcription of epithelial sodium channel (SCNN1/ENAC) and Na+/K+-ATPase subunit genes, promoting retention of sodium and water and excretion of potassium by the kidneys (Rossier et al., 2015). Therefore, it plays crucial roles in controlling of blood pressure and volume by increasing sodium reabsorption by the kidneys (Booth et al., 2002). In the target cells, aldosterone works by binding to and stimulating MR. Then, the transcription of a gene is triggered by the

aldosterone-MR complex binding to certain hormone response elements in the target gene region (Katsu & Baker, 2021).

Low blood pressure causes the release of aldosterone; when the angiotensin II or potassium levels in the blood rises, secretion of aldosterone increases; whereas sodium ion concentration in the blood rises, aldosterone secretion decreases (Scott et al., 2023). Due to its hydrophobicity, aldosterone permeates into the cortical collecting tubules and late distal tubules' main epithelial cells (Thomas & Harvey, 2019). After then, aldosterone and the MR interact in the cell's cytoplasm forming an aldosterone receptor complex, which travels into the cell nucleus to regulate the transcription of the targets they bind. This process enhanced secretion of hydrogen and potassium in addition to sodium reabsorption (Rossier et al., 2015). Specifically, it increases the growth of sodium-potassium pumps on the principal cells' basolateral membrane. Additionally, it causes an increase in the expression of epithelial sodium channels (ENaC), on the principal cells' apical membrane (Nesterov et al., 2012). Sodium is moved from the filtrate into the renal capillaries by the activity of ENaC and sodium potassium pumps, which reduces the amount of sodium lost in the urine. Since sodium attracts water, aldosterone ultimately raises serum sodium levels and increases the amount of fluid in the vasculature, which raises blood pressure (Grillo et al., 2019).

Glucocorticoids have been observed to inhibit the proliferation of numerous cells (Koyama et al., 2001). The levels of intracellular glucocorticoids are regulated by two isoforms of 11β-hydroxysteroid dehydrogenase (11βHSD). Glucocorticoids are deactivated by the type 2 isoform (11βHSD2), often found in mineralocorticoid target tissues, preventing the non-selective MR from being occupied by high amounts of

glucocorticoids (Funder et al., 1988). 11βHSD2 is responsible for the binding of aldosterone to MR (Rabbitt et al., 2003). In both normal and malignant breast tissues, higher levels of 11βHSD2 have been seen in association with the MR (Sasano et al., 1997). The studies evaluating the activity of 11βHSD2 suggested that it plays a crucial regulatory function in modulating the effects of aldosterone in breast cancers (C. H. Kim & Cho, 2010, p. 2). In a study investigating the 11βHSD2 in breast cancer cells, cell proliferation and 11βHSD2 activity were not affected by aldosterone treatment (10^{-8} M), while GR agonists stimulated 11βHSD2 activity (Koyama et al., 2001).

Aldosterone has been also shown to induce rapid effects by triggering the EGFR/ERK signaling pathway in breast cancer cells, depending on the participation of another receptor, GPER, and MR (Rigiracciolo et al., 2016). The G-protein estrogen receptor (GPER) activation is known to initiate various transduction pathways that play a role in the growth, infiltration, and movement of cancer cells, specifically involved in promoting estrogenic signaling effects in breast cancer (Filardo et al., 2000; Lappano et al., 2014; Prossnitz & Maggiolini, 2009). GPER has been shown to contribute to aldosterone-triggered responses, including the EGFR/ERK transduction pathway in a MR- and GPER-dependent manner in breast cancer cells, indicating that aldosterone induces direct interactions between MR and GPER, which involves in breast cell proliferation and migration (Rigiracciolo et al., 2016).

The steroid derivative spironolactone, which is used for the treatment of primary hyperaldosteronism, is the competitive aldosterone antagonist. It works by binding to the MR in the distal tubule of the nephron and the collecting ducts to block the effects and binding of aldosterone (Rankin, 2008). This process affects the normal signaling

pathways associated with aldosterone and diminishes its effects on sodium reabsorption and potassium excretion. Therefore, in the patients, spironolactone aids in the recovery of electrolyte balance and the regulation of blood pressure.

1.5.1.1.4 Combination of Drsp with E2 as Treatment

In the 1990s, studies started to investigate Drsp in many fields such as contraception and HRT (Muhn et al., 1995b; Oelkers et al., 1991; Pollow et al., 1992), due to the recognition of its pharmacological profile closely resembling that of endogenous progesterone, which leads to expectations of many additional health advantages for individuals using oral contraceptives or undergoing hormone replacement therapy (Muhn et al., 1995b). Therefore, in May of 2001, Yasmin® (Berlex, Inc.), a medication containing drospirenone/ethinyl estradiol, was first approved, and then a similar drug, Yaz® (Berlex, Inc.), was approved in March of 2006 (FDA label, Bayer Healthcare Pharmaceuticals Inc., 2006; FDA label, Exceltis USA Inc., 2019; Food and Drug Administration, 2006; Kennewell, 2007; Krattenmacher, 2000). Currently, they both are used as oral contraceptives and contain 3 mg drospirenone but 30 μg ethinyl estradiol combination in Yasmin and 20 μg ethinyl estradiol combination in YAZ. In 2005, Angeliq (Schering AG) containing 1 mg 17β-estradiol/2 mg drospirenone in 1:2 ratio was approved as HRT preparation for menopausal symptom management. Moreover, in May 2019, Slynd (Exeltis USA, Inc.), an oral contraceptive pill containing only 4 mg of Drsp, received FDA approval for pregnancy prevention (FDA label, Exceltis USA Inc., 2019). These widely used drugs make studying Drsp alone or in combination with estrogens very important for women health across all ages.

In an open trial that evaluated the use of Yasmin, researchers assessed several menstrual-related symptoms at different stages of the menstrual cycle, which included water retention, difficulty in concentrating, feeling unwell, unwanted hair changes and appetite increase (Parsey & Pong, 2000). Significant decreases in water retention and severity of increased appetite were observed while there were no significant changes in others. Overall, drospirenone was shown to have a positive effect on the above-mentioned common symptoms of the menstrual cycle.

Studies also has shown that long-term usage of Drsp:E2 combination provides a noticeable antiproliferative impact on the endometrium (Klipping et al., 2006) yet the molecular mechanisms underlying these changes are unclear. These combined oral contraceptives work by suppressing the hypothalamic–pituitary–ovarian axis which prevents the formation of follicles and the process of ovulation. Estrogen works in concert with Drsp by lowering the follicle-stimulating hormone (FSH) and luteinizing hormone (LH), levels which leads to suppression of hypothalamic–pituitary axis (Shaw et al., 2010). Estrogen also keeps the endometrium healthy and aids in reducing withdrawal bleeding (L. Miller & Hughes, 2003). Moreover, by inhibiting LH, Drsp also prevents follicular activity stimulation and ovulation. In the study in which the effect of different Drsp doses (0.5, 1, 2, and 3 mg) alone on the ovulation was investigated, the lowest incidence of ovarian maturation and the highest incidence of ovarian suppression was observed in the highest dose of Drsp (Rosenbaum et al., 2000). It has been discovered that drospirenone dosages of 1, 2, and 3 mg combined with 1 mg estradiol are safe and effective for treating climacteric symptoms in healthy women after menopause (Schürmann et al., 2004). In another analysis including women having polycystic ovary syndrome (PCOS), Drsp significantly reduced hirsutism and acne after 6 months of

continued therapy (Podfigurna et al., 2020). Additionally, LH, FSH, prolactin, and testosterone levels were statistically significantly decreased as a result of the medication.

While progesterones are used in drugs to stop ovulation, lessen uterine hyperplasia, and reduce the risk of gynecological cancer, both synthetic progestins and natural progesterones can negatively affect cognition and may even counteract some of the neuroprotective effects of estrogens. On the other hand, the cognitive effects of Drsp have been looked into in rats that have had their ovaries removed (Koebele et al., 2022). The effect of different doses of Drsp and the combination of Drsp with ethinyl estradiol therapy were compared and it was found that a moderate dose of Drsp increased spatial memory while the addition of ethinyl estradiol reversed the positive effects of drospirenone on working memory (Koebele et al., 2022). This might suggest that these two agents crosstalk.

1.5.1.2 Nuclear Receptor Signaling

Steroid hormones bind to a family of proteins called nuclear receptors (NRs), which are present in cells. NRs frequently undergo shape and form alterations upon binding with a ligand, which enables them to directly engage with cell DNA and regulate gene expression and include, estrogen receptor, progesterone receptor, mineralocorticoid receptor, androgen receptor and glucocorticoid receptor (Mangelsdorf et al., 1995; Weikum et al., 2018). The signaling mechanisms of steroid hormones and their respective NRs for the first three were summarized below in relation to breast cancer and if any relevance also to glioblastoma.

1.5.1.2.1 Estrogen Signaling and Its Relationship with Breast Cancer

Estrogen signaling is a complex process with impacts on various physiological and pathological situations. Traditionally, estrogen was thought to initiate signaling by binding itself to nuclear estrogen receptors, which then interact with DNA and function as transcription factors triggered by ligands. However, it is now known that estrogen signaling triggers a variety of cellular reactions through both genomic and non-genomic mechanisms (Revankar et al., 2005). Numerous biological processes are influenced by three different types of estrogen receptors (ERs): classical alpha (ERα) and beta (ERβ), and non-classical G protein-coupled estrogen receptor 1 (GPER1) (P. Chen et al., 2022). ESR1 on chromosome 6 (6q25.1) (Gosden et al., 1986) and ESR2 on chromosome 14 (14q23.2) (Enmark et al., 1997) encode ERα and ERβ, which are made up of 595 and 530 amino acids, respectively. The relative molecular weights are 66 and 59 kDa, respectively, for full-length ERα and ERβ. On the other hand, chromosome 7 encodes GPER1, which contain 375 amino acids and has a molecular weight of 41 kDa (Filardo & Thomas, 2012). These three predominant ERs are expressed in a wide variety of tissues. Whereas ERβ is found in the central nervous system, cardiovascular system, kidney, lung, colon, immune system and male reproductive system, ERα is mostly produced in reproductive tissues, breast, bone, liver, kidney and adipose tissue (Jia et al., 2015). Indeed, ERβ agonism has been an important means for treatment of glioblastoma in recent years (Pratap et al., 2023; Zhou et al., 2019). It was shown that ERβ agonists displayed growth inhibitory activities in glioblastoma by promoting apoptosis (Pratap et al., 2023).

Different immune cells, neurons, vascular endothelium, skeletal muscle and target effector organs are among the tissues where GPER1 is more extensively dispersed and

expressed (Olde & Leeb-Lundberg, 2009). Furthermore, it has been found that tissues from lung, ovarian, and breast cancers express GPER1 (Mizukami, 2010). Estrogens exert their affects in two way; genomic and non-genomic, by binding to the classic ERα and ERβ as well as on the non-nuclear receptor GPER1 (Hewitt et al., 2016).

In the regular genomic pathway, intracellular estrogen receptors (ERα and ERβ) bind to estrogen (for example, E2) and then use transcriptional mechanisms to control gene expression (Figure 4). After this attachment, the change in conformation occurs resulting in the E2-receptor dimerization complex, which is translocated to the nucleus where the it attaches to consensus estrogen response elements (EREs) sites on the DNA, enhancer regions inside or near promoters and 3′-untranslated regions of target genes, serving as transcription-regulating factors for genes (Fuentes & Silveyra, 2019; Klinge, 2001; Murphy, 2011). Then, the gene expression either increases or decreases as a result of the recruitment of co-regulators (Murphy, 2011). Estrogen-mediated gene products normally govern autophagy, proliferation, apoptosis, survival, differentiation, and vasodilation. However, in pathological situations, they could not perform as well. This "genomic effect" is slow due to the intracellular localization of ERα and ERβ, which means that activation usually lasts hours or longer (P. Chen et al., 2022). E2 also mediates the non-genomic signaling pathway by binding to membrane bound ERα, ERβ, and GPER1. By regulating ion channel opening or stimulating associated enzymes such as non-receptor tyrosine kinases (Src), PI3K, and mitogen-activated protein kinase (MAPK), this route instantly activates nuclear transcription factors (Mah et al., 2011; Marquez-Garban et al., 2011). This mechanism is known as a rapid "non-genomic effect" since it depends on the activation of different signaling cascades, not only gene regulation, and happens instantly (P. Chen et al., 2022).

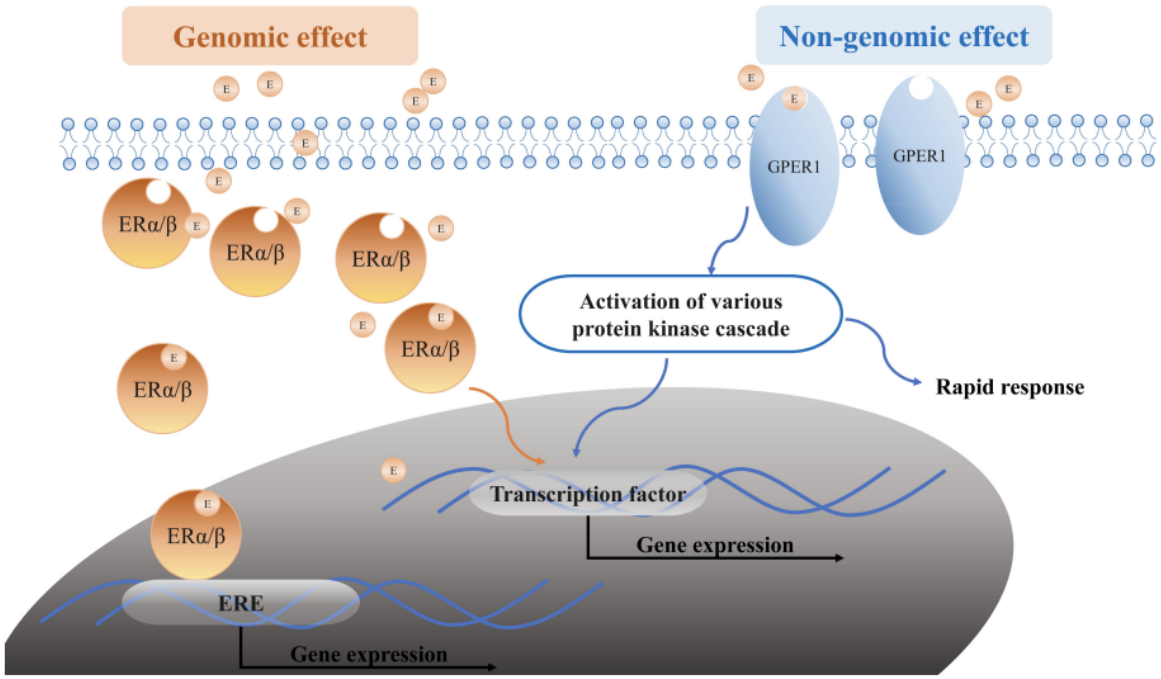


Figure 4. Mechanisms for estrogen signaling. In the graph, estrogen, or E2 (orange circle E), binds to ERα/ERβ and GPER1 to exert its genomic and non-genomic effects. Orange highlights the genomic impact, which is caused by the E2-receptor complex binding to EREs as it enters the nucleus. The non-genomic impact is displayed in blue. E2 binds to ERs in the membrane-like GPER1 and controls expression by adjusting the activation of associated enzymes or the opening of ion channels. E; estrogen or E2, ERα; estrogen receptor alpha, ERβ; estrogen receptor beta, GPER1; G protein-coupled estrogen receptor 1, ERE; estrogen response elements.

(This image was reproducible without the need of copy right permission and required citation and it was obtained from Chen, P., Li, B., & Ou-Yang, L. (2022). Role of estrogen receptors in health and disease. *Frontiers in endocrinology*, 13, 839005. https://doi.org/10.3389/fendo.2022.839005).

ERs have many roles in the growth of breast cancer tumors as well as in the survival and proliferation of cells. Anti-estrogen therapy is the "gold standard" for treating luminal breast cancer, which expresses ER (Szostakowska et al., 2019). ERs influence the development and course of breast cancer by starting a unique signaling cascade via their attachment to estrogens. ERα is overexpressed in breast cancer; it is five to eight times more abundant in the tissues of breast cancer than in healthy tissues (Huang et al., 2014). Furthermore, ERα promotes the development of tumors in breast cancer by interacting with estrogen. Accordingly, dysregulation of ERα expression, activity, or its coregulators and target genes accounts for the majority of breast malignancies, referred to as ERα+ or luminal (Clusan et al., 2023). ERα acts as a transcription factor that control several factors linked to the formation and proliferation of tumor cells, including IGF1R, vascular endothelial growth factor (VEGF), anti-apoptotic BCL-2 protein and cyclin D1 (P. Chen et al., 2022). Several post-translational modifications (PTMs) can happen on ERα, which affect its activity and aid in the initiation and progress of breast cancer. Depending on the location of change, phosphorylation can either stimulate or inhibit ER function (Yuan et al., 2106). In breast cancer cells, the phosphorylation of ERα serine 118 and serine 167 results in distinct gene expression patterns that eventually affect the development and morphology of tumors as well as the response of patients with breast cancer to hormone treatment (Anbalagan & Rowan, 2015). According to several studies, ERα activity in breast cancer cells via the interactions with coactivators is increased due to the phosphorylation of several serine residues and tyrosine (Arnold et al., 1995; Duplessis et al., 2011; Kato, 2001; Tharakan et al., 2008; Tremblay et al., 1999). Acetylation can also occur by histone acetyltransferases (HATs) and remarkably, ER transcriptional activity is increased by the acetylation of lysine 266/268 due to the stimulated binding of receptors

to the DNA, whereas ERα activity is inhibited by the acetylation of lysine 302/303 (M. Y. Kim et al., 2006).

Methylation is another genetic alteration that controls mainly silencing the transcription of ER genes and is associated with hormone treatment resistance (Tsuboi et al., 2017). Methylation is seen in gene promoters linked to certain biochemical processes in a number of malignancies, including breast cancer (Brooks et al., 2009). Research has shown that the promoters of ER/PR-negative tumors exhibit different levels of methylation than those of ER/PR-positive tumors, and DNA methylation indicators may have prognostic relevance for the treatment of breast cancer (Benevolenskaya et al., 2016; Fackler et al., 2011; L. Li et al., 2010).

Opposite to ERα, ERβ activation normally causes apoptosis to be induced and proliferation to be inhibited, however, these effects vary depending on the tissues under study, the cell environment, transcriptional coactivators, and whether or not ERα is co-expressed (Mal et al., 2020). Thus, ERβ expression in breast cancer cells is frequently associated with a favorable prognosis; however, some research suggests otherwise. In fact, previous work suggests that ERβ and its isoforms, together with certain coactivators including, TIF-2, NF-kB and AIB1, often control the growth and proliferation of cells in breast cancer (Choi, 2022; Choi et al., 2022). These particular coactivators have been linked to poor clinical outcomes and have been shown to be correlated with high expression of ERβ in subtypes of high-grade breast tumors, indicating that they could have a role in breast cancer cells' ERβ promotion (Girault et al., 2003).

On the other hand, GPER1, which is an another estrogen receptor although non-nuclear, plays a protective effect against breast cancer, and reduced expression of GPER1 in tumor tissue is associated with a lower chance of survival (Ignatov et al., 2013). GPER1 is present in around 60% of breast cancer tissues, and its expression is validated in triple-negative breast cancer (TNBC) (Treeck et al., 2020). Nonetheless, GPER1 expression is linked to a decreased response or resistance to tamoxifen treatment (Ignatov et al., 2020). In spite of this, GPER1's role in blocking NF-κB/IL-6 signaling, which results in the inhibition of migration and angiogenesis, makes it an important therapeutic target for TNBC (S. Liang et al., 2017). In summary, whereas ERβ's significance in the development of breast cancer is debatable, ERα and GPER1 exhibit promising roles.

1.5.1.2.2 Progesterone Signaling and Its Relationship with Breast Cancer

Progesterone operates through two distinct signaling mechanisms: the classical genomic receptor pathway (nuclear), and the non-classical/non-genomic receptor pathway (extranuclear). Since progesterone has lipophilic nature, it can easily pass through a target cell's membrane by diffusion and interacts with progesterone receptors located within the cytoplasm. There are two progesterone receptor isoforms (PR-A and PR-B), which coordinate distinct activities and found in tissues at different levels (Graham & Clarke, 1997; Kastner et al., 1990; Scarpin et al., 2009) (Kastner et al., 1990; Scarpin et al., 2009). These isoforms result from differential excision of one gene located on chromosome 11 in humans. In the classical mechanism, the transcribed cytoplasmic PRs are attached to a multiprotein chaperone complex and stay inactive (Pedroza et al., 2020). When progesterone attaches to the receptor’s ligand binding site, multiple conformational changes are triggered that lead to dimerization of two-ligand receptor protein complexes

and their relocation into the nucleus, where they bind to progesterone response elements (PREs) or hormone response elements (HREs) (Figure 5) (DeMayo & Lydon, 2020; Scarpin et al., 2009). This result in the recruitment of coregulators (activators or repressors) to the transcriptional complex in order to activate or suppress gene expression (Gellersen et al., 2009). Moreover, a third isoform, known as PR-C, has been identified in some studies (Condon et al., 2006; Wei et al., 1990, 1996). It is overexpressed in myometrial cells and lacks a DNA binding domain and the ability to bind to DNA. However, it can block PRs by binding to progesterone, thereby allowing progesterone to be used for the PR B isoform.

Estrogens are necessary for the biological effects of progestogens (at least in most tissues), as estrogens bound to ER is the key factor that regulate PR levels. (Jacobsen & Horwitz, 2012). Although, it is uncommon for normal PR-positive cells to lack ER expression, variations in PR and ER expression have only been found in cancer cells.

It has also been demonstrated that progesterone exerts its effects through non-genomic processes, such as the triggering the secondary messenger cascades through certain receptors such as progesterone membrane receptors (mPRs) and progesterone receptor membrane component 1 (PGRMC1) (Mueck et al., 2014). Non-classical mechanisms have been shown to rapidly activate progestin-activated signaling. It was shown that progestin R5020 rapidly stimulated c-Src, EGFR and MAPK-dependent phosphorylation on progesterone receptors in T47D cells (Faivre et al., 2008). In addition, via MAPK pathways, P4 may activate c-Src and enhance prolactin-mediated STAT (signal transducer and activator of transcription) activation in order to encourage cell division (Leehy et al., 2018). Another way of signaling of progesterone is through the receptors in

cell surface, which progesterone interacts with membrane-bound receptors and mediates signaling without the diffusing through the membrane. These membrane-bound receptors contain seven transmembrane domains, which have structural similarities with G protein-coupled receptors (GPCRs), are known as membrane progesterone receptors (mPRs), which have high affinities for progesterone. Once the progesterone binds to these receptors, extranuclear progesterone signaling initiates a complex network of intracellular processes, including as the activation of ERK1/2, MAPKs and an influx of intracellular Ca2+.

In the non-genomic receptor mechanism, effects related to progesterone are very fast compared to genomic mechanism, since progesterone can be bound by mPRs with fast rates of association and dissociation (García-Sáenz et al., 2023; Zhu et al., 2003). Progesterone receptor membrane components 1 and 2 (PGRMC1 and PGRMC2) are membrane-associated progesterone receptors (MAPRs) that do not conform to the same structural GPCR format (Cahill, 2007). They are characterized by unique protein type with a single transmembrane domain. It has been demonstrated that PGRMC1 interacts with cytochrome P450 (CYP) enzymes and is important for the synthesis of cholesterol. In a variety of cells lacking the nuclear PR, PGRMC1 can bind to SERBP1, and together they function as quick mediators for progestin activities, illustrating the effects of progesterone on the cell via surface receptors of cell (Bashour & Wray, 2012). In contrast to PGRMC1, less is known about PGRMC2. It has been demonstrated that PGRMC2 is mostly found in the nuclear envelope and endoplasmic reticulum (Gerdes et al., 1998; Parker et al., 2017).

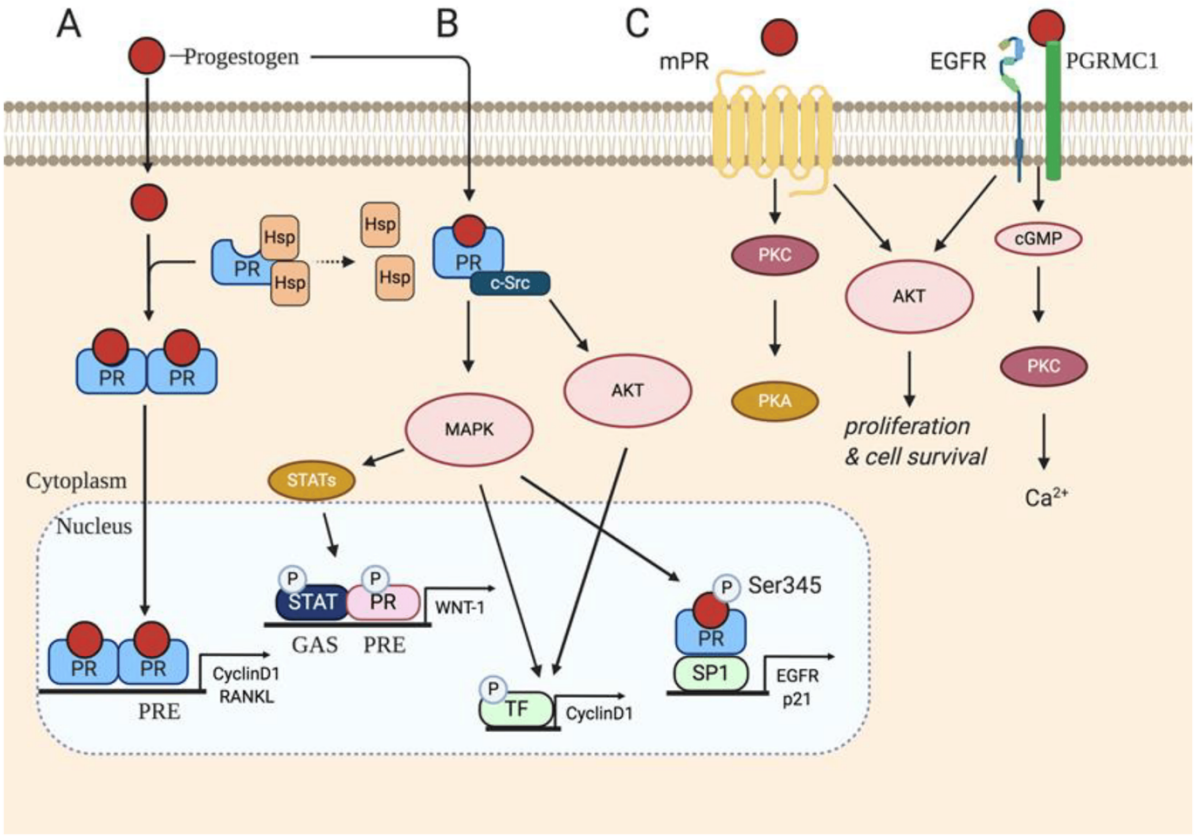


Figure 5. Classical and non-classical signaling of progesterone. A) Classical signalling of progesterone is mediatied by the diffusion of hormone through membrane and binding to receptor inside the cell. B) Rapid response in non-classical mechanism is mediated through the activatiton of c-Src, MAPK and AKT. C) The non-classical signaling can be initated by the binding of progesterone to mPRs and PGRMC1.

(This image was reproducible without the need of copy right permission and required citation and it was obtained from Pedroza, D. A., Subramani, R., & Lakshmanaswamy, R. (2020). Classical and Non-Classical Progesterone Signaling in Breast Cancers. Cancers, 12(9), 2440. https://doi.org/10.3390/cancers12092440).

The role and distributions of these two receptors (PR-A and PR-B) are different in different tissues. While PR-A has a more prominent function when P4 is absent, PR-B demonstrates higher transcriptional regulation when P4 is present (Jacobsen et al., 2005). In contrast to the similar expression of PRA and PRB in normal breast tissues, the aberrant

isoform ratios are found in breast and endometrial tumor cells (Mote et al., 2007). These receptors can also interact with other receptors such as ER, AR, MR and GR, and this interaction can results in agonistic or antagonistic effects (Jacobsen & Horwitz, 2012; Tung et al., 1993).

In the breast tumors, genes involved in the pathways related to cell proliferation and metastasis such as TNFRSF11A encoding the RANKL receptor, are transcribed by PR-A expression (J. A. Khan et al., 2012). Initial phases of the formation of breast tumors have been associated with the RANKL disruption in a progestin-responsive way (Gonzalez-Suarez et al., 2007; Grimm et al., 2016), while breast cancer cells express its receptor, RANK, which is essential for the proliferation, migration and differentiation of these cells (D. H. Jones et al., 2006). The activation of cyclin D1 can also be regulated by the RANKL/RANK signaling axis (González-Suárez, 2011). Progesterone promotes proliferation in ER+/PR+ cells through a cyclin D1-dependent, cell-intrinsic mechanism. It was shown that in ER+/PR+ breast tumors, a poor prognosis is linked to cyclin D1 overexpression and amplification (Brisken, 2013; Roy et al., 2010). Additionally, the interaction of PR with CK2 and FOXO1 facilitates a hormone-responsive reaction to progestins, which leads to interaction of PR with cyclin D1, thereby regulating the cell cycle (Diep et al., 2016; Grimm et al., 2016).

Increased levels of PR-A have been associated with poor tamoxifen treatment responses or resistance (Hopp et al., 2004; Mote et al., 2015). Furthermore, it has been reported that mifepristone, an anti-progestin drug, inhibits the proliferation of cells in tumors where PR-A is dominant but not PR-B (Rojas et al., 2017). Based on this research, the effectiveness of different therapies may be largely predicted by calculating the ratio of

PR-A to PR-B. Furthermore, it has been shown that an increased expression of PR-A in breast cancer cells reduces adhesion of cells and increases th cell migration into bone marrow stroma (Graham et al., 2005; McGowan et al., 2004; McGowan & Clarke, 1999).

Since natural progesterone is believed to have a inhibitory effect on growth and plays a crucial function in normal reproductive physiology, whereas synthetic progestins have been demonstrated to promote growth, the relevance of progestogens in breast tumors is debatable (Asi et al., 2016; NIH State-of-the-Science Panel*, 2005; Santen, 2003). In one HRT study, it was shown that postmenopausal women who received P4 with estrogen had no or a slightly elevated risk of breast cancer (A. Fournier et al., 2005). However, P4 has been shown to increase the risk of breast cancer by other studies (Blank et al., 2008; Shull, 1997).

Additionally, it was shown that progesterone alone increased the cell proliferation in breast cancer in vitro while a PR-antagonist significantly reduces E2-induced mammary development in vivo (Boopalan et al., 2015). These results are also consistent with the Women's Health Initiative (WHI) research, which demonstrated that progestogens are important in this context and that E2 alone does not increase the risk of breast cancer (Chlebowski et al., 2013). Studies have demonstrated that PRs may control certain target genes independently of ER, significantly affecting the survival and growth of mammary cancer cells (Hilton et al., 2012; J. A. Khan et al., 2012). As a result, PRs found in breast tumors that are PR-positive can be considered as possible targets for therapy. In preclinical studies, it was shown that anti-progestogens such as mifepristone or onapristone can effectively suppress the activity of PRs to control the growth of breast

cancer (Gaddy et al., 2004; Michna et al., 1989). These findings support the potential use of these kind of drugs to target PRs PR+ tumors.

1.5.1.2.3 MR Signaling and Its Relationship with Breast Cancer

The Mineralocorticoid Receptor (MR/NR3C2) belongs to the subfamily of steroid receptor family which also includes the PR, ER, AR and GR (Weikum et al., 2018). MR was first discovered as an aldosterone-binding receptor in rat kidney tissue as a key regulator of the reabsorption of sodium in the kidney (Funder et al., 1972). MR is expressed in many tissues and cell types; in particular, the distal convoluted tubules and cortical collecting ducts of the kidney express MR (Rocha et al., 2000). Additionally, it is also found in the liver, intestine, lung and non-epithelial structures including the heart and brain (Agarwal et al., 1993; Tesch & Young, 2017).

While the primary function of MR in mammals is to respond to aldosterone for regulating fluid balance and potassium levels (Shibata, 2017), while cortisol and corticosterone, the primary glucocorticoids in teleosts (Mommsen et al., 1999), also have a significant affinity for MR although being 100–1000 times more prevalent in plasma (Arriza et al., 1988; Gomez-Sanchez & Gomez-Sanchez, 2014). Selectivity of the MR for aldosterone or cortisol depends on the presence of the enzyme 11β-hydroxysteroid dehydrogenase 2 (11β-HSD2/HSD11B2) which is co-expressed with MR in epithelial cells (Krozowski & Funder, 1983). This enzyme is particularly abundant in aldosterone-sensitive cells, such as kidneys. Its primary function is to break down cortisol into cortisone, which is an inactive state and cannot activate MR, allowing aldosterone to activate the MR primarily in these cells and limit activation of the GR (C. R. W. Edwards & Stewart, 1991). In

general, glucocorticoids are the physiological ligands for the MR in tissues where HSD11B2 is not expressed (Iqbal et al., 2014). Therefore, in tissues that lack this enzyme but still contain MR, such as certain brain regions, cortisol can still activate MR, suggesting additional functions beyond its mineralocorticoid role (Berger et al., 2006; Rozeboom et al., 2007).

MR/NR3C2 displays a dynamic process of activation and regulation. MR is mainly found in the cytosol in its basal state (unliganded) (Rogerson et al., 2004). During this stage, MR is a component of a heterocomplex that also consists of immunophilins (such as FKBP52) and protein phosphatase 5 as well as chaperone heat shock proteins (HSPs), such as HSP90 which has a role in facilitating the binding of ligands to MR (Galigniana et al., 2010). When bound to ligands (aldosterone), MR undergoes conformational changes that cause it to dissociate from HSPs. Ligand-bound MR then organizes into homodimers and translocates from the cytosol to the nucleus (Grossmann et al., 2012; J. Yang & Fuller, 2012), and directly binds to its specific target DNA sequences, HREs (Galigniana et al., 2010). Through this association, MR is able to regulate the transcription of its target genes such as SGK1, EGFR and others, ultimately affecting their expression. It was also shown that HREs can bind GR (Payvar et al., 1983) and the crystal structure of the DNA-binding domain (DBD) of MR bounded to HRE closely resembles that of the GR (Hudson et al., 2014).

Progesterone also has a binding affinity to the MR similar to that of aldosterone and cortisol, but it has a low agonistic activity (Quinkler et al., 2002). In early studies, it was shown that progesterone is an antagonist of aldosterone in rat MR (Kagawa, 1958; Wambach & Higgins, 1978) and human MR (Landau & Lugibihl, 1958). These studies

highlighted that although progesterone has a high affinity for MR, it may not be a transcriptional activator of MR. Moreover, later studies using recombinant human MR revealed that progesterone, aldosterone, cortisol and corticosterone had similar binding affinities for human MR (Arriza et al., 1987) with a finding that progesterone was an antagonist of human MR (Katsu et al., 2018; Rupprecht et al., 1993). Additionally, in one study, it was reported that progesterone had a significant effect on the transcription of the MR gene, both *in vitro* and *in vivo* (Castrén et al., 1995). In *vitro* experiments showed that progesterone significantly increased MR promoter activity in neuroblastoma and pituitary cells, and it caused a dose-dependent increase in MR mRNA levels in hippocampal neuron cultures. *In vivo*, the chronic administration of progesterone, following estrogen priming, significantly increased the levels of MR mRNA in different subfields of the hippocampus. These results suggest that progesterone exerts a regulatory influence on MR expression, particularly in the hippocampus, and highlight the complex interplay between progesterone and estrogen in modulating corticosteroid receptors in the brain. Another study has also shed light on the physiological effects of progesterone acting as an antagonist to human MR (Geller et al., 2000). It demonstrated that progesterone activates human MR with a mutation known as S810L. It was shown in the same study that this mutation resulted in elevated blood pressure due to salt reabsorption, demonstrating how progesterone at normal levels could activate this mutant MR. Progesterone and other steroids without 21-hydroxyl groups, which are generally MR antagonists, function as potent agonists as a result of this mutation's continuous MR activity and change in receptor behavior.

Only a limited amount of studies have highlighted the unique roles of the mineralocorticoid receptor (MR) in breast tissue, indicating that these activities have not

been well investigated (Voutsadakis, 2016). According to these results, MR may have a role in progesterone receptor crosstalk, which promotes cell attachment and stops breast cancer cells from growing (Leo et al., 2004). Two important conclusions were drawn from a study (Muscat et al., 2013) in which the expressions of all 48 nuclear receptor superfamily members in both normal and malignant breast tissues were examined: (i) In contrast to normal breast tissues, malignant breast tissues expressed less MR, and (ii) among patients treated with tamoxifen for breast cancer, a separate predictor of metastasis-free survival was MR. The results of a study examining possible interactions between retinoic acid signaling and MR in breast cancer showed that MR and RAR signaling together suppressed the estrogen-dependent and independent proliferation of breast cancer cells by coordinating the regulation of a distinct collection of gene targets (Doan et al., 2020). The result of another study revealed that low-grade proliferative abnormalities in mammary glands have been correlated with the status of MR and GR together with the 11βHSDs (Cai et al., 2022). Additionally, it was posited a chance that the interaction between MR and PR had an impact on the growth of breast cancer cells.

1.5.1.2.4 Drsp and Its Associations with MR, ER, and PR in Breast Cancer

As a novel synthetic progestin, Drsp is commonly used in HRT with a combination of estrogen (Palacios et al., 2006). Since the use of estrogen-progestin HRT is known to increase the risk of breast cancer compared to estrogen-only HRT (Beral & Million Women Study Collaborators, 2003; Marjoribanks et al., 2012), the progestin component is an important factor that may alter this risk. However, unlike the older generation of progestins, studies examining DRSP's impact on breast cancer are few. In the most

proliferative MCF-7 cell line in response to E2, Drsp was found to increase the proliferation of the cells with same efficacy as other progestins but less potent than natural progesterone (P4) (Perkins et al., 2022). In mouse ex vivo model, progestins having anti-androgenic properties such as Drsp failed to stimulate PR target and Rankl which is a mediator of PR signaling in the mammary epithelium, in contrast to other progestins related to testosterone (Shamseddin et al., 2021). Furthermore, while there has been evidence of an increased risk of breast cancer due to P4, other studies could not find any association between P4 and breast cancer (reviewed in Perkins et al., 2018). The differential effect of progestogens on the proliferation indicates their effect might be depend on the cell model system that is used, since some studies have reported anti-proliferative effects of P4 (Formby & Wiley, 1999; Horwitz & Freidenberg, 1985), while others showed increased proliferation (Carvajal et al., 2005; Tian et al., 2018).

Moreover, while other progestins such as MPA and NET increased the expression of CTSD which has been associated with invasion and metastasis of breast cancer, Drsp could not affect it in T47D cells (ER+/PR+) (Perkins et al., 2022; Xing & Archer, 1998). Drsp also showed significant inhibition of proliferation in human benign epithelial breast cells (HMECs), while the combination of Drsp and estrogen in T47D cells significantly inhibited the E2-induced proliferation (Seeger et al., 2011). However, the clinical studies showed increased breast density and the proliferation of breast epithelial cells after the Drsp and estrogen combination (Hirschberg et al., 2019).

Studies revealed the positive effects of 2 mg of Drsp + 1 mg of estrogen combination on body weight and blood pressure due to anti-aldosterone activity of Drsp (Schürmann et al., 2004; Warming et al., 2004). While Drsp when combined with estrogen as HRT

counteracts the effects of E2 on water and salt retention by inhibiting the MR (Palacios et al., 2006), anti-aldosterone mechanism of Drsp was not investigated in the breast cancer cells. As an antimineralocorticoid, DRSP is about 8 times more effective than spironolactone, with a relative binding affinity for the MR that is 2 to 5 times greater than that of aldosterone (Krattenmacher, 2000).

While progestins can regulate the ER target genes, progesterone treatment causes a decrease expression of PR which is important effector of ER signaling and a clinical prognostic indicator of ER activity in breast cancer cells (Godbole et al., 2017). However, in another study, progestin was shown to upregulate PR expression in ER+/PR+ cells (Diep et al., 2016). Furthermore, progesterone and Drsp do not bind to the ER while Drsp shows high affinity to PR (Fuhrmann et al., 1996). However, molecular studies for Drsp in breast cancer are lacking and more studies are needed for the interaction of Drsp with receptors in the breast cancer cell.

Aims and Rationale

Utilizing now the most commonly used high-throughput sequencing technology, RNA sequencing, or RNA-seq, has completely changed our understanding of the transcriptomes, which are cell/tissue/condition specific. RNA sequencing methods are now widely used in the biological sciences, producing enormous amounts of high-throughput datasets. Analyzing large datasets might be difficult for researchers without considerable programming or statistical expertise. This emphasizes the need for highly flexible and user-friendly and accessible tools when evaluating and interpreting findings from massive sequencing data. Although there are a number of RNA-seq analysis tools available in the literature for those without bioinformatics expertise, there is no tool available that can perform multi-group differential gene expression, DGE, based feature/gene selection followed by individual functionalities for: 1) clustering, 2) annotation, and most importantly 3) prioritization.

There are several tools for clustering however there is not one that takes into accounts some clusters contain gene sets that are inversely correlated with each other, aka. *mirror clusters*. Moreover, there is not a general-purpose RNA-seq tool that can annotate mirror clusters using a network-based visualization. Finally, there is no tool yet that can prioritize gene clusters or individual genes based on a) “gene essentiality”, which can rank genes according to their importance in supporting cell survival; or b) “congruence between mRNA and protein levels”, which can rank genes accordingly how mRNA levels could predict the protein levels of a gene, in a tissue specific manner.

1) Accordingly, the first aim of this study is to design a Shiny-based application using the R programming language to analyze bulk RNA-seq data by performing clustering, DGE analysis, cluster annotation and gene prioritization using several publicly available datasets. In order to facilitate the comprehensive analysis of RNA-seq data, I aimed to incorporate a number of computational and analytical techniques with a particular focus on the identification, annotation and prioritization of clusters/genes that are associated with specific biological functions or diseases, as defined above. With the help of this Shiny application, researchers can have an easy-to-use tool for comprehensively examining RNA-seq data, which will eventually aid in the understanding of gene functions, disease causes, and possible therapeutic targets.
2) The second aim of this thesis is to apply this novel RNA-seq platform to other existing public datasets as well as to our own lab's RNA-seq datasets focusing on scientific fields in which "essential genes" could play important roles. The goal is to identify and prioritize genes of interest across these datasets. These fields were selected as a) neuronal stem cell and brain cancer regulator TrkB signaling, where TrkB is expressed in different isoforms; b) "senescence", where there are promising findings showing "essential gene" regulation could be important; c) breast cancer cell lines, exposed to multiple steroid hormones/regulators where likely to modulate cell proliferation and fate.

i) A first set of questions has included: a) whether receptor tyrosine kinase TrkB is oncogenic or tumorigenic in neuronal stem cells, which can be important leads in brain cancer development; b) If so whether the known spliced isoform without the kinase activity or the full length TrkB behaves differently in comparison to the full length form with respect to essential genes; c) if

complementary datasets in which TrkB was overexpressed or downregulated by siRNAs are used whether a meta-analysis signature or a gene-pair can be identified in association with TrkB gene in glioma.

ii) The enrichment/prioritization of "essential genes" could be important with respect to **senescence**, during which an irreversible shift to a state of growth arrest occurs (please see the introduction). Hence another set of scientific questions I have asked in this thesis are: a) Is senescence associated with a decrease in essential gene expression and what kind of "essential genes" are these? b) What happens to non-essential genes in senescence, are they up- or down-regulated and what are their roles? c) What is the effect of NTRK2 inhibition in senescent cells on the essential/nonessential genes and does this have any relevance to glioblastoma?

iii) To demonstrate the usefulness of CAP-RNAseq on patient data, I used RNA-seq data from whole blood samples of breast cancer patients after treatment with neoadjuvant chemotherapy. In this case, I aimed to find the effect of treatment on essential genes and possible biomarkers for recurrence.

iv) A last set of questions I asked have been based on the intricate crosstalk between steroid hormones that regulate transcriptomes of breast cancers and included: a) whether single or combinatorial use of estrogen and drospirenone (a progestin also targeting MR), prevalently used for birth control as well as a hormone replacement, affect "essential genes" and "non-essential genes" differently; and b) which genes can be diagnostic or therapeutic biomarkers for breast cancer. Since there is no available transcriptomics data in the literature that tested effects of drospirenone and estrogen alone or in combination with/without aldosterone, I have aimed to obtain RNA samples

from T47D (ER+/PR+/MR+) cells treated with these agents and perform an RNA-seq experiment. As validation I have aimed to obtain another RNA-seq dataset this time from MCF7 cells (ER+/PRlow/MRlow) overexpressing MR receptor (MCF-MR) to find the genes affected by these hormones in the case of MR overexpression, and test whether the clusters obtained from T47D corresponded to any MCF7-MR clusters using the RNA-seq tool aimed in the scope of this thesis. By identifying certain genes whose expression are modulated significantly as a result of these treatments, I aimed to provide insight into the possible contribution of essential and nonessential genes to the development of resistance to treatment and the progression of breast cancer.

3) A final aim has been to use CAP-RNAseq in our ongoing studies by using public or in house datasets in the context of a) essential genes in liver fibrosis and senescence; and b) TP53 mutant or wildtype TP53 overexpressing isogenic breast cancer cell line expression patterns.

The impact and novelty of this thesis lies in its development of a highly advanced and specific pipeline for a user-friendly and online RNA-seq application focusing on annotation and prioritization based on “gene essentiality” and congruence of mRNA-protein levels.

Oncogenic signaling pathways are many and can be studied using overexpression and knock-down/knock-out studies of receptors they govern. Brain cancers are deadly diseases in which TrkB has been important and several studies are performed to show that overactivation and downregulation of this receptor and its isoforms can lead to proliferation, or other phenotypes based on RNA-seq studies. There is a need for a

comparative meta-analysis of clusters of expression profiles and this thesis aims to provide a platform by which this can be feasible and provide important leads in brain cancer biomarker discovery.

Senescence, an inherent mechanism of the cells as well as its inducibility by oncogenes and other external factors, and its phenotypic profile suggests that “essential genes” play an important part, yet this is not well studied in a systematic manner and this thesis has the potential to shed a light to this important attribute of life.

As a complicated and diverse disease, breast cancer still requires a thorough investigation of its molecular causes in order to improve our comprehension and therapeutic approaches. Within this framework, the MR and ER crosstalk has been chosen and is expected to increase our understanding of estrogen-progestin interactions in the presence or absence of aldosterone, which can be an important modulator since it is equivalent to “salt” in our diet. Furthermore, a higher risk of breast cancer has been linked to the use of HRT during menopause, a condition that affects half of the world's population. It is therefore crucial to evaluate these hormones used in HRT, paying particular attention to the combination of estrogen and drospirenone, with or without aldosterone, in cells that overexpress MR. An earlier investigation carried out in our lab examined the relationship between MR and the antagonist Spironolactone (Spiro) and agonist Aldosterone (Aldo) and discovered a reversal of the effects of Aldo by Spiro (Seniye Targen, PhD Thesis, Bilkent University). This previous study and its findings have created a need to study the effects of the PR agonist and MR antagonist Drospirenone (Drsp) derived from Spiro. Although Drsp_E2 is used extensively in hormone replacement therapy, there are only a few extensive studies investigating its impact on breast cancer cell lines, especially when

combined with MR and Aldo there is no study examining the association of these hormones with modulation of “essential genes” in breast cancer. The treatment with multiple hormones results in multiple treatment and control groups and this necessitates an online RNA-seq analysis tool, specifically designed for this purpose and will be highly versatile for public and in-house datasets.

CHAPTER 2

MATERIALS AND METHOD

2.1 Development of CAP-RNAseq R Shiny Application

CAP-RNAseq is a web tool designed to facilitate more efficient analysis of RNA sequencing data. It offers users the possibility to upload their own raw count data from RNA-Seq experiments having replicated samples. It also contains demo datasets from different studies for those who want to familiarize themselves with its functionality or for educational purposes. The primary features of CAP-RNAseq are to provide clustering of the data to visualize and annotate the expression pattern of the genes and to prioritize the genes from the clusters based on their a) essentiality scores, b) similarity to cluster centroid, c) expression correlation and slope between the mRNA-protein levels, d) their log fold changes and/or base levels of expression. Moreover, this application was designed to perform comparative transcriptomics based on gene sets obtained from different experiments. The workflow of CAP-RNAseq is given in Figure 6.

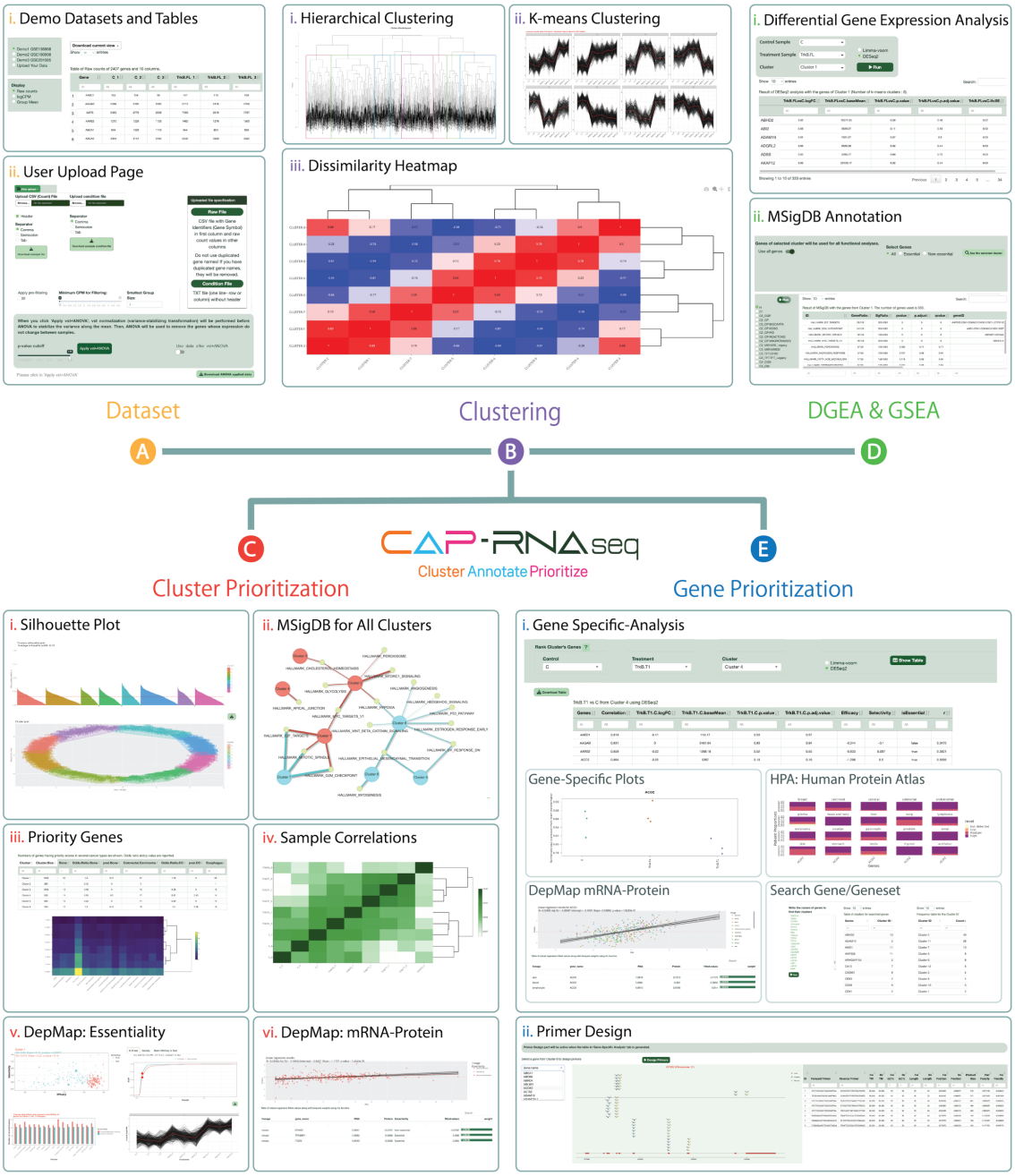


Figure 6. The workflow of CAP-RNAseq.

2.1.1 Publicly Available Datasets Integrated into CAP-RNAseq

CAP-RNAseq utilizes many different datasets from the literature that are then used to prioritize clusters and genes. Each of these is described below.

1. Target Priority Scores: The Project Score database, part of the Cancer Dependency Map (CDP) (Dwane et al., 2021), provides a list of genes with target priority scores across different cancer types. In this database, target priority scores were calculated by integrating various sources of information such as cell line fitness effects obtained from CRISPR-Cas9 experiments, genomic biomarkers, and frequency of somatic alterations in patient tumors (Behan et al., 2019). This score is a metric that indicates possible therapeutic targets based on their relevance and importance. The dataset consists of 2,879 protein-coding genes and 15 different cancer types, with each gene's score ranging from 0 to 100, representing its candidacy as a potential therapeutic target. The target priority score dataset was downloaded from https://score.depmap.sanger.ac.uk/downloads.

2. DepMap mRNA and protein expression data: One of the critical datasets integrated into CAP-RNAseq is derived from the Dependency Map (DepMap), a platform that contains information about the genetic dependencies of cancer cells (Behan et al., 2019; Tsherniak et al., 2017). DepMap includes data obtained from genetic perturbation techniques such as CRISPR and RNAi screens that help determine the genes essential for cancer cell survival and proliferation. Alongside the genetic dependence information, DepMap includes paired expression results for mRNA and protein data that have been incorporated in CAP-RNAseq. This makes it possible to determine the relationship between the genes' amounts of mRNA and protein expression. The depmap package in R

was used to seamlessly obtain TPM (transcripts per million for RNA) and proteomics data (Laurent Gatto, Theo Killian, 2023). Data on RNA expression levels were taken from the 22Q1 CCLE (Cancer Cell Line Encyclopedia) RNAseq result, which contains 19,177 genes in 1,393 cell lines, 33 primary diseases and 38 lineages. On the other hand, the proteomics data which includes information on 12,399 proteins in DepMap came from 20Q2 quantitative protein profiling by mass spectrometry (Nusinow et al., 2020). The use of mRNA and protein expression data allowed the calculation of correlation between mRNA and protein expression of each gene in R. First, the correlation coefficient between each gene's mRNA and protein expression across all samples was determined using Pearson's method. Furthermore, linear regression analysis was performed to gain insight into the relationship between the mRNA and protein levels of the genes. A new dataset was generated using this analysis and included all the information obtained from this analysis, such as the slope, intercept and associated p-values, as well as the R-square and adjusted R-square values used to evaluate the linear regression model's goodness of fit. This new dataset was then integrated into CAP-RNAseq and presented as a table where the distance correlation between each gene expression pattern and the cluster center was calculated.

3. Efficacy and Selectivity scores: Another dataset integrated into CAP-RNAseq includes the essentiality and dependency scores calculated by the shinyDepMap tool (Shimada et al., 2021), which integrated data from CRISPR and shRNA experiments of DepMap to assess the growth inhibition caused by gene knockdown or knockout, as well as the selectivity of this effect across different cell lines. The authors of the shinyDepMap calculated the dependency scores for 15,847 genes across 423 cell lines by considering both CRISPR and shRNA data with differential weighting. CAP-RNAseq incorporated

the default version of the data, which has a mixing ratio of 60:40 and q1 percentile. Other versions (e.g. 20:80, 40:60, 80:20 and q5, q10,q25) were also tried and no significant differences were observed. Two terms were also derived from combined effect scores in shinyDepMap: efficacy, which measures the extent of growth inhibition due to gene knockout or knockdown in sensitive cell lines; and selectivity, which characterizes the variation in efficacy between sensitive and resistant cell lines. The following datasets were downloaded from: https://doi.org/10.6084/m9.figshare.13653257.v1 for dependency scores and https://doi.org/10.6084/m9.figshare.13653260.v1 for efficacy/selectivity values.

4. Protein expression data of HPA: The Human Protein Atlas (HPA) was used to obtain another set of protein expression data, a database used by R's HPAanalyze package (Tran et al., 2019). It provides protein expression levels of 20,082 genes, determined by IHC staining, for 20 different cancer types and 63 normal tissue/cell types. Protein levels are categorized as "high", "moderate", "low" or "not detected".

2.1.2 Data Processing in CAP-RNAseq

CAP-RNAseq allows users to upload bulk (e.g., short-read Illumina or alike) RNA-seq count data together with a text file including condition/phenotype labels of samples. Depending on whether the RNA-seq data are sufficiently filtered, filtering may be required after upload. Filtering the genes is crucial in CAP-RNAseq since the high number of genes affects the processing speed of the application and the generation of redundant clusters. Initial filtering can be performed using counts per million (CPM) values, which remove the genes if the CPM values in some samples (the number is user-

defined) are less than the specified minimum value by the user. However, this may not be sufficient to filter the data to give meaningful and easily understandable clusters. Therefore, to address this, CAP-RNAseq utilizes ANOVA after applying the variance stabilization transformation (vst) method from DESeq2 package (Love et al., 2014) to select genes that show significant changes between any two or more groups.To obtain more significantly altered genes, the p-value threshold for the ANOVA filtering can be adjusted.

2.1.3 Clustering in CAP-RNAseq

CAP-RNAseq performs two unsupervised learning algorithms for clustering: hierarchical clustering and k-means clustering. The scaled logCPM values of data were used for hierarchical clustering with the parameters of user-selected linkage criteria and distance methods. It aids in determining the number of clusters that can be obtained from the data. On the other hand, genes are assigned to the designated number of cluster centroids based on their scaled logCPM values using k-means clustering, which is employed by R's kmeans function. The resulting clusters can be visualized as line or box plots showing each cluster's gene expression patterns, highlighting the clusters' centers as red lines.

Additionally, the number of differentially expressed genes inside each cluster can be visualized using CAP-RNAseq for chosen comparisons before going to the next parts. This is done by applying a cutoff value for absolute logFC values, and results are visualized in a two-column table showing the number of genes after filtering the results based on p-values or without any filtering. Moreover, silhouette analysis is performed with a user-selected cluster number range to check how many clusters are well clustered

in different cluster selections. This also provides insight into determining the cluster number used in k-means.

2.1.4 Mirror Clusters

After generating the clusters, the dissimilarity among cluster centroids is calculated by Pearson's method and presented in a heatmap. This heatmap enables users to identify clusters with inverse expression patterns, referred to as *mirror clusters*. The most dissimilar clusters are indicated by the lowest scores and CAP-RNAseq generates a table displaying these mirror clusters with the strongest negative correlations.

2.1.5 Cluster Prioritization

CAP-RNAseq uses several analyses to prioritize a cluster as described below:

1. Silhouette Analysis: After applying k-means clustering, Silhouette information is calculated for the resulting clusters using the "silhouette" function from the cluster package (Maechler et al., 2023). This analysis gives information about how well each gene fits into the cluster to which it has been assigned. CAP-RNAseq provides the silhouette width scores in a table displaying the smallest average dissimilarity between the cluster of each gene and its neighboring cluster. Moreover, factoextra package (Kassambara & Mundt, 2020) was used for the visualizations of silhouette width as ggplot2-based. All these analyses help users to identify the clusters that may include genes placed in the wrong clusters.

2. MSigDB Analysis: In CAP-RNAseq, functional annotation is utilized using the

Molecular Signatures Database (MSigDB), which is a comprehensive collection of gene sets (Liberzon et al., 2015), for all clusters simultaneously. The user can select the genes which will be used in the analysis based on their essentiality and/or their mRNA-protein correlation scores obtained from DepMap data. For example, analysis can be done only for essential genes or non-essential genes. By selecting a threshold for correlation coefficient, such as greater than 0.5, only genes with a relatively high correlation between mRNA and protein expression values can be included in MSigDB analysis.

All available gene set collections are retrieved from MSigDB using the msigdbr R package (Dolgalev, 2022) and enrichment analysis is performed by enricher function of clusterProfiler package (T. Wu et al., 2021). After the results are presented as a table, network visualization of these results can also be performed using the visnetwork package (Almende B.V. and Contributors & Thieurmel, 2022), where nodes represent the clusters and functional terms, and edges are the connectors between clusters and MSigDB terms, whose widths indicate the number of genes overlapping with each gene-set collection.

Since network visualization becomes more difficult and takes longer as the number of nodes increases, options for q-value filtering, cluster selection, and gene set collection are included. Moreover, a sample comparison is included here to color the cluster nodes based on the changes in expression levels. By taking average logCPM values for replicates and calculating the differences, cluster nodes are colored as red if this score is greater than zero indicating the upregulation, otherwise, they are colored as blue indicating the downregulation.

3. Priority scores: Following the generation of clusters, CAP-RNAseq computes and displays the number of genes that have target priority scores for each cancer type within each cluster in a tabular format. The table also provides odds ratios to score the significance of the association between genes with priority scores in a specific cancer type within the selected cluster compared to all genes across all clusters. This is accomplished by prioritizing a cluster that has a higher association score in the cancer type(s) of interest by utilizing the oddsratio function from the epitools package (Aragon, 2020). A stronger association between the genes having priority scores and the particular cancer type inside a cluster is shown by higher scores of odds ratios. This information can direct users to focus on clusters that are more likely to include more genes for an interested cancer type. The results can be visualized as a heatmap displaying the values of the Odds Ratio, log of Odds Ratio, or percentage with a priority score.

4. Sample Correlations: Within the selected cluster, the samples are subjected to Pearson correlation in order to examine the similarity between the replicates. This is visualized as an interactive heatmap using heatmaply package (Galili et al., 2018). This can be used to select clusters by looking at the similarities of replicates within it.

5. Efficacy and Selectivity from DepMap: The values of Efficacy and Selectivity were taken from the ShinyDepMap tool. For each cluster, the mean (+/- std) of efficacy and selectivity scores of genes, percentage of essential genes and odds ratios for essential genes from the Fisher's exact test are provided in a tabular format. When one of the clusters are selected from the table by clicking, efficacy and selectivity scores are displayed in a scatter plot, in which essential genes are color-coded with red, and nonessential genes are color-coded with blue. Additionally, this plot can be redrawn with

other several options like the log2 transformed average CPM , coefficient of variation on CPM, and log2 of range CPM, etc. Furthermore, comparing densities, median differences and density distribution for selected options can be performed by the Kolmogorov–Smirnov test, the Mann-Whitney U test and the density plot, respectively.

2.1.6 Differential Gene Expression and Functional Analysis

Differential Gene Expression Analysis (DGEA) can be performed for the genes of selected clusters using DESeq2 (Love et al., 2014) or limma-voom (Ritchie et al., 2015) for desired sample comparisons. This resulted in a table displaying the genes within the chosen cluster that were substantially impacted by the treatment together with statistical metrics such as p-value, corrected p-value, and log2 fold change (logFC). After that, functional analysis using MSigDB can also be performed in a cluster-specific manner. This analysis can be further refined by filtering the genes based on their logFC values and essentiality.

2.1.7 Gene Prioritization

After a cluster is prioritized, the gene/s of such cluster can also be prioritized using several options.

1. Distance correlation with the centroid of the cluster: To select a gene that can best represent the average expression profile of the selected cluster, a measure of distance correlation is calculated by dcor function of energy package of R between each gene and cluster centroid. The result displayed in a table is ranked from highest to lowest correlation, together with other scores of genes such as fold change, p-value,

efficacy/selectivity, and RNA-protein correlation values. When one of genes is selected from this table by clicking, it generates plots that show the normalized log2 expression of the gene from Voom transformation in all samples and expression pattern as observed in k-means clusters against to cluster centroid. One more plot is also generated that shows the dependency score of the gene in several tissues and cells. It also provides lineage-dependent essentiality on the same graph, and the more negative dependency score means that the cell depends on this gene more strongly.

2. HPA protein expression levels: After selecting a gene from the table showing distance correlation, its protein expression level for selected cancer types and/or normal tissues are shown in the stacked bar plots.

3. Correlating mRNA and protein levels: Once a gene is selected, expression data for RNA and protein obtained from DepMap is visualized in a scatter plot where the x-axis shows RNA while the y-axis shows protein levels. The plot also includes a linear regression line and information. Results can be visualized for all lineages or specific diseases or cells. In addition, applying robust regression using bisquare weighting helps to weight observations to find potential outliers.

2.1.8 Search a Gene/Geneset

This part is helpful in finding the cluster of an interested gene or genesets. In the case of multiple genes, it determines how many genes are present in each cluster, and it calculates the expected number based on the cluster size. Then, it performs a chi-square test to

compare the expected and observed results, and these findings are given as a frequency table.

2.1.9 Primer Design

After gene prioritization, primer design can be performed for the selected gene. The exon-intron information of the gene is taken from BioMart (Durinck et al., 2005, 2009) on the server side of the application, and two consecutive exons are randomly selected ten times. The sequence of selected exons is used to design primers flanking an intron. The primer design script was taken from the scAmpDesign package (Cribbs, 2022) using Primer3 (Koressaar & Remm, 2007; Untergasser et al., 2012) and modified to include the latest Primer3 files and to improve the table view and the number of primer pairs returned. The result of this section is displayed as an interactive plot and a table. In the plot, red lines along the x-axis indicate the exon sequences of the gene. The locations of the designed primers are shown above the exons. With the sequence and position information, the table provides multiple features of all designed primers, such as melting temperature (Tm), GC content, and penalty scores.

2.1.10 Publicly Available Demo Datasets

As example datasets, three publicly available RNA-seq data were given in CAP-RNAseq. The count data for each study were obtained from the Gene Expression Omnibus (GEO) database (accession numbers: GSE136868, GSE190998, GSE201085).

GSE136868 includes the overexpression of NTRK2 constructs, a pLJM1 lentiviral vector containing either one of two NTRK2 splice variants (TrkB.FL and TrkB.T1) or GFP

(control) was transfected into human neural stem cells (NSCs) (Pattwell et al., 2020). In the original study, it was found that TrkB.T1 has increased expression in human glioma and increases PDGF-induced Akt and STAT3 signaling pathways in vitro and promotes PDGF-driven gliomagenesis in vivo. Specifically, the RNA-seq results in the original paper indicates that the upregulated genes in TrkB.T1 compared to TrkB.FL are enriched in PI3K/Akt pathway and PI3K/ERBB4/ERBB2 based on the GO analysis.

In GSE190998, RNA-seq was used to profile proliferating WI-38 fibroblasts versus senescent WI-38 cells transfected with either control siRNA or siRNA targeting NTRK2 or BDNF (Anerillas et al., 2022b). BNDF containing samples were removed from the data that was used in CAP-RNAseq, since they were not related to our first demo dataset that analyze the splice variants of NTRK2. In the original study, it was found that increased expression of TrkB (NTRK2) was observed in senescent cells which secrets BDNF, the TrkB ligand, supporting the viability of senescent cells. The activation of TrkB by BDNF promoted the senescent cell survival, and blocking the BDNF signaling caused the decreases viability of senescent cells, which indicated the autocrine roles of TrkB and BDNF.

In GSE201085, RNA sequencing was performed on whole blood samples of 53 breast cancer patients after Neoadjuvant chemotherapy (NAC) completion and prior to surgery. Twenty-three of them received NAC, and they are classified according to whether they experienced breast cancer recurrence (n=7, RD-R) or did not experience recurrence for at least 3 years (n=16, RD-nR) (Axelrod et al., 2022). The remaining are the pathological complete responses (n=9, pCR) and those not received NAC (n=21). In the original study, an immunologic gene signature (PIRS) was found to be lowest in patients with RD with

recurrence and highest in those with the best outcomes (pCR). Moreover, the signature of IFN was identified in the comparison of pCR samples and RD samples, and significant enrichments was found for the terms of the complement pathway, IFN Alpha Response, and signature IFN Gamma Response.

Both GSE136868 and GSE201085 data were integrated into CAP-RNAseq after ANOVA-vst application was performed with p-value cutoff as 0.05. For GSE190998, a more stringent p-value cutoff of 0.001 was used in the ANOVA-vst analysis since using 0.05 gave 10,151 genes which is high.

2.1.11 Web Server Construction

CAP-RNAseq was written in the open-source R programming language and developed using the Shiny framework (W. Chang et al., 2023; R Development Core Team, 2023). It is hosted on http://konulabapps.bilkent.edu.tr:3838/CAPRNAseq/. All analytical workflows were constructed utilizing R libraries for genomics and computing. To enhance user experience and interface consistency, the frontend utilizes a range of R libraries, including Shiny, fresh, shinyWidgets, shinyBS, shinycssloaders and custom Javascript/CSS/HTML enhancements (Table 3).

Table 3. List of packages used in the development of CAP-RNAseq.

Package	Reference	Identifier
R (4.3.1)	(R Development Core Team, 2023)	https://www.r-project.org/
shiny (1.7.4.1)	(Chang et al., 2023)	https://shiny.rstudio.com/
shinyWidgets (0.7.6)	(Perrier et al., 2023)	https://CRAN.R-project.org/package=shinyWidgets
shinycssloaders (1.0.0)	(Sali & Attali, 2020)	https://CRAN.R-project.org/package=shinycssloaders
shinyjs (2.1.0)	(Attali, 2021)	https://CRAN.R-project.org/package=shinyjs
fresh 27.04.2024 15:59:00(0.2.0)	(Perrier & Meyer, 2020)	https://CRAN.R-project.org/package= fresh
shinythemes (1.2.0)	(Chang, 2021)	https://CRAN.R-project.org/package=shinythemes
Shinydisconnect (0.1.0)	(Attali, 2020)	https://CRAN.R-project.org/package=shinydisconnect
shinyBS (0.61.1)	(Bailey, 2022)	https://CRAN.R-project.org/package=shinyBS
edgeR (3.42.4)	(Robinson et al., 2010)	https://bioconductor.org/packages/release/bioc/html/edgeR.html
limma (3.56.2)	(Ritchie et al., 2015)	https://bioconductor.org/packages/release/bioc/html/limma.html
DESeq2 (1.40.2)	(Love et al., 2014)	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
dplyr (1.1.2)	(Wickham et al., 2023)	https://CRAN.R-project.org/package=dplyr
plyr (1.8.8)	(Wickham, 2011)	https://www.jstatsoft.org/v40/i01/
reactable (0.4.4)	(Lin, 2023)	https://CRAN.R-project.org/package=reactable
reactablefmtr (2.0.0)	(Cuilla, 2022)	https://CRAN.R-project.org/package=reactablefmtr
DT (0.28)	(Xie et al., 2023)	https://CRAN.R-project.org/package=DT
ggplot2 (3.4.2)	(Wickham, 2016)	https://ggplot2.tidyverse.org
plotly (4.10.2)	(Sievert, 2020)	https://plotly-r.com
biomaRt (2.56.1)	(Durinck et al., 2009)	https://bioconductor.org/packages/release/bioc/html/biomaRt.html
heatmaply (1.4.2)	(Galili et al., 2018)	https://github.com/talgalili/heatmaply
visNetwork (2.1.2)	(Almende B.V. and Contributors & Thieurmel, 2022)	https://cran.r-project.org/web/packages/visNetwork/index.html
clusterProfiler (4.8.2)	(Wu et al., 2021)	https://guangchuangyu.github.io/software/clusterProfiler/
energy (1.7-11)	(Rizzo & Szekely, 2022)	https://CRAN.R-project.org/package=energy
msigdbr (7.5.1)	(Dolgalev, 2022)	https://CRAN.R-project.org/package=msigdbr
HPAanalyze (1.18.1)	(Tran et al., 2019)	https://github.com/anhtr/HPAanalyze
cluster (2.1.4)	(Maechler et al., 2023)	https://CRAN.R-project.org/package=cluster
factoextra (1.0.7)	(Kassambara & Mundt, 2020)	https://CRAN.R-project.org/package=factoextra

2.2 Hormone Studies in Breast Cancer

2.2.1 Materials

2.2.1.1 Solutions, Kits and Reagents Used in the Experiments

Table 4. List of materials, kits and reagents utilized in cell culture experiments.

Product Name	**Company & Country**	**Catalogue Number**
Dulbecco's Modified Eagle Medium (DMEM) Low Glucose, without L-Glutamine, with Sodium Pyruvate	Biowest, France	B.L0064-500
DMEM, with 1.0g/l Glucose and Sodium Pyruvate, without L-Glutamine, without Phenol Red	PAN-Biotech, Germany	P04-001159
Dimethyl Sulfoxide (DMSO)	AppliChem, Germany	A3672
Lipofectamine® 2000	Thermo Fisher Scientific, USA	11668027
L-Glutamine, 200 mM	Biowest, France	B.X0550-100
3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide (MTT)	Thermo Fisher Scientific, USA	M6494
MEM eagle NEAA (100X)	Lonza, Switzerland	BE13-114E
Sodium Pyruvate Solution	Biowest, France	B.L0642-100
Trypsin/EDTA Solution	Lonza, Switzerland	BE-17-161E
Penicillin-Streptomycin Mixture	Lonza, Switzerland	DE17-602E
Opti-MEM® I Reduced Serum Medium	Gibco	31985-070
Dulbecco's Phosphate Buffered Saline w/o Calcium w/o Magnesium	Biowest, France	B.L0615-500
RPMI 1640 w/ Stable Glutamine	Biowest, France	B.L0498-500
RPMI 1640 Medium without L-Glutamine and without Phenol Red	BIO. IND.	BI01-103-1A
FBS South America, Gamma Irradiated	Biowest, France	S181G-500
Fetal Bovine Serum (FBS) South America, Charcoal Stripped	Biowest, France	S181F-100

Table 5. Hormones used in the experiments.

Product Name	**Company & Country**	**Catalogue Number**
Drospirenone	Sigma-Aldrich, Germany	SML0147-10MG
Aldosterone	Sigma-Aldrich, Germany	A9477

Estrogen was kindly provided by Dr. Murat Alper Cevher's laboratory at Bilkent University.

Table 6. List of materials, kits and reagents for RNA experiments.

Product Name	**Company & Country**	**Catalogue Number**
QIAzol Lysis Reagent	Qiagen, Germany	79306
Chloroform	Sigma-Aldrich, Germany	24216
2-Propanol	Sigma-Aldrich, Germany	24137
Ethanol	Sigma-Aldrich, Germany	32221
Sodium Acetate	Carlo Erba Reagents, Italy	366207
iScript cDNA Synthesis Kit	BIO-RAD	1708891
LightCycler® 480 SYBR Green I Master	Roche, Switzerland	4887352001
LightCycler® 480 Multiwell Plate 96, White	Roche, Switzerland	4729692001
AccuGENE™ Molecular Biology Water	Lonza, Switzerland	BE51200

Table 7. Laboratory materials, kits and reagents for ChipSeq experiments

Product Name	**Company & Country**	**Catalogue Number**
Formaldehyde	Sigma-Aldrich, Germany	252549-1L
SDS	EMD Millipore, USA	24730020
NaHCO3	Sigma-Aldrich, Germany	S5761-500G
Tris	Sigma-Aldrich, Germany	T1503-1KG
LiCl	Sigma-Aldrich, Germany	L9650-100G
NP-40	Sigma-Aldrich, Germany	NP40S-100ML
DOC	Sigma-Aldrich, Germany	D6750-10G
EDTA	Sigma-Aldrich, Germany	E5134-500G
Hepes	Sigma-Aldrich, Germany	H4034-25G
EGTA	Sigma-Aldrich, Germany	E3889-10G
NaCl	Sigma-Aldrich, Germany	S5886-500G
TritonX-100	Sigma-Aldrich, Germany	T8787-100ML
Protease Inhibitor Cocktail	Sigma-Aldrich, Germany	11873580001
Glycine	Sigma-Aldrich, Germany	G7403-100G
Rnase A	Sigma-Aldrich, Germany	10109169001
Proteinase K	Invitrogen, USA	25530049
Zymo DNA Clean&Concentrator	ZymoResearch, USA	D4034
Agarose	Sigma-Aldrich, Germany	A9539-100G
SyberSafe	Applied Biological Materials, Canada	G108
Dynabeads	Invitrogen, USA	11203D
tRNA	Sigma-Aldrich, Germany	R8508-1ML
BSA	Sigma-Aldrich, Germany	A9418-5G
Magnetic stand	Thermo Fisher Scientific, USA	12-321-D
TAE Buffer	Sigma-Aldrich, Germany	T9650-1L
Estrogen Receptor α (D8H8) Rabbi mAb_100 µl	Cell Signaling Technology, USA	8644S

Table 8. List of primers used in the experiments and their sequences.

Genes	**Primer Sequences**
TPT1 (Reference)	F: 5'- GATCGCGGACGGGTTGT -3' R: 5'- TTCAGCGGAGGCATTTCC -3'
MR (NR3C2)	F:5’-AGTGGAAGGGCAACACAACT-3’ R:5’-CTGCTCCTCGTGAATCCCTTT-3’
TFF1	F5’-CCATGGAGAACAAGGTGATCTGC-3’ R:5’-TTAGGATAGAAGCACCAGGGGAC-3’
PGR	F5’-GGAGGAGGAGGGAGGTATCA-3’ R:5’-CTGCTTGGAAGACTCAGGGA-3’
SGK1	F5’-AACCCTTCTCCTCCACCAAG-3’ R:5’-TTCCAAAACTGCCCTTTCCG-3’
ZBTB16	F5’-AGACGTACCTCTACCTGTGCTATGTGT-3’ R:5’-TGTCATAGTCCTTCCTTCATCTCACT-3’
GR (NR3C1)	F:5’-CTGGGGTGGAGATCATATAGACA-3’ R:5’-CATAAGATACCTGAAGCCTGTGT-3’
CCND1	F:5’-CTGCGAAGTGAAACCATCC -3’ R:5’- GCACTTCTGTTCCTCGCAGA-3’
S100P	F: 5’GGATGCCGTGGATAAATTGC-3’ R:5’ CGAACACGATGAACTCACTGAAG-3’
CCNE2	F:5’- GTAGCTGGTCTGGCGAGGTTT-3’ R:5’-GGGCTGCTGCTTAGCTTGTAA-3’
CHEK1	F:5’-TGGTCACAGGAGAGAAGGCA-3’ R:5’-CAGATAAACCACCCCTGCCA-3’
BCL2	F:5’-TGAACTGGGGGAGGATTGTG-3’ R:5’-CGTACAGTTCCACAAAGGCA-3’
BAX	F:5’-GGGTTGTCGCCCTTTTCTAC-3’ R:5’-CTGGAGACAGGGACATCAGT-3’
RRM2	F:5’-TCCGGTTCTTTTGCGTCGAT-3’ R:5’-TCCGATGGTTTGTGTACCAGG-3’
FKBP5	F:5’-CACCATTCCCCACTCTTTTG-3’ R:5’-TGATGAAGGTGCCAAGAACA-3’

NDRG1	F:5’- CTCCTGCAAGAGTTTGATGTCC-3’ R:5’- TCATGCCGATGTCATGGTAGG-3’
BCL6	F:5’-GGCCTTAATAAACTTCCGGC -3’ R:5’-AGAGCCCATAAAACGGTCCT -3’
TGFB2	F:5’-GAGTGCCTGAACAACGGATT-3’ R:5’-TTCGCCTTCTGCTCTTGTTT-3’
GREB1_ChIP	F: TGTGTGGAGAGCACTGTGAC-3’ R: AACTGGAGCTTCATCTGCCC-3’
PGR_ChIP	F: TAATGCCTCTCCCATGTCTCCA-3’ R: GTCCCAGAATGTCAGCCTCAT-3’
TPRG1 _positivecontrol_ChIP	F: CCAGTGTTGCTGTCTGTTGG-3’ R: ACCGTGACCCAAAACTTCCT-3’
Neg. Ctrl2	F: GGTGAGATGGATTGGTAAGACCC-3’ R: AAAGCCAGATTTGACCAAGCA-3’
Neg. Ctrl3	F: CATTTTAAGCAGACCCACGC-3’ R: GGCTCAATGACCTTCCGAAC-3’

2.2.1.2 Equipments

Table 9. List of equipment used in the experiments.

Instrument Name	**Company, Country**
NanoDrop ND-1000	Thermo Fisher Scientific, USA
PCR Thermal Cycler 2720	Applied Biosystems, USA
LightCycler® 96 Instrument	Roche, Switzerland
Biological Safety Cabinet	NuAire, USA
Air-Jacketed Automatic CO2 Incubator	NuAire, USA
Synergy HT Microplate Reader	BioTek, USA
Amersham™ Imager 600	GE Healthcare Life Sciences, USA

2.2.1.3 Plasmids

The pcDNA3.1 plasmid (empty vector) was gratefully donated by Dr. Ebru Erbay's group at Bilkent University, while the MR-STAG-pcDNA3.1 plasmid (control vector) was kindly provided by Dr. Nourdine Faresse from the University of Zurich, Switzerland.

2.2.2 Methods

2.2.2.1 Cell Culture Experiments

The MCF-7 cell line was grown in DMEM with 10% FBS, 1% nonessential amino acids, 2% L-Glutamine and 2% sodium pyruvate.Additionally, 1% penicillin-streptomycin was added to inhibit bacterial contamination. On the other hand, the ZR-75-1 and T47D-1 cell lines were cultured in RPMI containing glutamine along with 10% FBS, 1% nonessential amino acids, 2% sodium pyruvate and 1% penicillin-streptomycin. The cells were cultivated in T75 flasks, and when confluence of cells reached almost 70-80%, they were passaged. All cell lines were sustained in a constant temperature of 37°C within a 5% CO2 incubator.

2.2.2.2 Transfection of Cells with MR

Transfection was applied only to MCF-7 and ZR-75-1 cell lines following their second passage. To prepare the cells for transfection, 200,000 cells were placed in each well of 6-well plates. After 24 hours in order for cells to adhere, the transfection process was conducted following the protocols provided by Lipofectamine 2000 manufacturers. The transfection mixture was prepared using two plasmids: the pcDNA3.1 vector as empty

plasmid and MR-Stag-pcDNA3.1 vector as control plasmid. These vectors were prepared by Seniye Targen.

In separate Eppendorf tubes, each one microgram of plasmid was combined with 250 µl of serum-reduced OPTI-MEM I, and two µl of Lipofectamine was combined with 250 µl of OPTI-MEM I, separately. After all tubes had been let to sit at room temperature for 5 minutes, the contents of both tubes (lipofectamine and one of vectors) were mixed by pipetting up and down to form the plasmid and Lipofectamine complexes. Then, the mixture containing the plasmid and Lipofectamine was incubated at room temperature for 20 minutes to form DNA-lipofectamine complexes. During this incubation time, the liquid in the 6-well plates was removed and replaced with 1500 µl of media without antibiotics. After the incubation period, 500 µl of the prepared mixture was slowly added to the plates and allowed to incubate for 4-6 hours before changing the media.

2.2.2.3 Hormone Treatments

All cell lines were subjected to serum deprivation for 24 hours before applying hormone treatments to eliminate the masking effects of hormones. This makes sure that there are no hormones left over that could interfere with the subsequent treatments. The cells were washed twice with PBS and then, cell-specific mediums (RPMI for T47D and DMEM for MCF-7) supplemented with 5% Charcoal FBS phenol red-free medium were given to the wells.

All stock hormones (drospirenone (Drsp), aldosterone (aldo) and estrogen (E2)) were stored at -20°C. Prior to first use, these hormones were mixed with ethanol (EtOH) to achieve stock concentrations of 30 mM, 10 mM and 2 mM, respectively. Then, in each

use, working solutions from stock concentrations. were then prepared at concentrations of 20 μM for Drospirenone, 10 μM for Aldosterone, and 2 μM for Estrogen in a medium containing 5% charcoal-stripped FBS without phenol-red. Subsequently, the these solutions were more diluted to achieve the desired final concentrations of the hormones. After that, various combinations of these drugs were prepared based on the specified final concentrations according to the experimental design, including Drsp, Aldo, E2, Drsp_Aldo, Drsp_E2, Aldo_E2, and Drsp_Aldo_E2 samples. These treatments were administered to the cells in the 6-well plate. Then, the plates were incubated for 24 hours in the incubator.

2.2.2.4 Collection of Cell Pellets

After 24 hours of incubation of cells with the hormone treatments, cells were washed with PBS to get rid of any residual media. The cells underwent trypsinization by being exposed to 1x Trypsin-EDTA for 3–4 minutes. After the observing cells had detached, 5 ml of media was introduced into each well to counteract the trypsin, and then the cells were gathered into 15 ml falcons. Then, these tubes were spun at 1500 rpm for 5 minutes, leading to the creation of a mass of cells at the bottom. After removing the supernatant, cells were rinsed two times with 1 ml of cold PBS. Centrifugation was used in each wash to re-pellet the cells and remove the supernatant. The last wash was followed by snap-freezing the cell pellets in liquid nitrogen to maintain their integrity and storing them at -80°C until use in additional experiments.

2.2.2.5 RNA Extraction

Total RNA isolation was carried out using the QIAZOL lysis reagent following the manufacturer's instructions, and the procedure was carried out under a chemical hood. Before the extraction procedure, the frozen cell pellets were thawed on ice. Seven hundreds µl of lysis reagent were pipetted into each pellet and well mixed. After this mixture was homogenized, it was put into Eppendorf tubes. Each tube was then filled with 140 µl of chloroform and inverted 15 times up and down. After the ten-minute incubation period of tubes at room temperature, they were spun at 13,000 rpm for 20 minutes at 4°C for lysis.

After centrifugation, the mixture separated into three phases: a lower organic layer, a middle layer containing white DNA strands, and an upper aqueous layer with RNA material. Approximately 300 µl of the aqueous phase were carefully transferred to freshly labeled Eppendorf tubes by being careful not to disrupt the other phases. Then, to precipitate the RNA, an equal volume of isopropanol (around 300 µl) was given to each tube containing the aqueous phase, and the mixture was mixed via inversion five to ten times to ensure that the solution was thoroughly mixed. Following 10 minutes of incubation of tubes at room temperature, centrifugation was performed for 12 minutes at 4°C using 13,000 rpm. After this centrifugation, RNA precipitates were easily observed at the base of the tubes, and the supernatants were cautiously removed without disturbing the precipitates.

Firstly, the RNA precipitates were rinsed with 1 ml of 75% ethanol and then spun at 8,000 rpm for 8 minutes at 4°C, leading the pellets becoming easier to see. Then, the supernatant was discarded, and pellets underwent another washing step with 1 ml of pure ethanol,

followed by an 8-minute spin at 8,000 rpm at 4°C. After ethanol removal, the pellets were air-dried in a chemical hood before being dissolved in 30 to 50 µl of nuclease-free water, with the amount depending on pellet size. RNA concentrations were then assessed using NanoDrop.

2.2.2.6 cDNA Synthesis

The iScript™ cDNA synthesis kit was employed to produce cDNA from extracted RNA, in accordance with the manufacturer's guidelines. Kits contain the following 3 essential materials: 5x iScript reaction mix, iScript reverse transcriptase and nuclease-free water.

For each RNA sample, the reaction was set up by mixing 4 µl of iScript mix, 1 µl of reverse transcriptase and 1 µg of RNA. Then, to obtain a final reaction volume of 20 µl, an appropriate amount of nuclease-free water was added. To use as a negative control, samples without reverse transcriptase were also used in parallel for each experiment. Finally, the reactions were put in thermocycler and subjected to the following conditions: 5 minutes at 25ºC, 20 minutes at 46ºC and 60 seconds at 95ºC. Subsequently, the cDNA was diluted in a ratio of 1:20 and stored at -20°C.

2.2.2.7 qRT-PCR Protocol

LightCycler® 480 SYBR Green I Master mix was used in qRT-PCR (Quantitative Real-time Polymerase Chain Reaction) investigations for the quantitative measurement of gene expression. Five µl of Master mix from the kit, one µl of nuclease-free water, one µl of each forward (10 µM) and reverse (10 µM) primer specific to a gene and 2 µl of diluted

cDNA were mixed as 10 µl for each well. Initially, 8 µl of mixture (excluding cDNA) was pipetted into each well of a 96-well plate. Next, 2 µl of 1:20 diluted cDNA was added to the side of the wells to reach the final volume of 10 µl. Then, the plate was prepared by sealing, and subjected to centrifugation at room temperature for 1.5 minutes at 1500 rpm to eliminate any air bubbles and ensure the mixture's uniform distribution. The plate was then placed into the LightCycler® 96 Instrument to start the run with the conditions were described in Table 10. In the qPCR experimental design, each sample/treatment has two technical replicates and two biological replicates. To account for any possible oligonucleotide contamination, a negative control group, consisting of nuclease-free water instead of cDNA, was used for each primer set. The gene TPT1 (Tumor Protein, Translationally Controlled 1) was chosen as the reference/housekeeping gene for the aim of normalizing gene expression.

Table 10. Reaction conditions for qRT-PCR experiments.

Step	**Temperatures (°C)**	**Time**	**Number of Cycle**
Pre-Incubation	95 °C	5 min	1
Amplification	95 °C	10 sec	40-45
	58-60 °C (depending on Tm of primer)	20 sec	
	72 °C	20 sec	
Melting Curve	95 °C	5 sec	1
	55 °C	1 min	
	95 °C	Acquisition per 5°C	
Cooling	40 °C	30 sec	1

After the run finished, cycle threshold (CT) results were obtained using the "Abs Quan (Absolute Quantification)" analysis in LightCycler® 96 System program. Then, the relative fold gene expression of the samples was determined using the delta-delta Ct ($2^{-\Delta\Delta Ct}$) technique (Livak & Schmittgen, 2001). The first step in this method is to calculate ΔCt for each sample by deducting Ct value of the reference gene from the Ct value of the target gene. Then, $\Delta\Delta Ct$ is obtained by subtracting the ΔCt values of untreated/control sample from the ΔCt values of treated/experimental samples. Finally, calculation of logarithmic values to base 2 for $2^{-\Delta\Delta Ct}$ gives the relative fold change of gene expression compared to the reference sample.

2.2.2.8 Western Blot

After the drug treatment, cells pellets were collected and lysed with freshly prepared 60 ul RIPA lysis buffer as described in Table 11. The tubes containing cell pellet and RIPA buffer were kept on ice for thirty minutes, and vortexed every five minutes for ten seconds. Then, the collected lysates were centrifuged at 13,000 RPM for 15 minutes at +4°C, and supernatants were transferred into new tubes. The total protein levels were measured according to BCA values using the BCA Protein Assay Reagent kit (Thermo Scientific, USA). The preparation of standards and solutions followed the guidelines provided by the manufacturers. Synergy HT Microplate Reader was used to measure absorbance levels following a 30-minute incubation period at 37°C. Using the regression line method between the standards' concentration and corresponding absorbance, the total quantity of protein was determined.

All proteins used in the study were separated in 10% acrylamide gel with 15 µg total protein amount and transferred to PVDF membrane with semi-dry system. A new tube was filled with 15µg of protein from each sample. Using RIPA, the concentration was adjusted for each sample up to the maximum volume. The protein solution was then combined with 4X SDS Loading Dye and allowed to denature for 5 minutes at 95°C. Proteins that were not denatured were kept at -80°C. The TGX Stain-Free™ FastCast™ Acrylamide Kit (Bio-Rad, 1610183) was used to preapte a 10% SDS page in accordance with the manufacturer's instructions. For the separating the gel, 4 ml of solution A and B were mixed together and 40 µl of freshly made APS (10%) and 4 µl of TEMED were added. For the stacking gel, 1.5 ml of solution A and 1.5 ml of solution B were mixed, and 15 µl of APS and 3 µl of TEMED were added. After the gels were hardened and placed into the tank filled with 1X running buffer, 2 µl of protein ladder was loaded into first well and remaining wells were loaded with10 µl of protein samples. Using an EC300XL2 power supply, the gel was run at 80V for stacking and 120V for separating. When the loading dye was removed from the gel, run was halted and proteins were transferred to the PVDF membrane using the Anode I, Anode II, and Cathode solutions as described in Table 11. The membranes were blocked with 10% BSA for one hour, incubated with primary antibodies overnight, followed by incubation with secondary antibodies for 1.5-2 hours. Table 12 lists the primary antibodies used in these experiments, along with their working dilutions. The protein bands were visualized using ECL substrate with Amersham Imager 600 chemiluminescence mode. The band intensities of the proteins were calculated using the ImageJ program and the protein levels were expressed as a multiple of the control by proportioning the protein levels to the loading control GAPDH protein and the control values of each experimental setup.

Table 11. Solution and Buffers used in Western Blot.

Solutions and Buffers	**Components**
RIPA Buffer (1 ml)	50 µl of 1M Tris-HCl, 10 µl of 10% SDS, 75 µl of 2M NaCl, 40 µl of 25X PI, 40 µl of 20X PhosSTOP, 10 µl of NP-40, and the volume was increased to 1 ml with ddH2O
TBS (10X)	24 g of Tris and 88 g of NaCl dissolved in 900 ml of ddH2O. After adjusting pH to 7.6 with HCl, the volume was increased to 1000 ml using ddH2O
TBS-T (Tween 20) (0.2%) (1X)	2 ml of Tween20 was mixed with 100 ml of 10X TBS into 900 ml of ddH2O
PI (25X)	2 tablets (cOmplete, EDTA-Free) were dissolved in 840 µl of ddH2O
PhosStop(20x)	2 tablets were dissolved in 1000 l ddH2O
SDS Loading Dye (4X)	50 µl of β-mercaptoethanol was added into 450 µl 4x Laemmli buffer
10% APS	0.1 g APS was dissolved in 1000 µl of ddH2O
Running Buffer (10X)	30.3 g Tris, 144.1 g Glycine and 100 ml 10% SDS, was mixed into ddH2O up to 1000ml
Anode I Buffer	18.15 g of Tris into 100 ml of MetOH and filled up to 500 ml with ddH2O.
Anode II Buffer	1.5 g of Tris into 100 ml of MetOH and filled up to 500 ml with ddH2O
Cathode Buffer	2.62 g of 6-aminocaproic acid was mixed with 100 ml of MetOH and filled up to 500 ml with ddH2O
Blocking Solution (5%)	5 g of BSA is added into 100 ml of 1x TBS-T

Table 12. Antibody list used in Western Blot.

Antibody	**No.**	**Brand and Dilution**	**Incubation**
Anti-mouse	#7076P2	Cell Signaling, 1:5000	2 hours at RT, in 5% BSA
Anti-rabbit	#7074S	Cell Signaling, 1:5000	2 hours at RT, in 5% BSA
BAX	#5023P	Cell Signaling, 1:1000	16 hours at 4°C, in 5% BSA
BCL-2	#2870	Cell Signaling, 1:200	16 hours at 4°C, in 5% BSA
Total-CHK1	sc-8408	Santa Cruz, 1:1000	16 hours at 4°C, in 5% BSA
MR	rMR1-18 1D5	DSHB/University of Iowa, 1:100	16 hours at 4°C, in 5% Milk
GAPDH	sc-47724	Santa Cruz, 1:5000	16 hours at 4°C, in 5% BSA
GR	#12041S	Cell Signaling, 1:1000	16 hours at 4°C, in 5% Milk

2.2.2.9 RNA Sequencing Procedure

Sequencing of RNAs was obtained as a purchased service from BMLabosis, Türkiye. All samples were subjected to a quality control assessment by BMLabosis before the sequencing procedure, and they all passed the quality control standards. Illumina NovaSeq platform was used to run the RNA samples with a 150-bp paired-end sequencing and 20M read depth parameters. After fastq files were delivered to us, all samples were aligned to the hg38 (human) genome assembly and underwent quality control.

Every analysis was carried out utilizing the web-based Galaxy platform (https://usegalaxy.org/). The raw files (fastq) were uploaded to Galaxy platform and underwent quality controls using FastQCRead Quality reports tool available on Galaxy (S. Andrews, n.d.).The reports of this tools confirmed that none of the samples failed or needed trimming. Then, the reads were aligned to the hg38 genome using HISAT2 (version 2.2.1) (D. Kim et al., 2015), and raw counts were obtained using the FeatureCounts tool on Galaxy using standard settings (Y. Liao et al., 2014). After obtaining count data for all samples, the rest of the analyses were performed using the R programming language. With the help of DESeq2 package of R, genes having zero or low expression were filtered out from the data and differential gene expression analysis was performed.

2.2.2.10 Statistical Analysis

In this thesis, GraphPad Prism 9 (GraphPad Software, San Diego, CA, USA) was used for graphical demonstrations, and R was utilized to perform all statistical analyses. All

qPCR results were normalized TPT1 gene and relative fold gene expressions were calculated using $2^{-\Delta\Delta Ct}$ method. One-way and Two/Three-Way ANOVA analysis were used for qPCR results showing dose effect and MTT results, respectively. Pairwise comparisons were performed using Tukey multiple comparison method for qPCR results and Dunnett's multiple comparisons test for Western blot results.

CHAPTER 3

RESULTS

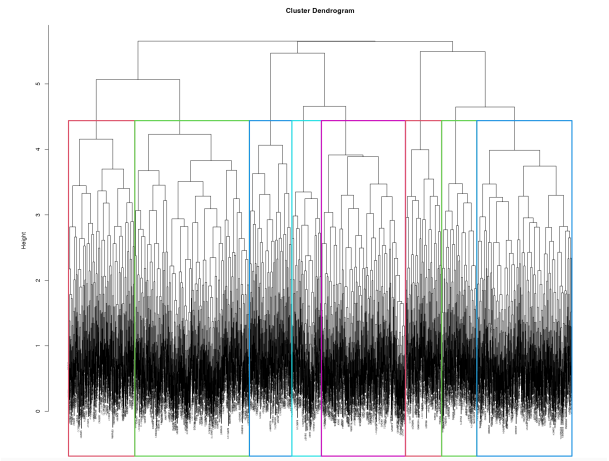
3.1 General Architecture of the CAP-RNAseq App

CAP-RNAseq is modular and has the following tabs: Clustering, Cluster Prioritization, DGEA and GSEA and Gene Prioritization (Figure 6). Using the demo 1 dataset I have explained the general workflow of these tabs in the contexts of GSE136868 dataset.

3.1.1 CAP-RNAseq Demonstration Using Demo 1 Dataset

The expression patterns of genes in GSE136868 data were analyzed in CAP-RNAseq. First, the optimum cluster number was determined using hierarchical clustering and silhouette information. The dendrogram can be cut at any point in hierarchical clustering tab to visualize possible clusters. In this case, the choice of 8 clusters seemed more meaningful (Figure 7A). Furthermore, the findings of the silhouette analysis for cluster numbers 4 through 15 indicated that there were more well-clustered clusters in the clusters with numbers 8 and 13 (Figure 7B). Then, when the graph was checked, the average silhouette width was highest in 8 when compared to 13 (Figure 7C). Therefore, 8 was selected as the cluster number to be used in k-means analysis module.

A)



B)

Clusters	# of well clustered clusters
Cluster number: 8	3
Cluster number: 13	3
Cluster number: 4	2
Cluster number: 5	2
Cluster number: 6	2

C)

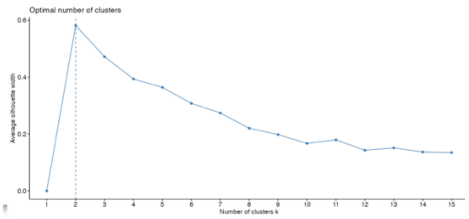


Figure 7. Analysis for finding the optimal number of clusters. A) Cluster dendrogram of Hierarchical Clustering cut for 8 clusters. B) The table shows the number of well-clustered clusters at each cluster number after performing a silhouette analysis for cluster numbers from 4 through 15. C) The graphical visualization of average silhouette width for cluster numbers until 15.

After deciding on 8 clusters, k-means clustering was performed using the following parameters: Number of Clusters: 8, Number of Iterations: 200000 and Number of Random Set: 20. The eight clusters were visualized as line plots showing the expression of each gene as a black line across all samples (Figure 8). The cluster with the highest number of genes that had absolute logFC values greater than 0.5 was found to be the cluster 6, while clusters 4 (among significant DEGs) and 8 (among all DEGs) included the highest number of genes (Table 13).

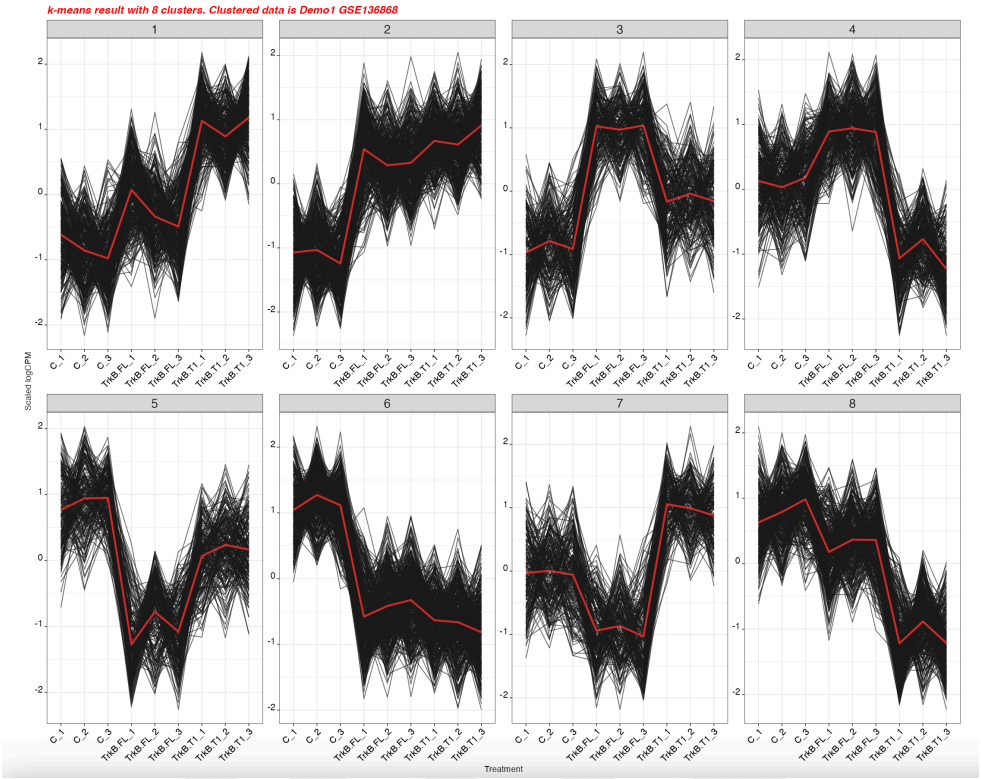


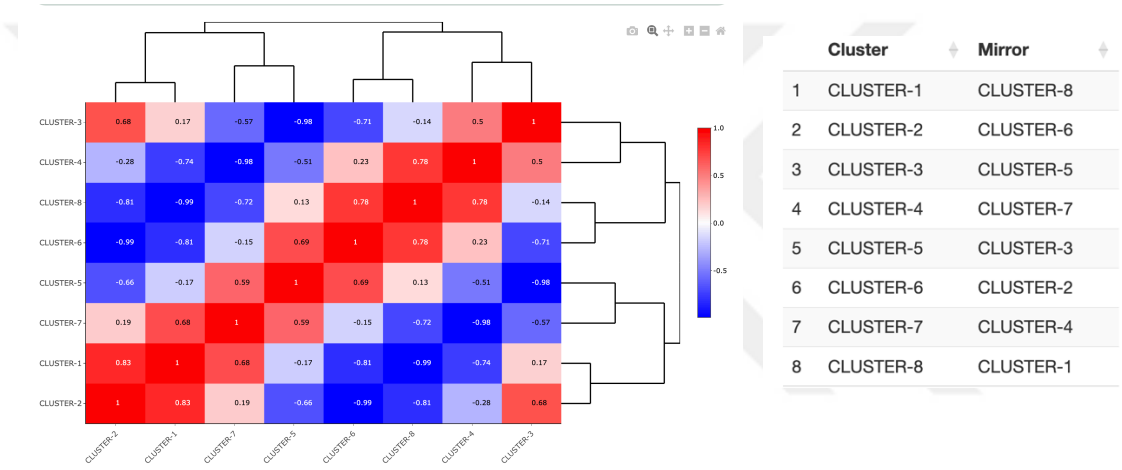
Figure 8. Expression patterns in GSE136868 presented by k -means clustering. The cluster centroid pattern is shown by red lines.

Table 13. The number of differentially expressed genes in each comparison using an absolute log-fold change threshold of 0.5.

	TrkB.FL vs Control		TrkB.T1 vs Control		TrkB.T1 vs TrkB.FL	
	DEGs with p-value<=0.05	All DEGs	DEGs with p-value<=0.05	All DEGs	DEGs with p-value<=0.05	All DEGs
Cluster 1	0	3	14	14	5	8
Cluster 2	12	14	16	16	1	1
Cluster 3	12	15	0	1	6	8
Cluster 4	1	3	6	11	14	16
Cluster 5	30	32	9	12	3	10
Cluster 6	79	90	103	107	2	11
Cluster 7	1	7	1	3	10	12
Cluster 8	1	6	30	35	10	27

After that, one can obtain the mirror clusters, a feature unique to CAP-RNAseq. This module uses Pearson's correlation coefficients to find the dissimilarity between the centroids of the clusters. According to the heatmap and the table, clusters 1 and 8, clusters 2 and 6, clusters 3 and 5, and clusters 4 and 7 were found to be the mirror clusters where each pair showed opposite expression patterns (Figure 9).

A) **B)**



	Cluster	Mirror
1	CLUSTER-1	CLUSTER-8
2	CLUSTER-2	CLUSTER-6
3	CLUSTER-3	CLUSTER-5
4	CLUSTER-4	CLUSTER-7
5	CLUSTER-5	CLUSTER-3
6	CLUSTER-6	CLUSTER-2
7	CLUSTER-7	CLUSTER-4
8	CLUSTER-8	CLUSTER-1

Figure 9. Dissimilarity heatmap and a table for mirror clusters. A) Heatmap shows the correlation values, in which negative values (blue) indicate the lowest similarities and positive values (red) indicate the highest similarities. B) The mirror clusters are shown in a table to be easily identified.

3.1.2 Gene Set Annotation of Demo 1

Next, gene set enrichment analyses based on MSigDB can be run for all clusters and using a selected gene set collection to identify enriched gene sets and pathways. When clusters were annotated using MSigDB and Hallmark gene set for demo 1 data; each cluster's

enriched Hallmark keywords and the genes contained in it can be visualized in a table form. A visual analysis was also conducted using the “network” tab based on the top 10 terms with q-values of 0.05, and cluster nodes were colored according to the differences in logCPM values of the genes between treatment and control groups (in this case; TrkB.T1 vs TrkB.FL).

Among the clusters, cluster 1 and cluster 2, in which genes presented increased expressions upon the overexpression of isoforms, were enriched in several important gene sets, including hypoxia, glycolysis, G2M checkpoints, E2F and MYC targets, MTORC1 signaling, and mitotic spindles (Figure 10). This suggested that higher the TrkB expression cells were likely to divide. On the other hand, genes of cluster 8 (mirror cluster of cluster 1) showing a decreasing pattern inTrkB.T1 samples with respect to the control were enriched in myogenesis and epithelial-mesenchymal transition (EMT). Moreover, the genes of cluster 6 (mirror cluster of cluster 2) were enriched in the P53 pathway, EMT, hedgehog signaling, IL2_STAT5 signaling, angiogenesis, and other related terms. This suggested dividing cells might downregulate pathways relevant with cell migration as well as immune response and apoptosis. For clusters 2 and 6 which contained the genes that have similar expressions in both TrkB.T1 and TrkB.FL samples, hypoxia appeared as the shared term in both cluster; this suggested that hypoxia pathway was likely to have inhibitors and activators and the outcome of hypoxia need to be studied with respect to such changes as whole.

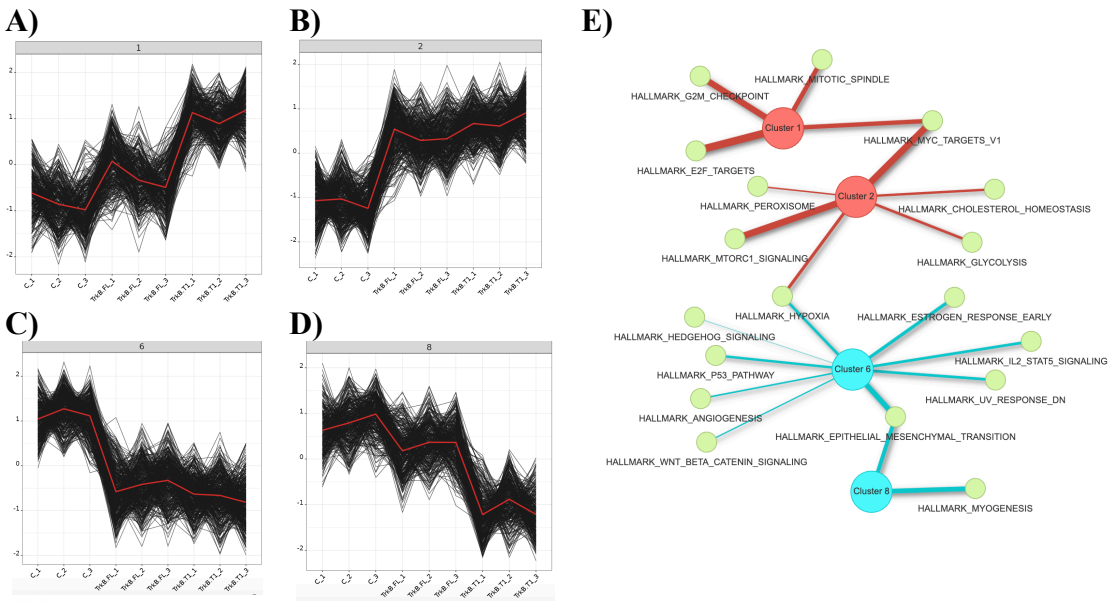


Figure 10. Analysis of clusters 1, 2, 6 and 8. A-B-C-D) Expression pattern of clusters. E) The network visualization of enrichments of genes in each cluster using Hallmark gene set collection of MSigDB. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between TrkB.T1 and TrkB.FL samples.

In clusters 3 and 5, where the most significant changes were observed in TrkB.FL samples when compared to the control, genes enriched in the UV response and the EMT were more abundant in cluster 5, while cluster 3 contained genes enriched in oxidative phosphorylation, MYC targets, cholesterol homeostasis, and adipogenesis (Figure 11). In the last pair of mirror clusters with similar patterns to clusters 3 and 5, E2F targets, the G2M checkpoint, and the mitotic spindle appeared in cluster 7, and only the apical connection was in cluster 4.

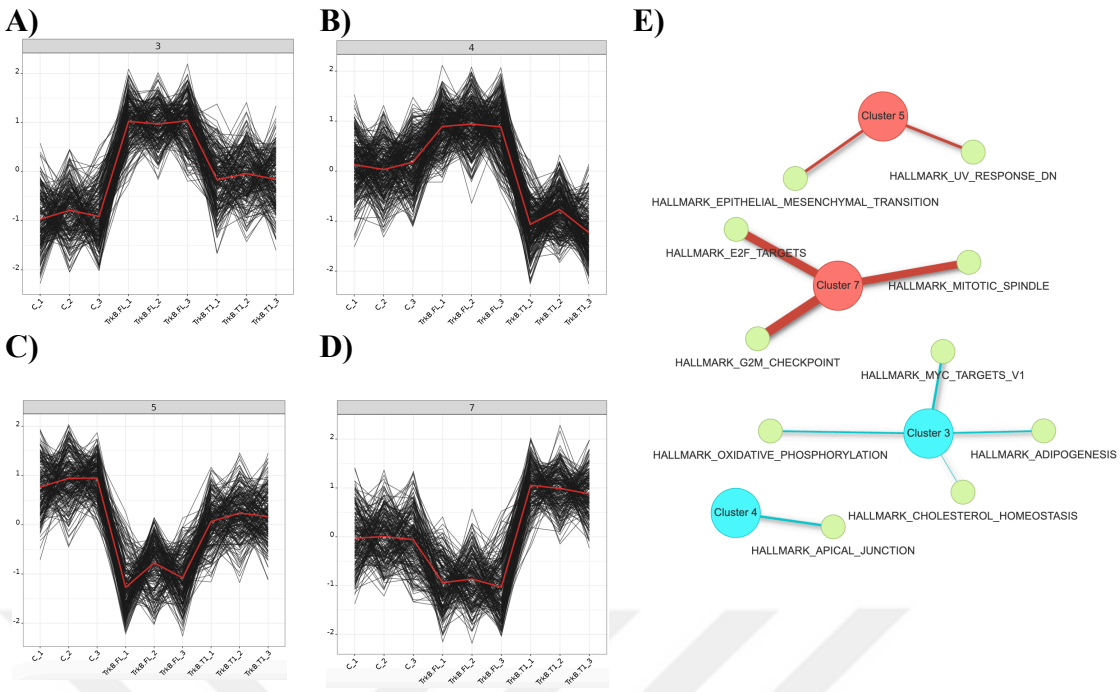


Figure 11. Analysis of clusters 3, 4, 5 and 7. A-B-C-D) Expression pattern of clusters. E) The network visualization of enrichments of genes in each cluster using Hallmark gene-set collection of MSigDB. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between TrkB.T1 and TrkB.FL samples.

3.1.3 Essential Gene Analysis of Demo1 Dataset

Next, the essential gene analysis was performed in each cluster described above. In the table 14, the mean efficacy and selectivity values along with standard deviations were shown. This table also included the number of essential genes and their percentage. Lastly, it contained the results of Fisher's exact tests which helped determine whether the number of essential genes in each cluster differed from those in all clusters in a nonrandom manner. The highest percentage of essential genes was found in cluster 1, which had the highest odds ratio and significance, while the lowest percentage of essential genes and odds ratio were found in cluster 6 (Table 14).

Table 14. Essentiality information of clusters for Demo 1 dataset. SD: standard deviation; OR: Odds Ratio; FET: Fisher's Exact Test; p.val: p-value.

Clusters	Mean of Efficacy	SD of Efficacy	Mean of Selectivity	SD of Selectivity	Number of Essential Gene	Total Number of Genes	Percentage of Essential Genes	OR of FET	p.value of FET
Cluster 1	-1	0.44	0.14	0.22	123	285	43.16	1.42	0.01
Cluster 2	-1.05	0.47	0.07	0.21	138	354	38.98	1.19	0.15
Cluster 3	-1.05	0.52	0.03	0.21	79	184	42.93	1.4	0.04
Cluster 4	-0.86	0.41	0.05	0.19	86	221	38.91	1.19	0.24
Cluster 5	-0.91	0.43	0.14	0.29	59	180	32.78	0.91	0.62
Cluster 6	-0.72	0.35	0.11	0.27	65	363	17.91	0.41	0
Cluster 7	-1	0.48	0.1	0.23	70	168	41.67	1.33	0.08
Cluster 8	-0.78	0.42	0.06	0.2	93	288	32.29	0.89	0.43

For a more detailed examination of the cluster essentiality, scatter plots were drawn for log2 of average CPM on the x-axis and CV of CPM values on the y-axis (Figure 12). Log2 of range CPM values were indicated as the size of dots, representing genes in the cluster. In cluster 1, most of the upregulated genes had higher expression values but lower coefficient of variations and the non-essential genes had significantly higher median CVs compared to essential genes (Figure 12A, B). Cluster 1's mirror, cluster 8, also displayed similar results on a scatter plot with its downregulated genes (Figure 12C, D). Moreover, cluster 2 also contained a significantly high number of essential genes with low CVs while having higher expression levels (Figure 12E, F). In its mirror, cluster 6, genes with higher CVs exhibited low to moderate expression levels, and median CVs were higher in the non-essential genes of cluster 6 (Figure 12G, H). In general, it was observed that non-essential genes had a higher CV, supported by Mann Whitney U test and lower average log2 CPMs. Also, by demonstrating on the plots, the log2 range of CPM values which showed the log2 difference between maximum and minimum, were lower in these non-essential genes, provided a novel look at differential expression analysis of datasets with more than two groups. This kind of plotting provides an alternative to MA plots, which will be focused on further in the discussion. Interestingly, the essential genes had higher log2 range CPMs and log2 average CPMs, yet lower CVs, which were normalized to mean expression.

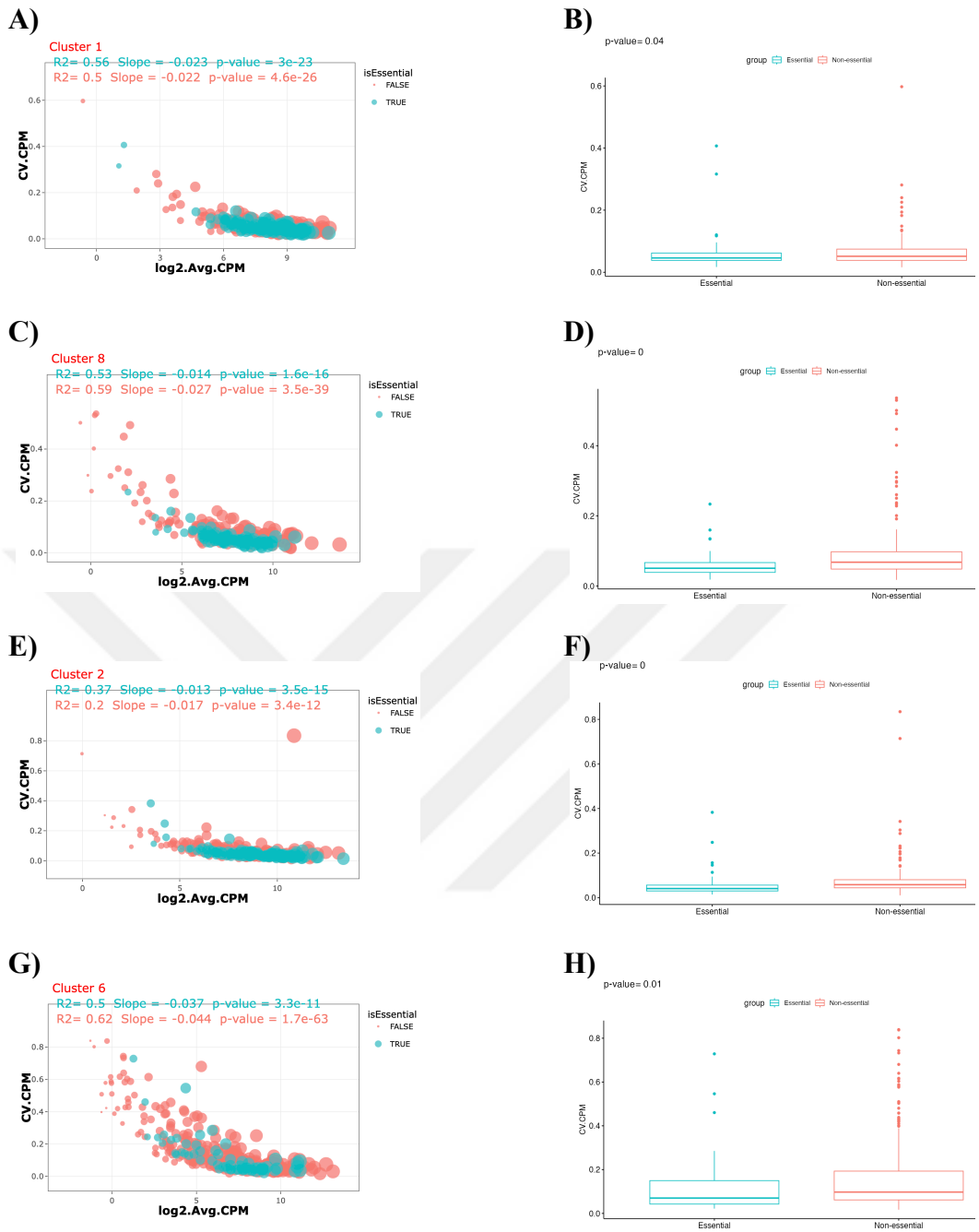


Figure 12. Analysis of cluster 1 and cluster 2 along with their mirror clusters 8 and 6, respectively. Each graph displays the log2 of average CPM values on the x-axis and CV of CPM values on the y-axis of genes of cluster 1 (A), cluster 8 (C), cluster 2 (E), and cluster 6 (G). The size of each dot displays a gene's log2 of range CPM value. The median CV for both essential and non-essential genes is displayed using the Mann Whitney U test in cluster 1 (B), cluster 8 (D), cluster 2 (F), and cluster 6 (H).

From these plots, CCND2, TNFSF10 and AK4 were identified in cluster 6, cluster 1 and cluster 2, respectively, as the essential genes with the highest variation. The CV, log2 average CPM and log2 range CPM of these genes were shown in the table 15 below.

Table 15. CV, log2 average CPM and log2 range CPM values of selected genes.

Genes	**CV**	**log2 average CPM**	**log2 range CPM**
CCND2	0.72	1.26	2.34
TNFSF10	0.40	1.28	1.46
AK4	0.38	3.50	3.69

3.1.4 Analysis of Correlation Between mRNA-Protein Levels of Demo 1 Dataset

Cluster 6 was further investigated using the DepMap data found in the DepMap: mRNA~protein module under the Cluster Prioritization tab of CAP-RNAseq (Figure 6) to find the correlation between mRNA-protein expression values of genes in glioblastoma disease. It was observed that there was no strong correlation, but the essential genes were located below the regression line, indicating they had relatively lower correlation coefficients when compared to non-essential genes. Notably, CCND2 gene exhibited the lowest RNA and protein expression values in DepMap data for glioblastoma (Figure 13).

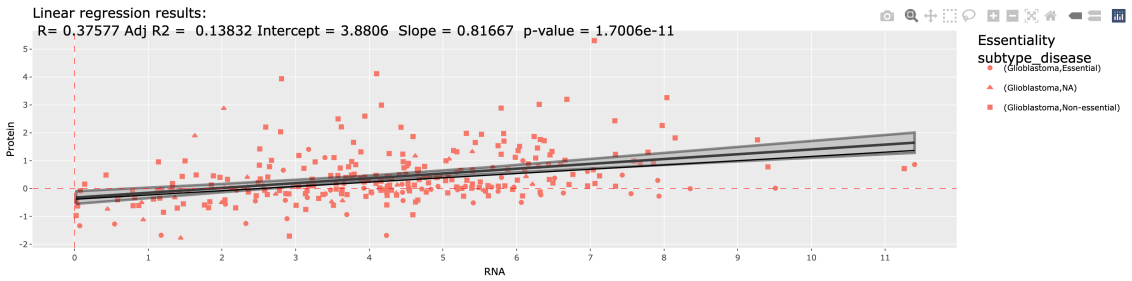


Table of robust regression fitted values along with bisquare weights using rlm function

Search

subtype_disease	gene_name	lineage	RNA	Protein	Essentiality	↑ fitted.values	weight
Glioblastoma	ZNF423	central_nervous_system	0.0215	-0.9695	Non-essential	-0.3827	0.91903
Glioblastoma	ST8SIA5	central_nervous_system	0.0286	-0.4562	Non-essential	-0.3816	0.99866
Glioblastoma	LAMP5	central_nervous_system	0.0424	-0.6245	Non-essential	-0.3795	0.98563
Glioblastoma	CCND2	central_nervous_system	0.0704	-1.3363	Essential	-0.3751	0.79049

Figure 13. Correlation result between mRNA-protein expression values of genes of cluster 6 in glioblastoma disease.

3.1.5 Gene Prioritization for Demo1 Dataset

Using the DESeq2 method, further analysis was performed on the genes of the clusters visualized in Table 16 and Figure 14 to examine differential gene expression for the three prioritized genes with high variation between groups. In the comparison of TrkB.T1 vs control groups, CCND2 was highly downregulated in TrkB.T1 samples while other genes (TNFSF10 and AK4) showed increased expressions. Based on the efficacy values, the most essential gene was CCND2, and it was also a more selective gene. The r value, which indicated the correlation between mRNA-protein values in DepMap data using all lineages, was also found to be highest for CCND2 (Table 16).

Table 16. Differential gene expression analysis results comparing TrkB.T1 and Control Samples using DESeq2 method for the identified genes.

	Cluster 6	**Cluster 1**	**Cluster 2**
Genes	CCND2	TNFSF10	AK4
Correlation with cluster centroid	0.925	0.824	0.954
TrkB.T1 vs Control logFC	-2.41	1.15	1.25
TrkB.T1 vs Control baseMean	28.13	28.65	133.13
TrkB.T1 vs Control p.value	0	0	0
TrkB.T1 vs Control p.adj.value	0	0	0
Efficacy	-1.153	-0.694	-0.449
Selectivity	0.68	0.411	-0.061
isEssential	TRUE	TRUE	TRUE
r for mRNA~protein	0.85	0.6458	0.7797
R2	0.7225	0.41704	0.60797
Adj.R2	0.71556	0.40947	0.60691
Intercept	-0.97794	-1.6638	-2.3742
Slope	0.61084	0.49243	0.61806
pval	1.072E-09	1.309E-07	4.92399e-77

The expression of the CCND2 gene was found to decrease not only in TrkB.FL samples compared to control but also showed a more pronounced decrease in TrkB.T1 samples (Figure 14A). However, TNFSF10 and AK4 showed an opposite pattern to CCND2, with increased expression levels in both TrkB.FL and TrkB.T1 samples (Figure 14C, E), indicating a possible association with TrkB isoforms. Utilizing DepMap data, mRNA ~ Protein correlation was visualized for each gene across all lineages. CCND2 had the highest correlation but was found in fewer samples of DepMap data (Figure 14B),

suggesting that there is a potentially robust relationship between mRNA and protein expression. The lowest correlation among all three genes was found in TNSF10 but it was found in more samples than CCND2 (Figure 14D). On the other hand, correlation for AK4 was more than TNSF10 and was found in more samples in DepMap data than TNSF10 (Figure 14E).

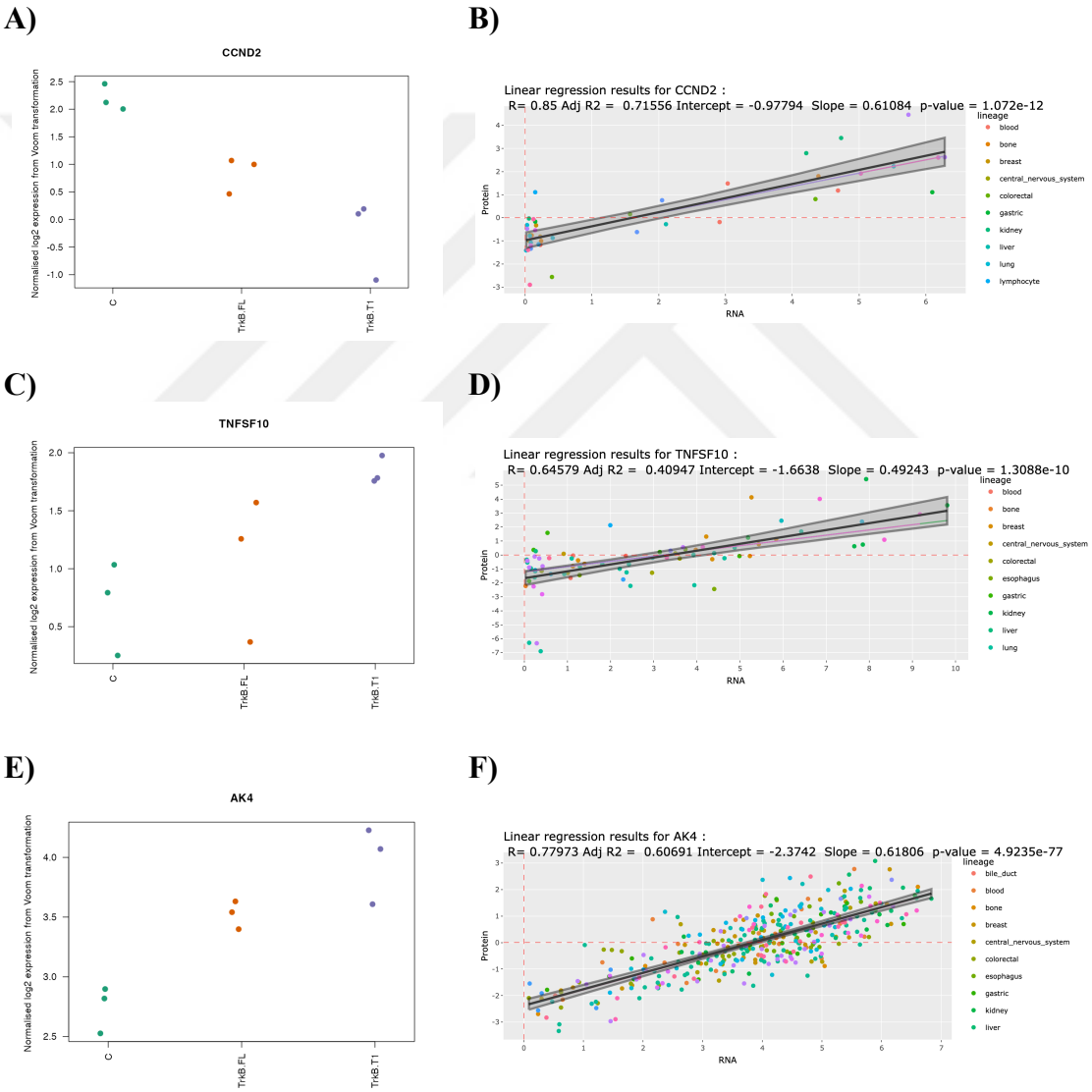


Figure 14. Expression pattern and mRNA ~Protein correlation of each gene. (A, C, D) Graphs showing the normalized log2 expression of genes in each sample. (B, D, F) mRNA ~Protein correlation of each gene in all lineages using DepMap data.

3.1.6 Protein Expression of Prioritized Genes in Demo1 Data Using HPA Module

In order to gain insight into the protein expression levels of these genes in various cancer and normal tissues, CAP-RNAseq's HPA analysis module was used. The results demonstrated that CCND2 was produced at low levels in neuronal cells and neuropil of the cerebral cortex, while its expression remained generally low across almost all types of cancer, including glioma (Figure 15A, B). However, TNFSF10 was found to be highly expressed in glioma patients, while no information regarding its expression in normal cells was found in the HPA database (Figure 15C). Moreover, AK4 was expressed at low to moderate levels in cancers, except testis cancer where its expression was notably high. While a small proportion of glioma patients expressed AK4 at moderate levels, its expression was also found at moderate levels in normal neuronal cells and glial cells of the cerebral cortex (Figure 15D, E).

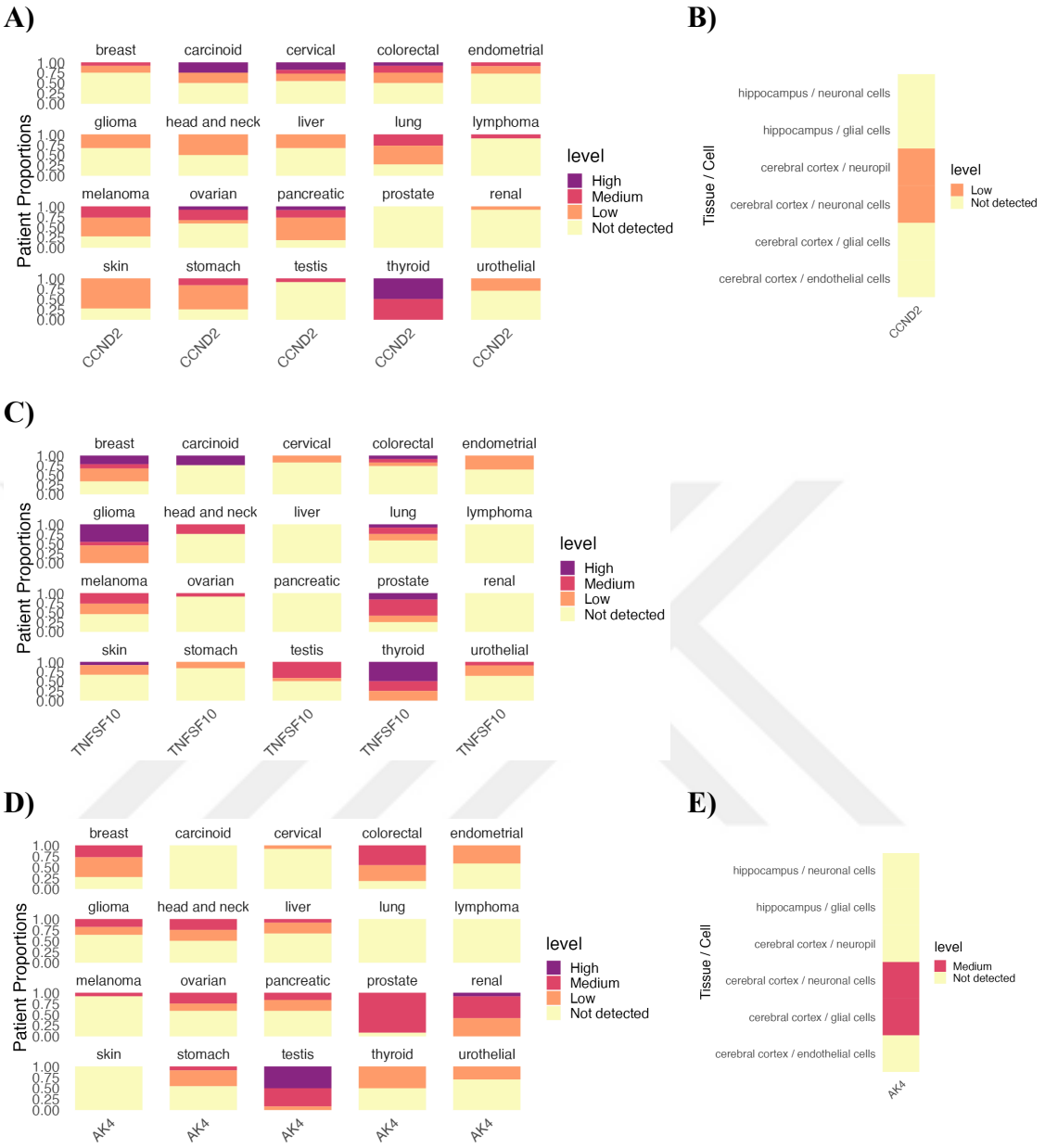


Figure 15. Protein expression levels of CCND2 (A, B), TNFSF10 (C) and AK4 (D, E) in all cancers and normal tissues using the HPA dataset.

Based on these findings, CCND2 was further prioritized, and I hypothesized that NTRK2 and CCND2 gene expression could be associated with each other in gliomas. The findings that CCND2 expression was significantly decreased in the TrkB.T1 overexpressing samples and CCND2 protein level was low in glioblastoma pointed to this potential relationship between NTRK2 and CCND2 genes in gliomas. I have then obtained using cbioportal.org the expression values of NTRK2 and CCND2 in TCGA LGG (Brain Lower Grade Glioma (TCGA, PanCancer Atlas) containing 514 samples with mRNA data) and GBM data (Glioblastoma Multiforme (TCGA, PanCancer Atlas) containing 592 samples with mRNA data) and plotted the expression values and the correlation using Pearson method (Figure 16). I have found that there was a significant negative correlation between NTRK2 and CCND2 in LGG but not significant in GBM. It is known that whole gene NTRK2 (TrkB.FL) expression is similar in LGG, GBM and normal tissues, while TrkB.T1 is known to be the predominant isoform in glioblastoma samples CAP-RNAseq revealed that overexpression of NTRK2 lead to a decreased expression of CCND2, which aligned with the low protein levels observed in glioma patients. These insights improve our knowledge of the molecular processes underlying gliomas and point to possible treatment targets that require more investigation.

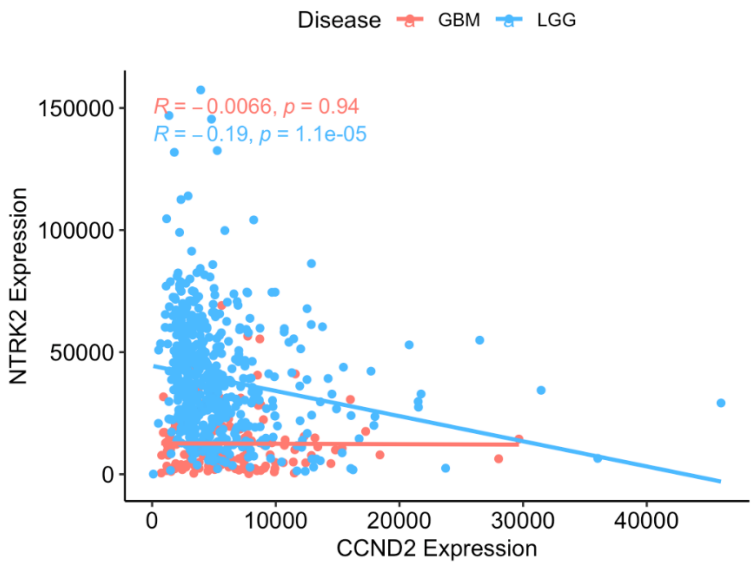


Figure 16. The correlation of NTRK2 mRNA expression with CCND2 expression in LGG and GBM patients.

3.2 Analysis of Senescence and "Essential Genes"

The second demo dataset, GSE190998, was generated in research where it was discovered that TrkB has a role in preserving the survival of senescent cells, WI-38 fibroblasts. After applying vst+ANOVA, 12 clusters were obtained using k-means clustering on the 4,412 remaining genes. Clusters 2, 3, 5, 6, and 8 exhibited an elevated expression pattern in TrkB silencing samples compared to senescent control samples (Figure 17). Among them, cluster 2 had the majority of genes with differential expression based on a threefold logFC value of 0.5 (Table 17). The number of differentially expressed genes with the same threshold was highest in cluster 11 when comparing TrkB silencing samples or senescent control samples with proliferating control samples.

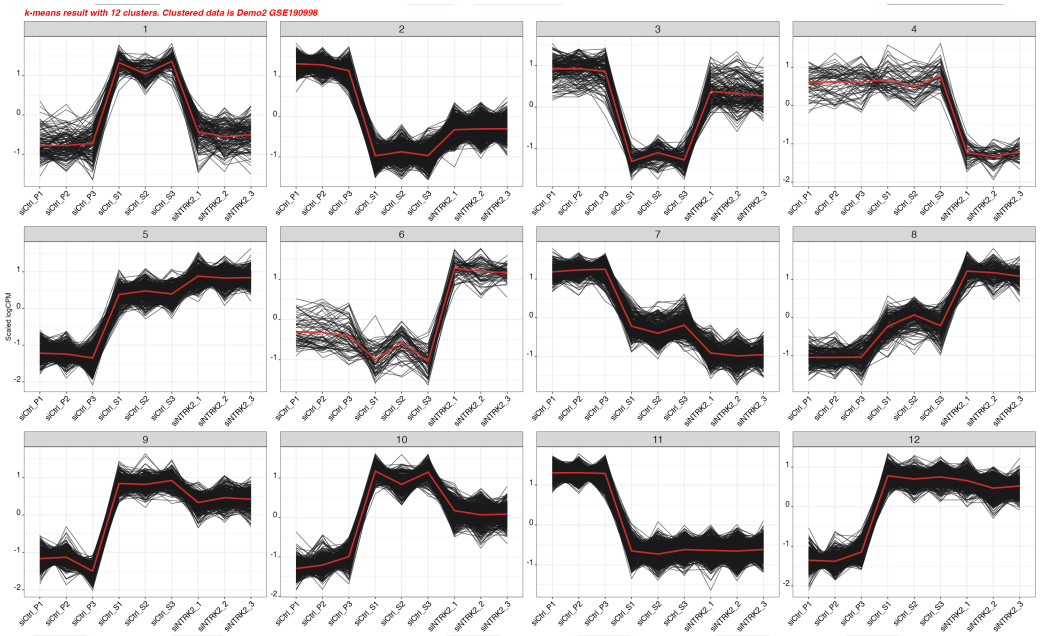
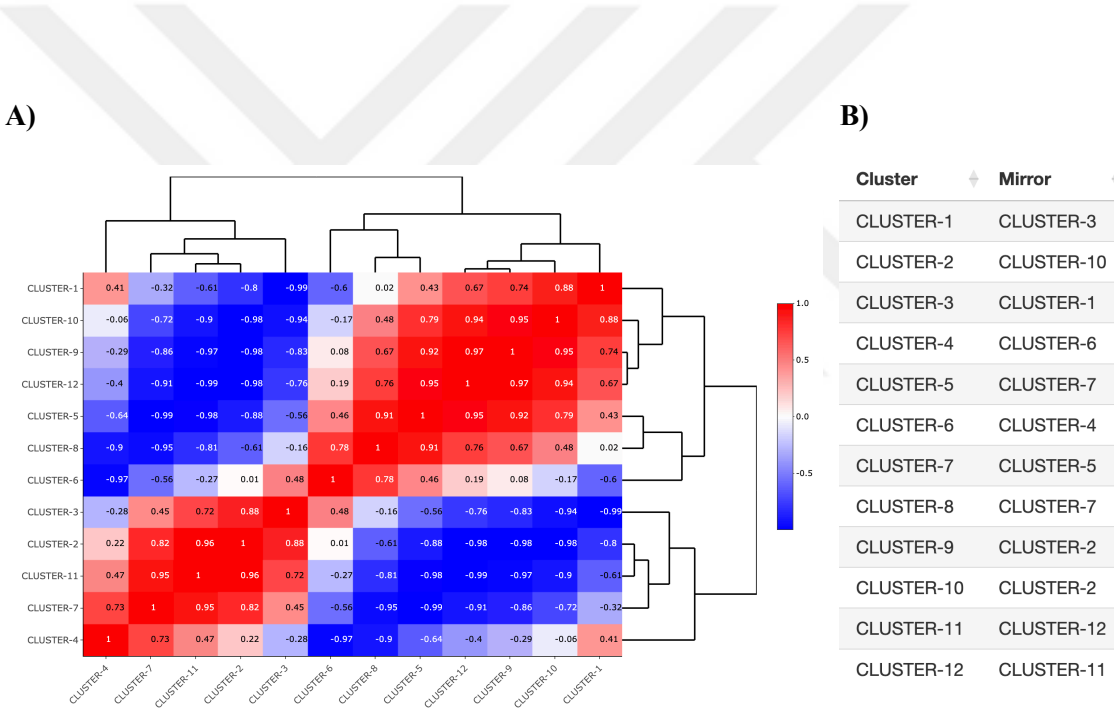


Figure 17. Expression patterns in GSE190998 presented by k -means clustering. The cluster centroid pattern is shown by red lines.

Table 17. The number of differentially expressed genes in each comparison using an absolute log-fold change threshold of 0.5.

	siNTRK2 vs siCtrl_S		siNTRK2 vs siCtrl_P		siCtrl_S vs siCtrl_P	
	DEGs with p-value<=0.05	All DEGs	DEGs with p-value<=0.05	All DEGs	DEGs with p-value<=0.05	All DEGs
Cluster 1	37	37	2	2	50	50
Cluster 2	112	117	442	442	583	583
Cluster 3	40	40	8	9	73	73
Cluster 4	30	30	24	24	1	1
Cluster 5	19	24	365	365	297	298
Cluster 6	33	33	16	16	1	1
Cluster 7	25	31	295	295	158	159
Cluster 8	53	55	132	132	34	35
Cluster 9	34	45	223	223	285	285
Cluster 10	70	70	124	126	309	309
Cluster 11	3	3	710	710	700	700
Cluster 12	10	15	445	445	484	484

Then, the dissimilarity heatmap was displayed to obtain mirror clusters (Figure 18). In the table of essentiality metrics, cluster 2 showed a significantly higher odds ratio compared to other clusters, indicating a greater proportion of essential genes (Table 18). Cluster 11, which contained the most essential genes, also showed the second-highest odds ratio, and its mirror cluster, 12, has the second-lowest odds ratio. Cluster 5, on the other hand, had the lowest odds ratio and contained a high number of non-essential genes compared to the number of essential genes. Moreover, cluster 2 and its mirror cluster 10 had the lowest mean of selectivity values.



Cluster	Mirror
CLUSTER-1	CLUSTER-3
CLUSTER-2	CLUSTER-10
CLUSTER-3	CLUSTER-1
CLUSTER-4	CLUSTER-6
CLUSTER-5	CLUSTER-7
CLUSTER-6	CLUSTER-4
CLUSTER-7	CLUSTER-5
CLUSTER-8	CLUSTER-7
CLUSTER-9	CLUSTER-2
CLUSTER-10	CLUSTER-2
CLUSTER-11	CLUSTER-12
CLUSTER-12	CLUSTER-11

Figure 18. Dissimilarity heatmap and a table for mirror clusters in Demo 2 data. A) Heatmap shows the correlation values, in which negative values (blue) indicate the lowest similarities and positive values (red) indicate the highest similarities. B) The mirror clusters are shown in a table to be easily identified.

Table 18. Essentiality information of clusters for Demo 2 dataset. SD: standard deviation; OR: Odds Ratio; FET: Fisher's Exact Test; p.val: p-value.

Clusters	Mean of Efficacy	SD of Efficacy	Mean of Selectivity	SD of Selectivity	Number of Essential Genes	Total Number of Genes	Percentage of Essential Genes	OR of FET	p.value of FET
Cluster 1	-0.68	0.34	0.1	0.15	35	106	33.02	1	1
Cluster 2	-0.98	0.43	0.07	0.19	291	624	46.63	1.78	0
Cluster 3	-0.84	0.41	0.11	0.26	44	137	32.12	0.96	0.93
Cluster 4	-0.94	0.47	0.23	0.28	12	47	25.53	0.7	0.35
Cluster 5	-0.75	0.36	0.11	0.22	95	398	23.87	0.64	0
Cluster 6	-0.84	0.34	0.19	0.25	19	54	35.19	1.1	0.77
Cluster 7	-0.86	0.43	0.11	0.23	94	311	30.23	0.88	0.35
Cluster 8	-0.75	0.31	0.15	0.32	41	172	23.84	0.64	0.01
Cluster 9	-0.75	0.31	0.1	0.2	64	293	21.84	0.57	0
Cluster 10	-0.77	0.4	0.08	0.19	108	323	33.44	1.02	0.9
Cluster 11	-0.94	0.46	0.09	0.29	317	784	40.43	1.38	0
Cluster 12	-0.74	0.4	0.1	0.21	119	507	23.47	0.62	0

3.2.1 Analysis of Specific Clusters in Terms of Essentiality

Cluster 2 and 5 were selected for detailed examination. The scatter plots were drawn with log2 of average CPM on the x-axis and CV of CPM values on the y-axis (Figure 19). Log2 of range CPM values were indicated as the size of dots. Cluster 2 contained a large number of essential genes with lower gene expression in senescence control (siCtrl_S) samples compared to proliferating control (siCtrl_P) and slightly higher expression in siNTRK2 samples compared to siCtrl_S (Figure 19A). Among these genes, those with

low to moderate expression levels had larger CVs. The Mann-Whitney U-test also indicated that essential genes exhibited a greater coefficient of variation of CPM in comparison to non-essential genes within cluster 2 (Figure 19B). On the other hand, genes in cluster 5 showed increased expression in both siCtrl_S and siNTRK2 compared to siCtrl_P samples, and only a few of the essential genes had a higher coefficient of variation of CPM (Figure 19C). The Mann-Whitney U-test also revealed that non-essential genes displayed the largest coefficient of variation of CPM (Figure 19D). CCND2 was also found as the essential gene in cluster 5 with the largest coefficient of variation of CPM and log2 of range CPM.

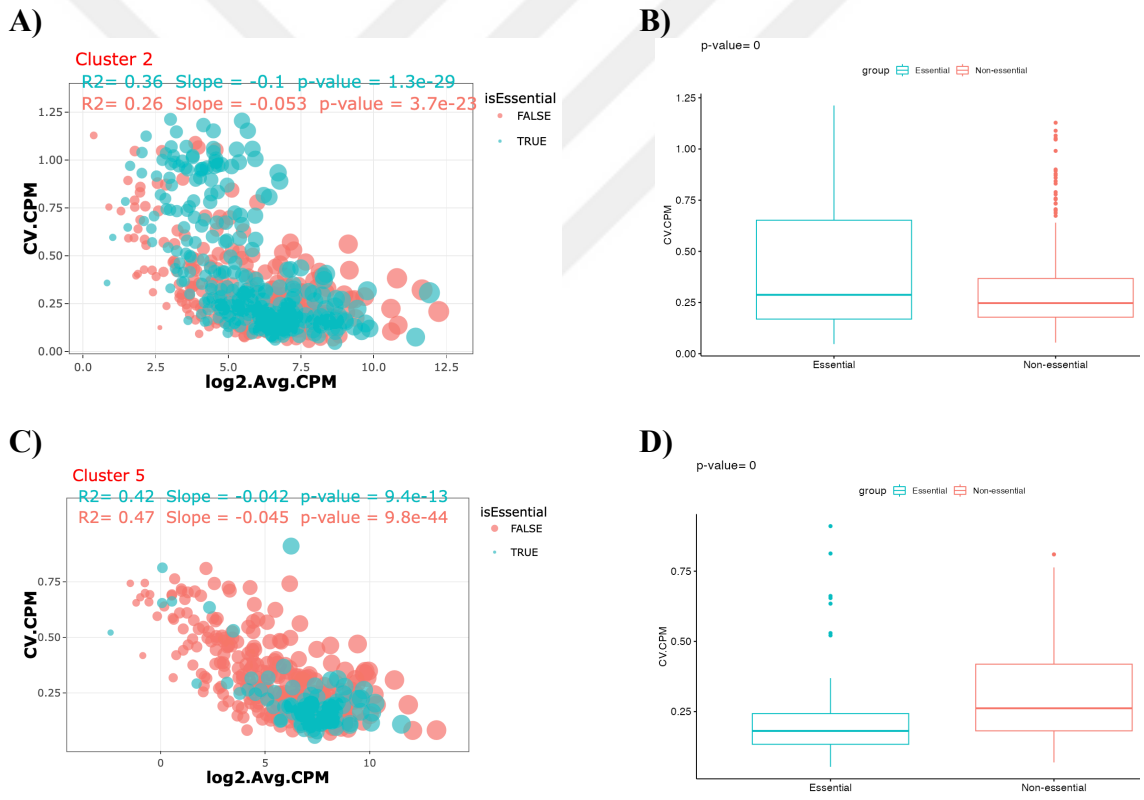


Figure 19. Examination of cluster 2 and cluster 5. Each graph presents the log2 of average CPM values on the x-axis and CV of CPM values on the y-axis for the genes of cluster 2 (A) and cluster 5 (C). The log2 of range CPM values are displayed by the size of the dots.

The median CV is displayed using the Mann Whitney U test for both essential and non-essential genesin cluster 2 (B) and cluster 5 (D).

3.2.2 Gene Set Annotation for Essential and Non-Essential Genes

MSigDB analysis was performed for essential and non-essential genes, separately (Figure 20). Cluster 10 was the only cluster that contained upregulated essential genes in siCtrl_S samples compared to proliferating control siRNA samples. Other clusters, which displayed enriched terms with q-values less than 0.05, contained the downregulated essential genes. Cluster 10 exhibited enrichment of essential genes in two hallmark terms: the peroxisome, which includes the genes encoding components of peroxisome, and UV response which contains genes upregulated in response to ultraviolet (UV) radiation. Conversely, downregulated essential genes in other clusters were enriched in terms such as MYC targets, DNA repair, G2M checkpoints, MTORC1 signaling, cholesterol homeostasis, mitotic spindle, and E2F targets. Interestingly, essential genes in cluster 11, downregulated in both siCtrl_S and siNTRK2 samples, were responsible for DNA repair and G2M checkpoint. Moreover, non-essential genes, which showed decreased expression in clusters 3, 6 and 11, and increased expression in clusters 5 and 8, shared the EMT term. Specifically, upregulated non-essential genes were enriched in coagulation, xenobiotic metabolism and KRAS signaling in clusters 8, 10 and 12, respectively. On the other hand, downregulated non-essential genes were also responsible for G2M checkpoint and mitotic spindle assembly in cluster 2, and interferon alpha and gamma responses in cluster 7.

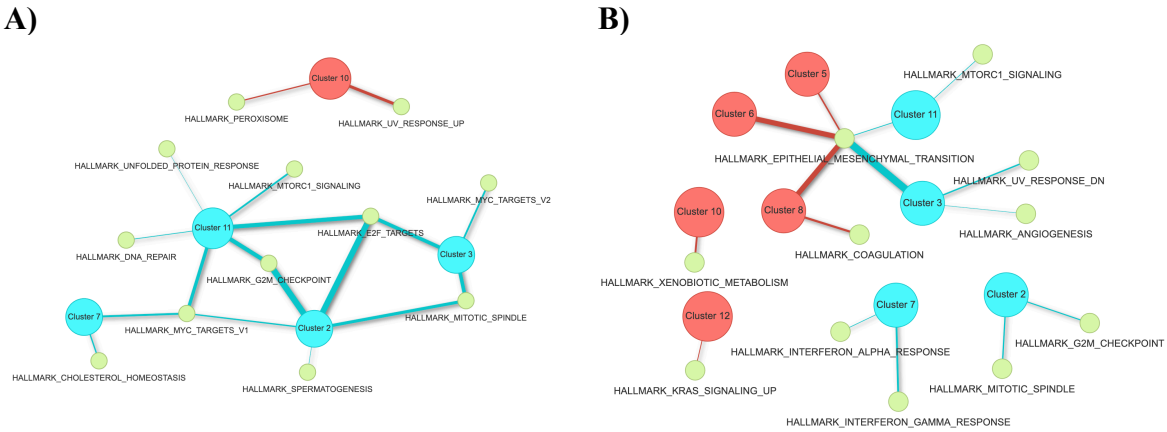


Figure 20. MSigDB networks for Hallmark gene set collection. Network visualization with enriched terms having q-values less than 0.05 for essential genes (A) and non-essential genes (B) in the clusters. The number of genes that overlap with each gene set collection was shown by edge widths. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between siCtrl_S and siCtrl_P samples.

3.2.3 Specific Analysis of Cluster 2

DGE analysis was performed to compare the siNTRK2 treatment sample with the siCtrl_S control sample, in order to thoroughly examine cluster 2, which displayed the highest percentage of essential genes. Subsequently, an MSigDB analysis was conducted on 117 genes from this cluster with logFC values greater than 0.5, using REACTOME and GO terms. The analysis revealed terms related to cell division (Table 19), which means that the expression of genes involved in cell division was reduced in senescent cells compared to proliferating cells, whereas inhibition of NTRK2 in senescent cells caused an increase in their expression.

Table 19. First 20 results of the enrichment analysis of upregulated genes in cluster 2 for the REACTOME and GO terms.

ID	GeneRatio	BgRatio	p value	p adjust	q value
GOBP_CHROMOSOME_SEGREGATION	47/106	350/18448	0	0	0
GOBP_NUCLEAR_CHROMOSOME_SEGREGATION	43/106	289/18448	0	0	0
GOBP_SISTER_CHROMATID_SEGREGATION	39/106	207/18448	0	0	0
GOBP_MITOTIC_NUCLEAR_DIVISION	42/106	296/18448	0	0	0
GOBP_MITOTIC_SISTER_CHROMATID_SEGREGATION	36/106	173/18448	0	0	0
GOBP_ORGANELLE_FISSION	48/106	496/18448	0	0	0
REACTOME_CELL_CYCLE_CHECKPOINTS	35/106	292/18448	0	0	0
REACTOME_MITOTIC_PROMETAPHASE	28/106	203/18448	0	0	0
REACTOME_M_PHASE	33/106	415/18448	0	0	0
REACTOME_RESOLUTION_OF_SISTER_CHROMATID_COHESION	23/106	126/18448	0	0	0
REACTOME_MITOTIC_SPINDLE_CHECKPOINT	21/106	111/18448	0	0	0
GOBP_MICROTUBULE_CYTOSKELETON_ORGANIZATION_INVOLVED_IN_MITOSIS	22/106	153/18448	0	0	0
REACTOME_MITOTIC_METAPHASE_AND_ANAPHASE	25/106	236/18448	0	0	0
GOBP_SPINDLE_ORGANIZATION	24/106	208/18448	0	0	0
GOBP_MITOTIC_SPINDLE_ORGANIZATION	20/106	126/18448	0	0	0
GOBP_REGULATION_OF_CHROMOSOME_SEGREGATION	18/106	91/18448	0	0	0
REACTOME_RHO_GTPASES_ACTIVATE_FORMINS	20/106	140/18448	0	0	0
GOBP_CHROMOSOME_SEPARATION	18/106	97/18448	0	0	0
GOBP_METAPHASE_ANAPHASE_TRANSITION_OF_CELL_CYCLE	16/106	65/18448	0	0	0
GOBP_MITOTIC_CELL_CYCLE_PHASE_TRANSITION	28/106	440/18448	0	0	0

The CCND2 gene was first identified in the demo 1 dataset, which sparked interest and further investigation in Cluster 5 of Demo 2 dataset. Its expression was found to be higher in senescent cells compared to proliferating cells, and even more so with the inhibition of

NTRK2 (Figure 21). The CAP-RNAseq application revealed a new association between CCND2 and NTRK2 genes. The inhibition of NTRK2 in senescent cells led to a significant increase in CCND2 expression, while overexpression of NTRK2 had the opposite effect, as observed in the demo 1 dataset.



Figure 21. Expression plots of the CCND2 gene in each sample of Demo 1 data (A) and Demo 2 data (B).

3.2.4 The Effect of NTRK2 Inhibition in Senescent Cells

To check the effect of NTRK2 inhibition, I performed the MSigDB analysis once again, but this time using all genes and coloring the nodes according to the average logCPM differences between siNTRK2 and siCtrl_S (Figure 22). Apoptosis was revealed for clusters 1 and 10, both of which showed upregulation in senescent cells but less upregulation in siNTRK2 cells compared to proliferating cells. Most of the genes enriched in apoptosis in cluster 10 were the genes that decrease the apoptosis when they are expressed. Moreover, genes in cluster 10 were also enriched in the genes that are upregulated in response to UV while genes in cluster 3 which included the upregulated genes by siNTRK2 compared to siCtrl_S were enriched in the genes that are downregulated in response to UV.

Moreover, EMT was appeared in the clusters 3, 5, 6, and 8 which shows upregulation by siNTRK2 compared to siCtrl_S. The cluster 2 showed the enrichment in the terms related to cell cycle progression and division. In the cluster 7 which contained genes downregulated more by siNTRK2, responses to interferon alpha and gamma terms were revealed.



Figure 22. MSigDB networks for Hallmark terms having q-values less than 0.05 for all genes. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between siNTRK2 and siCtrl_S samples.

3.2.5 Cross-Dataset Mapping

Finally, Demo 2 dataset's genes from cluster 5 were mapped to the Demo 1 dataset to uncover the clusters that exhibit an enrichment of these genes (Figure 23). The genes of cluster 5 displayed a substantial increase in their expression levels within senescent cells, and this effect was further amplified when NTRK2 was inhibited.

As a result of this analysis, two tables were generated. The first table shows the clusters to which each gene belongs, while the second table shows the frequency of appearance

for each cluster. Based on the results in the tables, it was revealed that most of these genes were predominantly found in cluster 6, which includes the CCND2 gene, and cluster 8 of the Demo 1 dataset. The genes in these clusters displayed the expression patterns that were the inverse of the expression pattern of cluster 5 in the demo 2 dataset, which showed a decrease in the gene expression levels when NTRK2 was overexpressed. This observation demonstrates the complex interplay between NTRK2 and the gene expression patterns within these clusters.

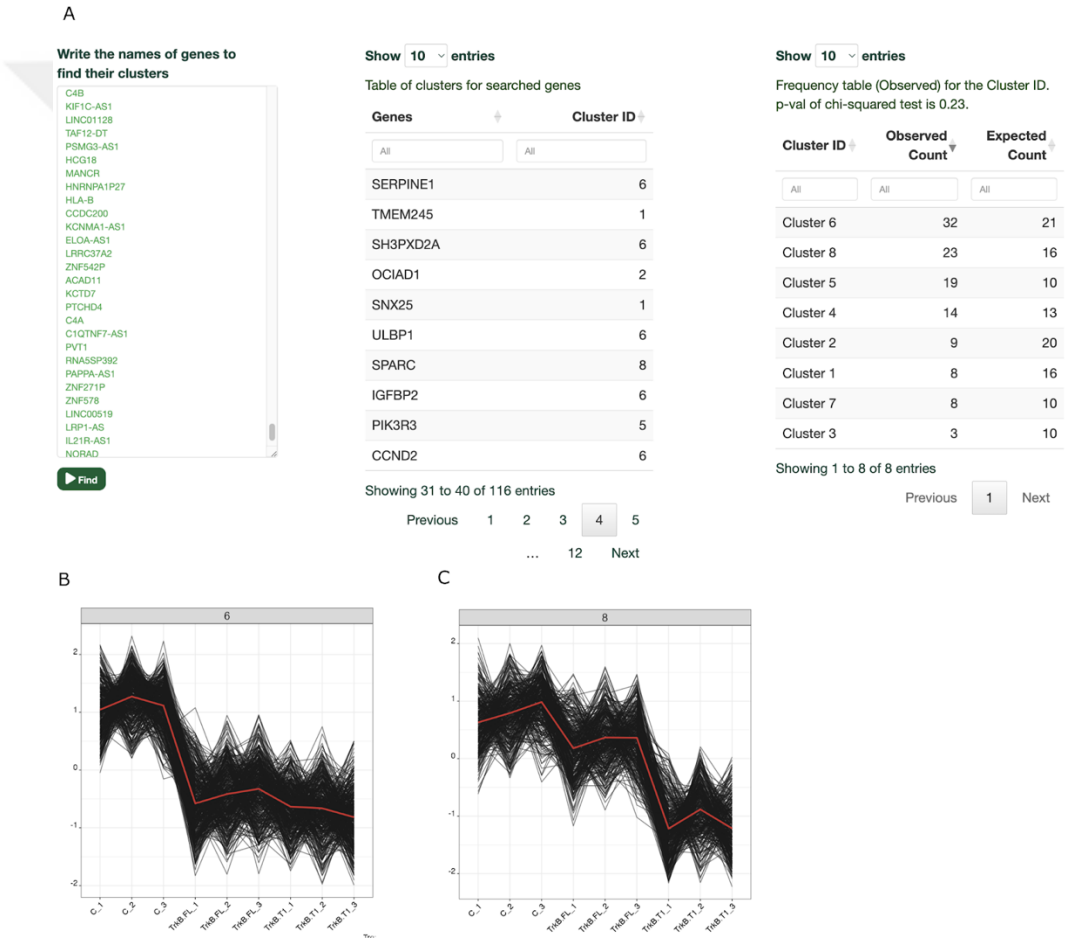


Figure 23. Cross-comparison of gene expression from Demo 1 and Demo 2 datasets. A) The Demo 2 dataset's genes from cluster 5 were searched and mapped to the Demo 1 analysis. B) Gene expression patterns within cluster 6 and C) within cluster 8 of the Demo 1 dataset.

3.3 Essential Genes and Recurrence Free Breast Cancer

Another demo dataset, GSE201085, was analyzed to prioritize novel biomarkers for potential use in diagnosing and treating breast cancer. This data was obtained from blood samples collected from a group of patients diagnosed with breast cancer. There are 53 samples in total, with 32 of them having received neoadjuvant chemotherapy (NAC) and the remaining 21 samples not having received this treatment. Of the patients who received NAC, 23 were classified as residual disease (RD). Out of these 23 patients, 7 were further categorized as RD with recurrence (RDR), indicating that their cancer had returned after treatment, and 16 as free of recurrence (RDnR), meaning that their cancer had not returned for at least three years. Nine were classified as pathologic complete response (pCR), indicating that their cancer was removed entirely.

To re-analyze this data, 4,846 genes filtered by vst+ ANOVA at p-value less than 0.05 were obtained and integrated into CAP-RNAseq. Eight clusters representing different gene expression profiles were generated based on group means and visualized as box plots, clearly representing their distribution and pattern (Figure 24). Among these clusters, it was observed that clusters 1, 6, 7, and 8 contained genes that had increased expression in the samples classified as pCR compared to those categorized as RD. This finding suggests that genes in these clusters could play a functional role in successfully eradicating cancer cells.

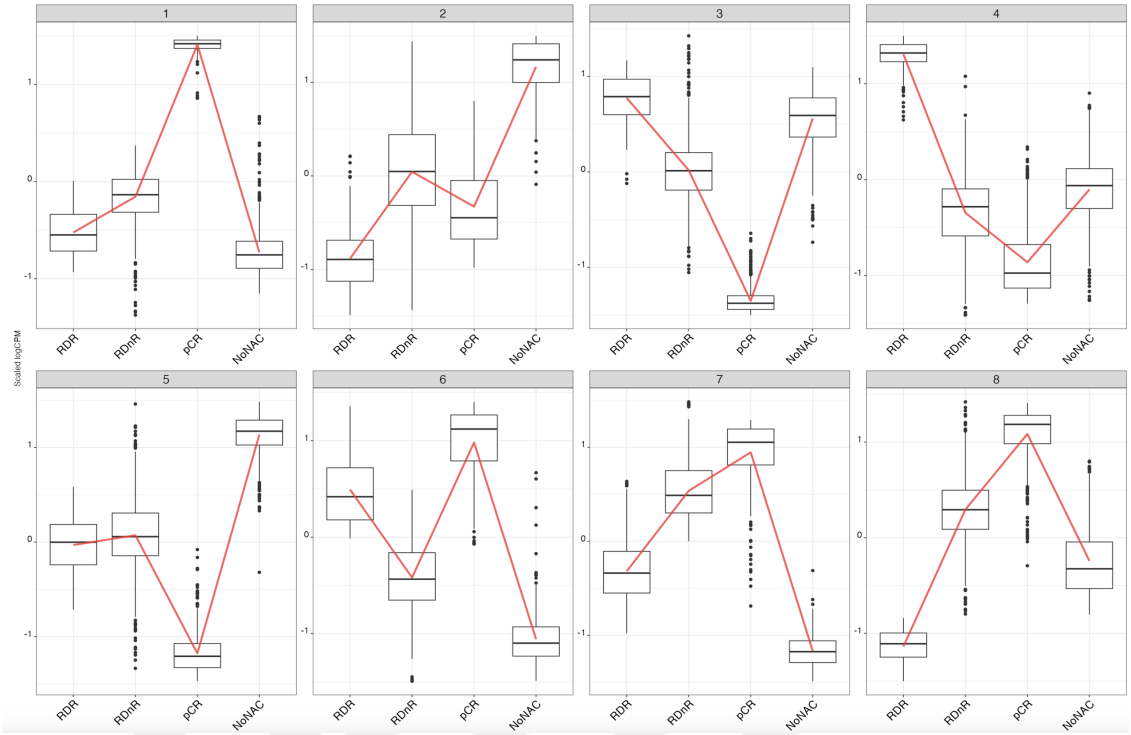


Figure 24. K-means clustering of GSE201085 data using group means values and presented as box plot. Red lines indicate the pattern of clusters.

3.3.1 Gene Set Annotation for All, Essential and Non-essential Genes

Additional analyses with MSigDB using hallmark gene sets were performed to gain further insight into the biological processes and pathways associated with these gene clusters. Significant pathways (q value <0.05) were identified only in clusters 1, 7, and 8 when all genes were included. Cluster 1, which exhibited the highest upregulation of genes in pCR samples compared to RDR samples was found to be enriched in several key pathways related to immune response, energy metabolism, cell growth, survival and inflammation, including interferon alpha and gamma response, complement, oxidative phosphorylation, IL6/JAK/STAT3 signaling, and MTORC1 signaling pathways (Figure

25). On the other hand, cluster 8 showed enrichment in TNFA signaling, apoptosis, which signifies the importance in inflammatory responses and programmed cell death. Other terms revealed in cluster 8 indicated roles of genes in cell cycle regulation with the enrichment in P53 pathway, cell differentiation processes with the enrichment in NOTCH signaling, and immune regulation and cellular responses to oxygen deprivation with the enrichment in IL2/STAT5 signaling and hypoxia, respectively. In addition, cluster 7 shared common pathways with cluster 8, including TNFA signaling and apoptosis.

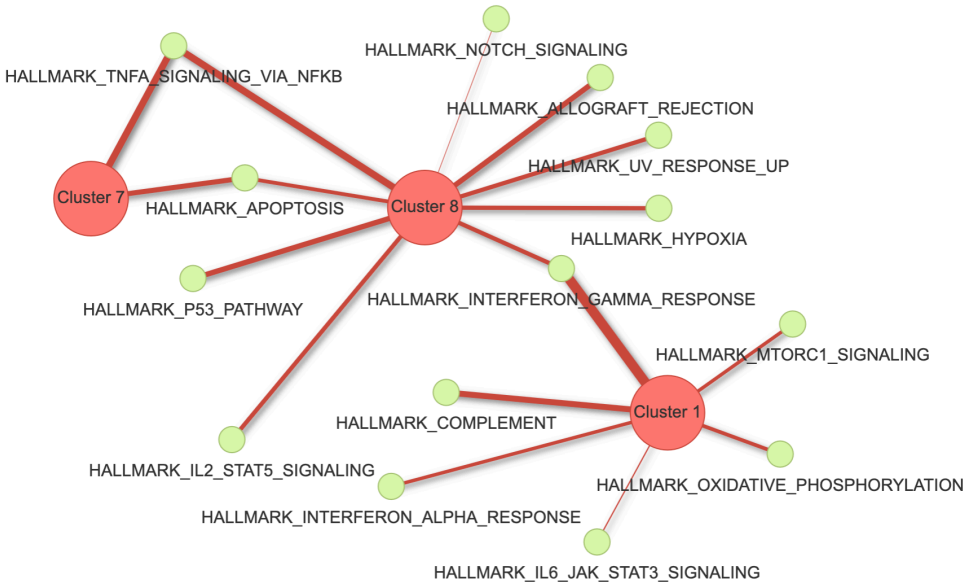


Figure 25. MSigDB analysis of clusters for Hallmark gene set collection with network visualization of enriched terms having q-values less than 0.05 for all genes. Red nodes indicate the increase in gene expression in pCR samples compared to RDR samples.

When the enrichment of essential and non-essential genes was analyzed separately, it revealed that essential genes in cluster 4 where a significant decrease in gene expression was observed in pCR samples compared to RDR samples were associated with G2M checkpoint and late estrogen response, whereas non-essential genes showed enrichment

in only IL2/STAT5 signaling (Figure 26A). Similarly, essential genes in cluster 5, also displaying a decrease, were enriched in E2F targets. However, essential genes in cluster 1, which exhibited an increase in pCR samples compared to RDR samples were linked to pathways including MTORC1 signaling, hypoxia, oxidative phosphorylation, MYC targets and interferon gamma response. Essential genes in cluster 7 and 8 were enriched in common term, TNFA signaling, with cluster 7 also showing a specific enrichment in PI3K/AKT/MTOR signaling. On the other hand, the non-essential genes in the cluster 1 were enriched in interferon alpha and gamma response, IL6/JAK/STAT signaling and complement terms, with the last term also shared by cluster 7. Specifically, non-essential genes in cluster 7 were enriched in the apoptosis and TNFA signaling, while those in cluster 8 were linked to NOTCH signaling and allograft rejection (Figure 26B).

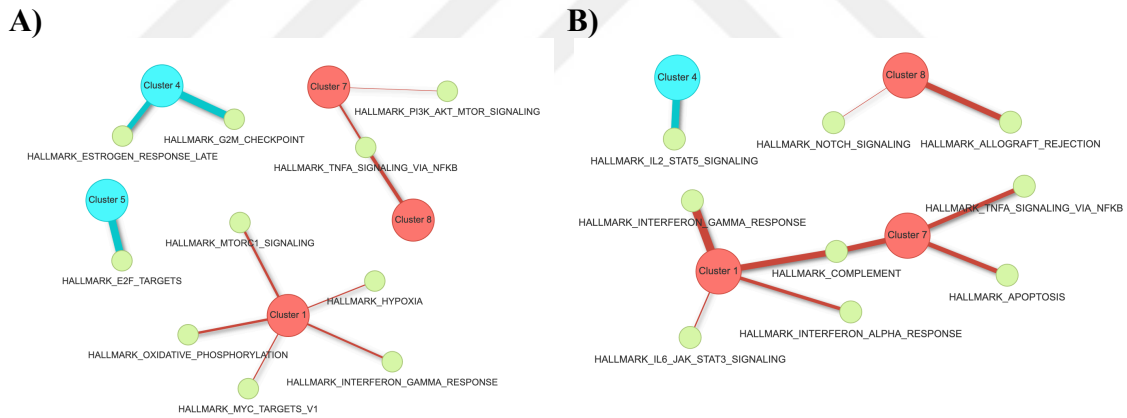


Figure 26. MSigDB networks for Hallmark gene set collection. Network visualization with enriched terms having q-values less than 0.1 for essential genes (A) and non-essential genes (B) in the clusters. Red nodes indicate the increase in gene expression in pCR samples compared to RDR samples, whereas blue node indicate a decrease.

3.3.2 Possible Biomarkers for the Recurrence

To find a biomarker for the recurrence in response to NAC, cluster 4 was further analyzed since its genes exhibited upregulation in RDR samples compared to all other samples. This suggests that the genes in cluster 4 may play a crucial role in tumor recurrence. Among these genes, the IL1R2 gene showed the most significant increase in expression in RDR samples compared to samples showing a complete pathological response (pCR) after NAC (Figure 27A). However, there was no well-established correlation between IL1R2 mRNA and protein in breast cancer (Figure 27B). Despite this, the protein level of IL1R2 was highly expressed in breast cancer and not detected in normal breast tissues, as observed in HPA analysis (Figure 27C-D). Moreover, another gene, ENPP5, also displayed a significant increase in both mRNA and protein expression levels in breast cancer (Figure 28A). The correlation between the mRNA and protein expression levels of ENPP5 was found to be high in breast cancer patients (Figure 28B). In the HPA analysis, ENPP5 also showed moderate protein expression in the majority of breast cancer patients, indicating its potential as a biomarker, while a quarter of patients had high expression of ENPP5, suggesting heterogeneity in its expression pattern among breast cancer patients (Figure 29C-D). Also, ENPP5 was not detected in normal breast tissue, further supporting its potential as a biomarker for breast cancer.

As a result of the analysis of cluster 4, the genes IL1R2 and ENPP5 were identified as potential biomarkers of breast cancer recurrence. In order to determine their clinical utility and potential for targeted therapies, these genes need to be studied further.

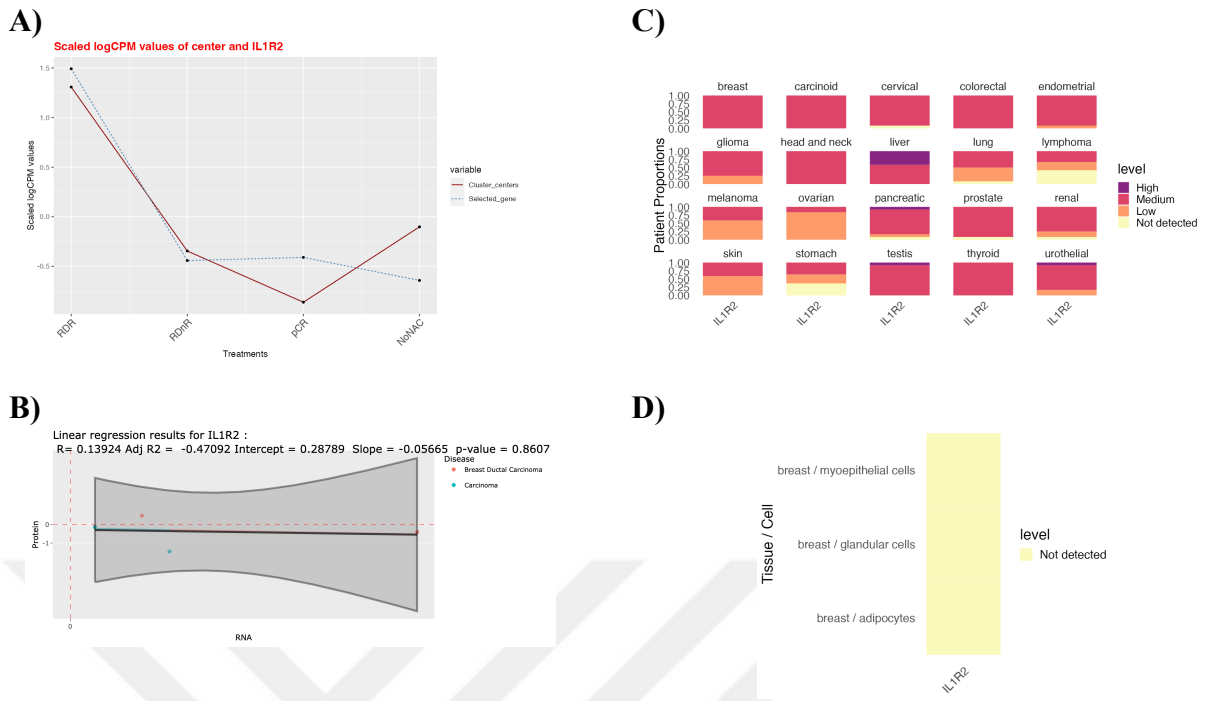


Figure 27. Analysis of IL1R2 gene. A) Plot of scaled logCPM values of IL1R2 gene against its cluster centroid. B) Correlation plot for mRNA and protein levels in breast cancer obtained from DepMap data. C) Protein levels of IL1R2 gene in all cancers by HPA. D) Protein level of IL1R2 gene in normal breast tissues.

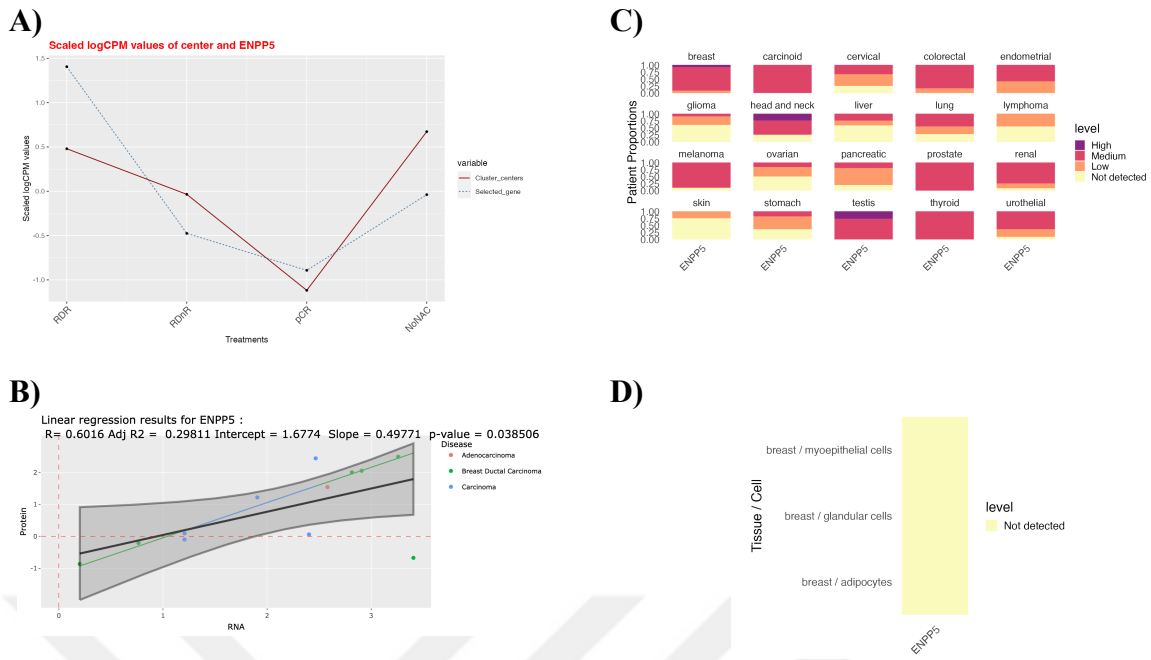


Figure 28. Analysis of ENPP5 gene. A) Plot of scaled logCPM values of ENPP5 gene against its cluster centroid. B) Correlation plot for mRNA and protein levels in breast cancer obtained from DepMap data. C) Protein levels of IL1R2 gene in all cancers by HPA. D) Protein level of ENPP5 gene in normal breast tissues.

3.4 Drospirenone, Estrogen, Aldosterone Driven Expression Changes in T47D Cells Analyzed by Mirror Clusters

Other analyses on CAP-RNAseq were performed to explore the complex interactions among steroid hormones and their regulatory effects on the transcriptomes of breast cancer cells. Firstly, RNA-seq data obtained from T47D cells were investigated.

3.4.1 RNA-seq Quality

After RNA isolation, samples with a volume of at least 40 µL and concentration more than 50 ng/µL were sent to the company for sequencing. Table 20 lists each sample's RNA Integrity Number (RIN) values. RNA sequencing for T47D samples was done in two batches. In the first batch, 4 groups (Control, Drospirenone, Drospirenone: Estrogen and Drospirenone: Estrogen: Aldo) were sent as a total of 8 samples with 2 biological replicates from each. In the second batch, 1 more replicate was added for each group in the first batch, and one sample of Aldo and E2 samples were added and sent for sequencing.

Table 20. RIN values of samples from T47D cells.

	Sample Name	RIN
First batch	Control_1	9.8
	Control_2	10.0
	Drsp_1	9.6
	Drsp_2	9.9
	Drsp_E2_1	9.8
	Drsp_E2_2	10.0
	Drsp_E2_Aldo_1	9.8
	Drsp_E2_Aldo_2	9.7
Second batch	Control	10
	Drsp	10
	Drsp_E2	10
	E2	10
	Aldo	10
	Drsp_E2_Aldo	10

For bioinformatics analysis of RNA-seq results performed at two different times, it is necessary to eliminate the batch effect. For this purpose, the combat function of the sva package (Leek et al., 2012) was used. This function uses a negative binomial regression to model the read data and estimates parameters representing batch effects. It then provides corrected data by mapping the original data to an expected distribution without any batch effects. Before performing this function, I filtered the data by removing zero and low-expressed genes from our T47D RNA-seq data. Then, the Combat function was run with the filtered raw reads, and the PCA plot of the results was drawn (Figure 29). Aldo, E2, and samples labeled as 3 belong to the second RNA-seq data, whereas samples labeled as 1 and 2 belong to the first data. In the PCA drawn after Combat, it was observed that the replicates of DE2 and DEA samples were very well separated. Control and Drsp samples were also very well separated, especially in PC1. E2 is the determinant of the separation at PC1, while Aldo appears to be effective in the presence of Drsp_E2, but not when alone.

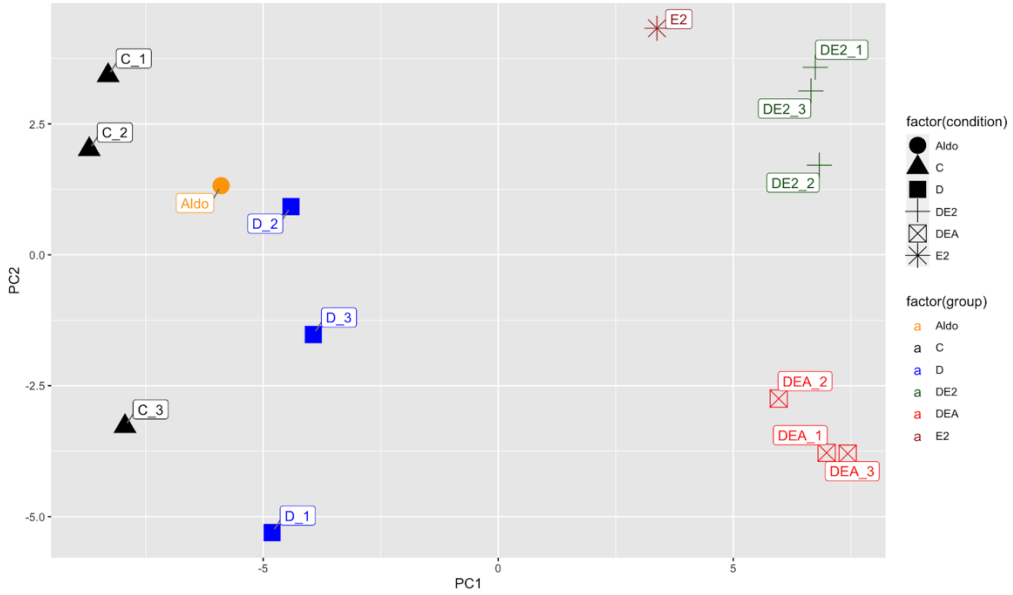


Figure 29. PCA result of two RNA-seq data after combat.

3.4.2 Analysis of T47D Data in CAP-RNAseq

After removing zero and lowly expressed genes from the data, 14,135 genes were retained. After uploading to CAP-RNAseq, CPM filtering was applied to the data, setting a threshold of a minimum CPM of 0.2 in at least three samples. Then, to select the genes that show significant changes in expression between any two groups, vst+ANOVA was applied, utilizing a stringent p-value cutoff of 0.001. This approach yielded a subset of 1,571 genes that demonstrated significant differential expression between any two groups. The identified genes were then subjected to cluster analysis, resulting in 12 different expression clusters (Figure 30). In the rest of the analysis, the data were organized into pairs of mirror clusters to facilitate the interpretation process (Figure 31).

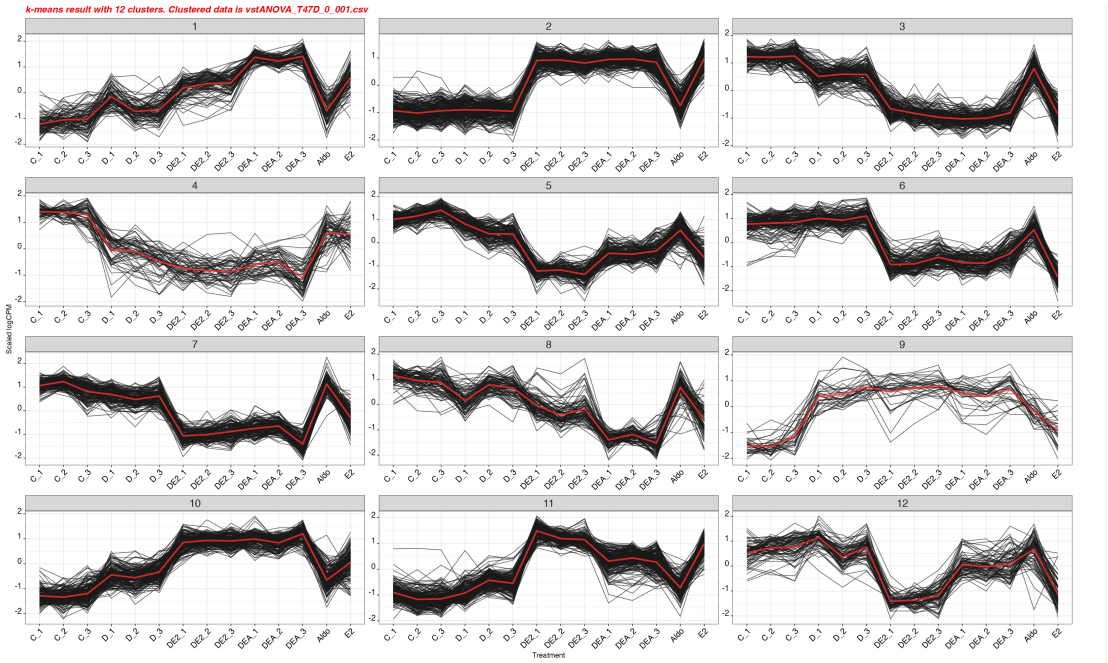
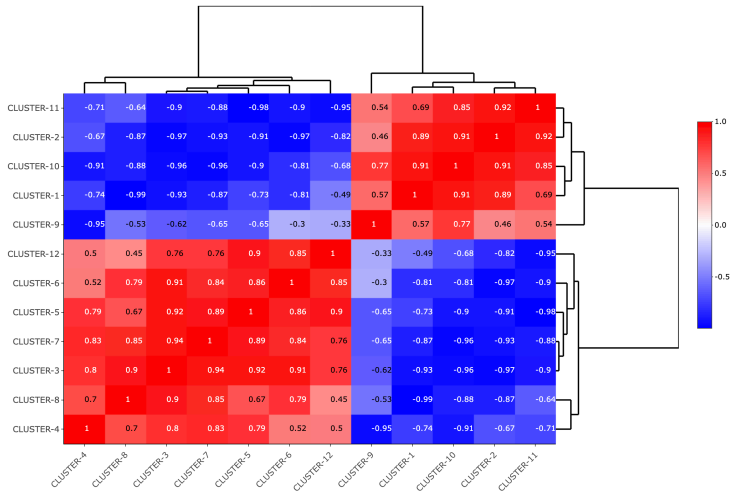


Figure 30. k-means clustering result of 1,571 genes in T47D RNA-seq data with 12 clusters. The cluster centroid pattern is shown by red lines.

3.4.3 Mirror Clusters and Essentiality Analysis of Clusters

Using Pearson's correlation, mirror clusters were identified in T47D data. According to the heatmap and the table, Clusters 1 and 8, Clusters 2 and 6, Clusters 3 and 2, Clusters 4 and 9, Clusters 5 and 11, Clusters 7 and 10 and Clusters 11 and 5 were found to be the mirror clusters where each pair showed opposite expression patterns (Figure 31).

A)



B)

Cluster	Mirror
CLUSTER-1	CLUSTER-8
CLUSTER-2	CLUSTER-6
CLUSTER-3	CLUSTER-2
CLUSTER-4	CLUSTER-9
CLUSTER-5	CLUSTER-11
CLUSTER-6	CLUSTER-2
CLUSTER-7	CLUSTER-10
CLUSTER-8	CLUSTER-1
CLUSTER-9	CLUSTER-4
CLUSTER-10	CLUSTER-7
CLUSTER-11	CLUSTER-5
CLUSTER-12	CLUSTER-11

Figure 31. Dissimilarity heatmap for 12 clusters to find the mirror clusters. A) Heatmap visualization. B) Table visualization of mirror clusters.

The cluster with the highest number of essential genes was cluster 11 with 121 essential genes, which had also highest odds ratio (Table 21). It is followed by cluster 2 (n=73) and then cluster 10 (n=43). In contrast, the lowest number of essential genes were found in cluster 9 with only 1 essential gene, followed by cluster 4 with 10 essential genes. The highest efficacy mean was also found in the cluster 11, correlated with the highest number of essential genes. However, the highest mean of selectivity was seen in cluster 12.

Table 21. Essentiality information of clusters for T47D dataset. SD: standard deviation; OR: Odds Ratio; FET: Fisher's Exact Test; p.val: p-value.

Clusters	Mean of Efficacy	SD of Efficacy	Mean of Selectivity	SD of Selectivity	Number of Essential Genes	Total Number of Genes	Percentage of Essential Genes	OR of FET	p.value of FET
Cluster 1	-0.9	0.4	0.15	0.21	26	82	31.71	01.03	0.9
Cluster 2	-0.89	0.42	0.12	0.23	73	201	36.32	1.26	0.15
Cluster 3	-0.55	0.17	0.1	0.21	31	144	21.53	0.61	0.02
Cluster 4	-0.65	0.17	0.21	0.24	10	53	18.87	0.51	0.07
Cluster 5	-0.65	0.23	0.16	0.23	33	131	25.19	0.74	0.17
Cluster 6	-0.63	0.31	0.17	0.22	18	114	15.79	0.41	0
Cluster 7	-0.55	0.13	0.12	0.21	31	134	23.13	0.66	0.06
Cluster 8	-0.73	0.45	0.2	0.4	23	76	30.26	0.96	1
Cluster 9	-0.5		0.23		1	26	3.85	0.09	0
Cluster 10	-0.78	0.43	0.08	0.19	43	150	28.67	0.89	0.58
Cluster 11	-1.07	0.44	0.03	0.19	121	197	61.42	3.51	0
Cluster 12	-0.66	0.18	0.29	0.28	17	62	27.42	0.83	0.58

3.4.4 Annotation Networks of Genes Based on Essentiality in Mirror Clusters

MSigDB analysis was performed for essential and non-essential genes, separately, using Hallmark term (Figure 32). Essential genes in cluster 1, 2 10 and 11, which all exhibited increasing pattern compared to control samples, shared common terms, such as E2F targets and G2M checkpoint, while mitotic spindle term was found in both cluster 1 and 10, and MYC targets term was shared by cluster 2 and 11 (Figure 32A). Particularly, essential genes in cluster 2 were also enriched in early and late estrogen response, indicating that these genes in cluster 2 were responsible to estrogen pathway. MTORC1 signaling and MYC targets were also appeared in the cluster 11. On the contrary, essential

genes downregulated in D_E2 samples of cluster 8 and 12 were enriched in protein secretion and peroxisome for cluster 8 and UV response for cluster 12.

Furthermore, non-essential genes in clusters 2, 9, and 10, which all exhibited increasing pattern compared to control samples, shared the estrogen response, while cluster 9 was particularly enriched in xenobiotic metabolism and androgen response (Figure 32B). Genes in cluster 11 were also found to be enriched in MYC targets. On the other hand, the non-essential genes downregulated in D_E2 samples compared to control were linked with WNT beta Catenin signaling, IL2/STAT5 signaling and HEME metabolism for cluster 12, and adipogenesis for cluster 3.

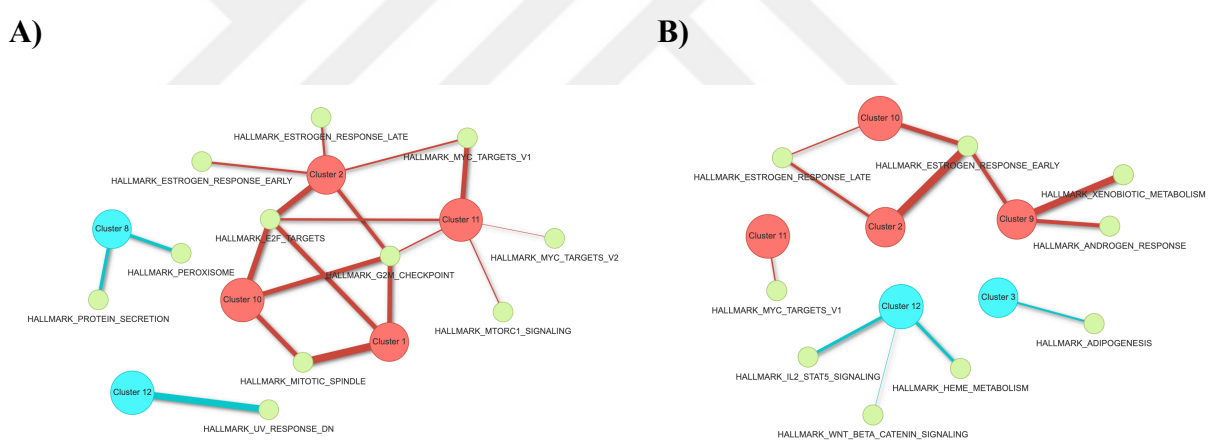


Figure 32. MSigDB networks for Hallmark gene set collection. Network visualization with enriched terms having q-values less than 0.05 for essential genes (A) and non-essential genes (B) in the clusters. The number of genes that overlap with each gene set collection was shown by edge widths. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between D_E2 and control samples.

3.4.5 Identification of Genes Affected by Drsp in Mirror Clusters 4 and 9

Clusters 4 and 9 showed opposite patterns (Figure 33A, B). In cluster 4, genes were downregulated in the samples containing Drsp regardless of E2 and Aldo, compared to control samples, while in cluster 9, genes were upregulated in these samples. However, in samples containing only E2 or Aldo, these genes were not affected as much as in other samples. These clusters revealed the genes that were affected by Drsp. In the MSigDB analysis of cluster 9 using Hallmark and REACTOME pathways, genes upregulated by Drsp were enriched in the gene sets related to fatty acid metabolism, xenobiotic metabolism, early estrogen response and androgen response (Figure 33C). However, in this cluster, there is only one essential gene (MYO10) (Figure 33D).

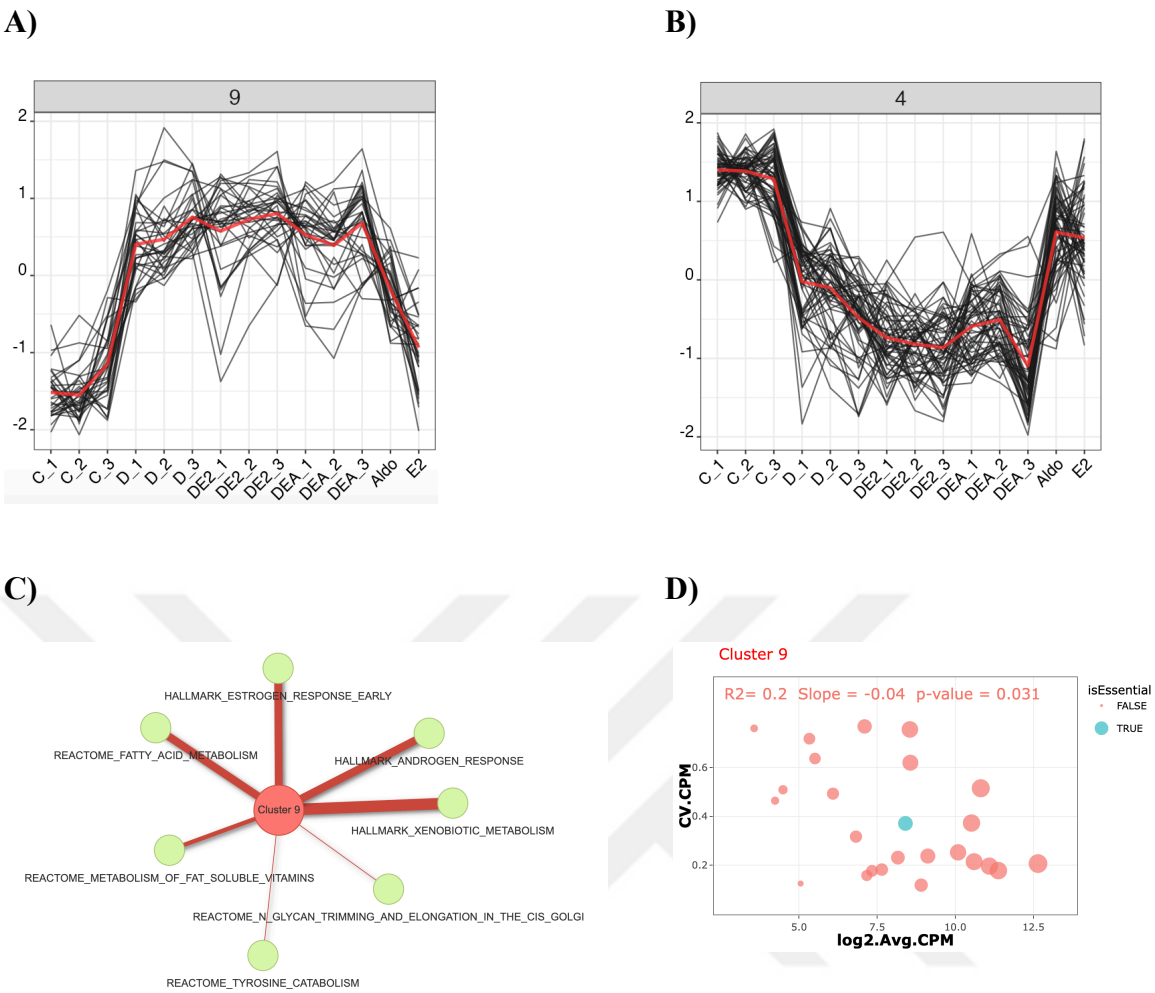


Figure 33. Cluster 9 and 4. A) Expression pattern of cluster 9. B) Expression pattern of cluster 4. C) MSigDB analysis of cluster 9 using Hallmark and REACTOME gene sets with q-value< 0.1 for top 7 terms. D) Scatter plot showing the log2 of average CPM in the x-axis and coefficient of variation of CPM in the y-axis of genes in cluster 9. The size of the dots shows the log2 of range CPM.

In the deeper analysis of genes in cluster 9 under the “Gene Prioritization” tab, DESeq2 was used to perform a differential expression analysis with the aim of comparing the effects of Drospirenone with the Control group (Table 22). The resulting table was then arranged to display the genes that showed the highest correlation (r) between mRNA and protein expression levels and display the first 20 genes are shown. Columns for p values were removed since all of them were less than 0.05.

Table 22. The result of DESeq2 analysis of genes in cluster 9.

Genes	Correlation	D vs C logFC	D vs C baseMean	Efficacy	Selectivity	isEssential	r
PLLP	0.938	1.12	674.08	-0.27	0.108	false	0.834
ARRB1	0.635	0.73	6464.75	-0.331	0.043	false	0.8197
FKBP5	0.963	1.91	10735.5	-0.167	0.053	false	0.8139
AKR1C3	0.854	0.89	134.15	-0.221	-0.033	false	0.8089
CD59	0.717	0.57	12964.64	-0.228	0.091	false	0.775
RBP2	0.873	1.97	70.6	-0.182	0.106	false	0.7419
AZGP1	0.991	4.05	2222.21	-0.258	-0.232	false	0.7199
FAH	0.941	0.86	1705.86	-0.319	0.198	false	0.7193
AQP3	0.966	4.52	814.54	-0.287	-0.041	false	0.7157
SORBS1	0.932	2.08	402.38	-0.175	0.251	false	0.6827
ACBD7	0.913	0.5	1619.69	-	-	-	0.6355
MYO10	0.924	1.31	2008.76	-0.498	0.227	true	0.6133
NETO2	0.813	0.51	960.86	-0.217	0.098	false	0.5955
MAN1A1	0.94	1.5	8702.58	-0.198	-0.004	false	0.5691
ELOVL5	0.967	0.64	15823.89	-0.205	0.07	false	0.5603
HECTD3	0.795	0.33	2860.33	-0.393	-0.063	false	0.3087
MRFAP1	0.917	0.52	38183.46	-0.14	-0.055	false	0.2335
ZBTB16	0.971	10.84	162.47	-0.202	0.12	false	-
SGK1	0.982	3.49	239.61	-0.256	-0.028	false	-

The top gene in this table was the PLLP gene, which shows the highest correlation between mRNA and protein levels and also a higher distance correlation with a log2 fold change of 1.12. While Aldo and E2 slightly increased the expression of this gene, more increases were observed in the Drsp-containing samples (Figure 34A). In HPA analysis, the expression of this protein was found at low to medium amounts in normal breast tissues but only a few breast cancer patients showed low levels of it (Figure 34B-C).

Another prioritized gene from this table was FKBP5 which is found in the pathways related to androgen response, early and late estrogen response, and transcriptional regulation by MECP2. FKBP5 showed an expression pattern similar to that of the PLLP gene (Figure 34D). In HPA analysis, it showed medium expression in all normal breast tissues and a high level of protein expression in breast cancer patients (Figure 34E-F).

The last gene prioritized from this table was the MYO10 gene, which is the only essential gene in this cluster. The expression of this gene in T47D dataset was similar to the PLLP and FKBP5 genes (Figure 34G). However, in HPA data, most of the breast cancer patients showed medium levels of expression while only small subsets had high and low expression levels (Figure 34H-I).

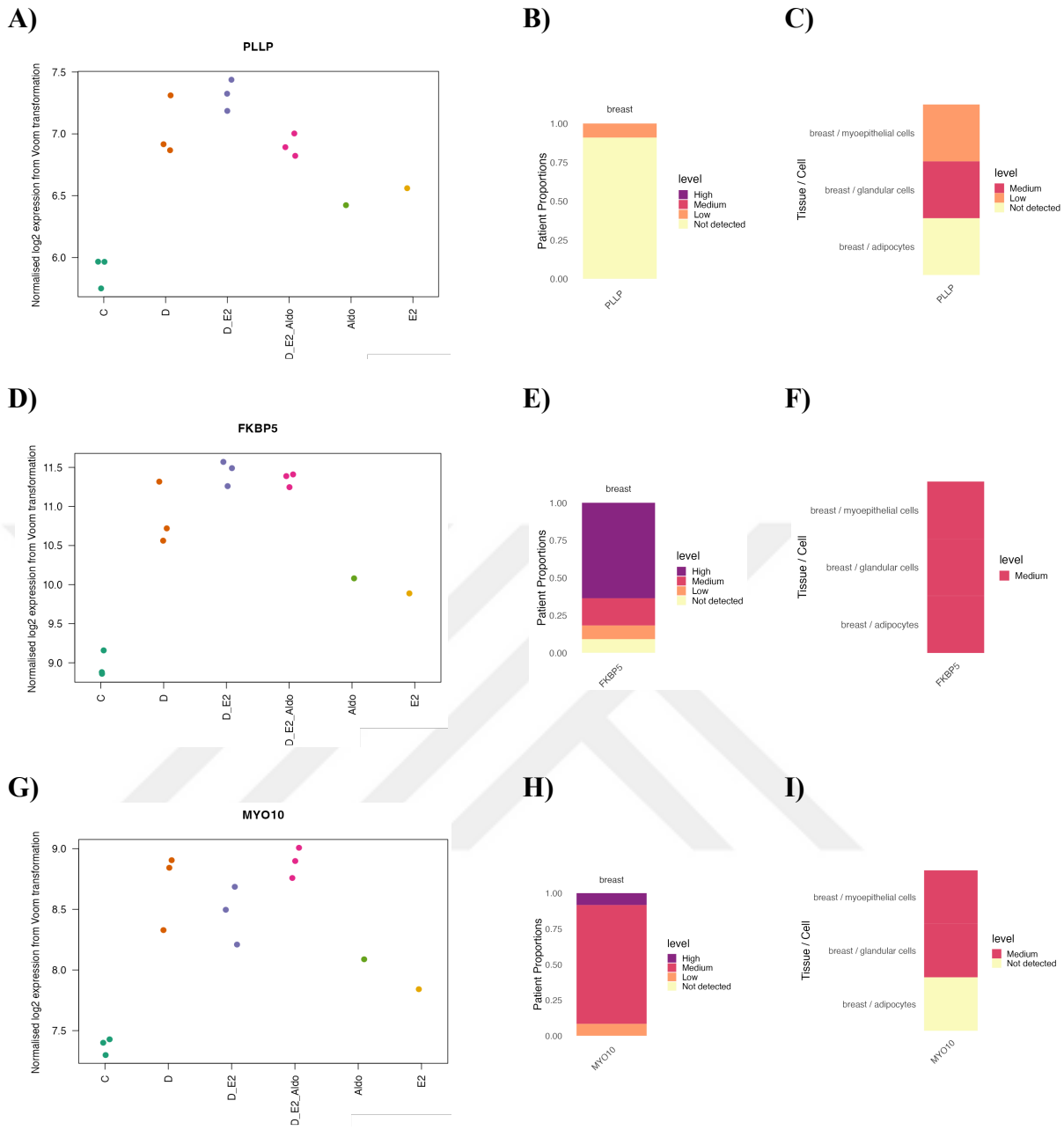


Figure 34. Log2 expression graphs and HPA results of PLLP (A-B-C), FKBP5 (D-E-F) and MYO10 (G-H-I) genes.

In the analysis of cluster 4, which was a mirror cluster of cluster 9, genes that were downregulated by Drsp-containing samples but not affected much by E2 alone were prioritized. MSigDB analysis of these genes using KEGG, REACTOME, and GO gene sets revealed WNT signaling pathway many times among the first ten terms (Table 22). Additionally, other significant terms that emerged included pathways in cancer, regulation of muscle system and contraction, endoplasmic reticulum organization, JNK cascade, and ubiquitin-dependent protein catabolic process.

Table 23. First 10 results of the enrichment analysis of the genes in cluster 4 for the KEGG, REACTOME, and GO terms.

ID	Gene Ratio	BgRatio	p value	p adj	q value	geneID	Count
KEGG_PATHWAYS_IN_CANCER	8/52	325/18497	0	0.004	0.003	AKT1/CCNE2/DAPK3/FZD2/KITLG/MECOM/SMAD3/TCF7L1	8
GOBP_CELL_CELL_SIGNALING_BY_WNT	9/52	455/18497	0	0.004	0.003	AMOTL1/DAPK3/FZD2/LRRK2/SHISA2/SMAD3/SOX2/SPIN1/TCF7L1	9
GOBP_REGULATION_OF_WNT_SIGNALING_PATHWAY	7/52	337/18497	0	0.022	0.018	DAPK3/LRRK2/SHISA2/SMAD3/SOX2/SPIN1/TCF7L1	7
GOBP_REGULATION_OF_MUSCLE_SYSTEM_PROCESS	6/52	250/18497	0	0.027	0.022	ANXA6/DAPK3/PKP2/PLCE1/PRKG1/SMAD3	6
GOBP_REGULATION_OF_MUSCLE_CONTRACTION	5/52	170/18497	0	0.027	0.022	ANXA6/DAPK3/PKP2/PLCE1/PRKG1	5
GOBP_ENDOPLASMIC_RETICULUM_ORGANIZATION	4/52	90/18497	0	0.027	0.022	LRRK2/RAB18/REEP1/SPTSSB	4
GOBP_JNK_CASCADE	5/52	175/18497	0	0.027	0.022	ARHGEF6/LRRK2/MECOM/SH3RF1/SMAD3	5
GOBP_POSITIVE_REGULATION_OF_PROTEASOMAL_UBIQUITIN_DEPENDENT_PROTEIN_CATABOLIC_PROCESS	4/52	93/18497	0	0.027	0.022	AKT1/CLU/LRRK2/SH3RF1	4
REACTOME_DEACTIVATION_OF_THE_BETA_CATENIN_TRANSACTIVATING_COMPLEX	3/52	42/18497	0	0.035	0.028	AKT1/SOX2/TCF7L1	3
GOBP_CANONICAL_WNT_SIGNALING_PATHWAY	6/52	313/18497	0	0.035	0.028	DAPK3/FZD2/LRRK2/SMAD3/SOX2/TCF7L1	6

Among them, SOX2 appeared as one of the highly correlated essential genes between mRNA and protein levels (Table 24). In the expression graph of SOX2, it was observed that E2 does not affect much the expression of genes, but all other samples (Drsp, DE2, DEA, and Aldo alone) can decrease it significantly (Figure 35A). However, as a result of HPA, the protein level of SOX2 was found in only a minority of breast cancer patients, while its expression is absent in most patients and normal breast tissue samples (Figure 35B-C).

Table 24. DESeq2 result of two genes selected from cluster 4.

Genes	Correlation	D vs C logFC	D vs C baseMean	D vs C p value	D vs C p adj value	Efficacy	Selectivity	isEssential	r
SOX2	0.822	-1.15	33.86	0	0.08	-0.738	0.541	true	0.7312
CLU	0.803	-0.64	27610.98	0	0	-0.236	-0.097	false	0.7647

Another prioritized gene from cluster 4 was CLU which had high expression in the data. While its expression did not change by E2 alone, it was downregulated in all other samples (Figure 35D). In HPA analysis, the expression of CLU was not found in normal breast tissues and most of the breast cancer patients, but only a small subset of the patients showed low to medium levels (Figure 35E-F).

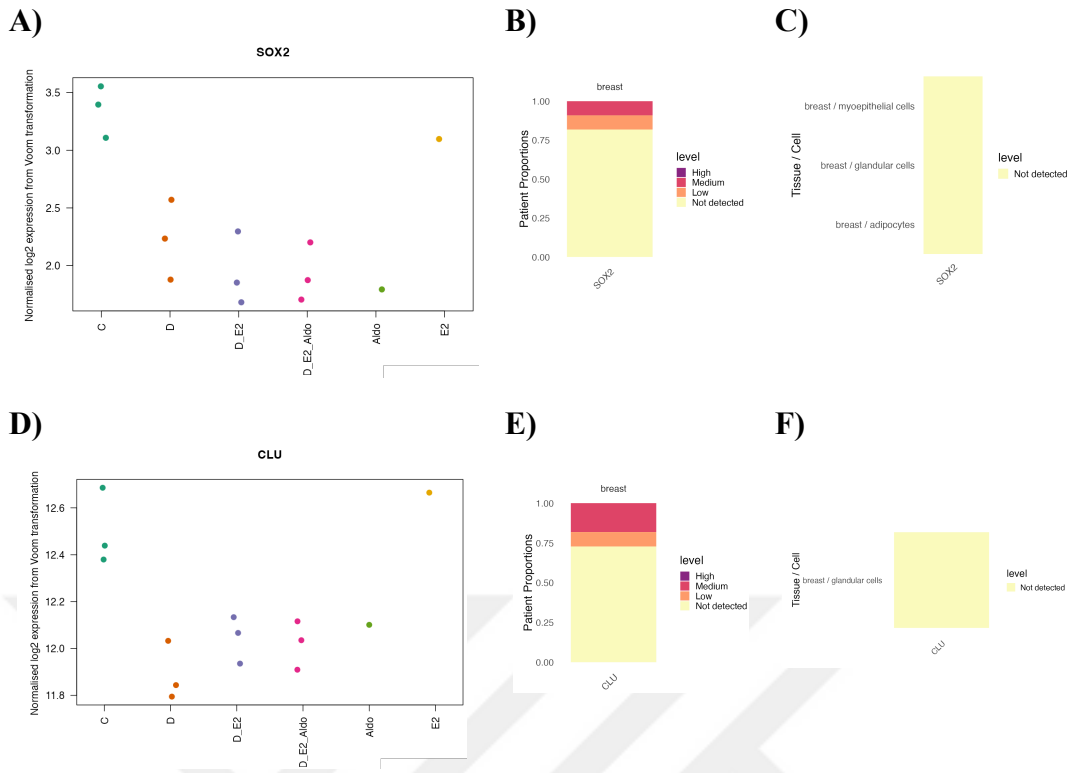


Figure 35. Log2 expression graphs and HPA results of SOX2 (A-B-C), and CLU (D-E-F) genes.

3.4.6 Synergistic Effect of Drsp and E2 in Mirror Clusters 5 and 11

Genes in cluster 5 were slightly reduced by Drsp alone, but the addition of E2 to Drsp (D_E2) caused a significant reduction in the expression, while E2 alone did not reduce expression as much as DE2 did (Figure 36A). This suggests that Drsp and E2 may show a synergistic effect. However, the addition of Aldo to D_E2 (DEA) reduced the decrease caused by D_E2 treatment, suggesting that aldosterone may further influence the gene expression alterations induced by Drospirenone and estrogen since Aldo alone could not decrease the expression as much. Genes in cluster 11 showed the opposite pattern, which synergistic effect of Drsp and E2 can be observed by the increase in D_E2 samples (Figure 36B).

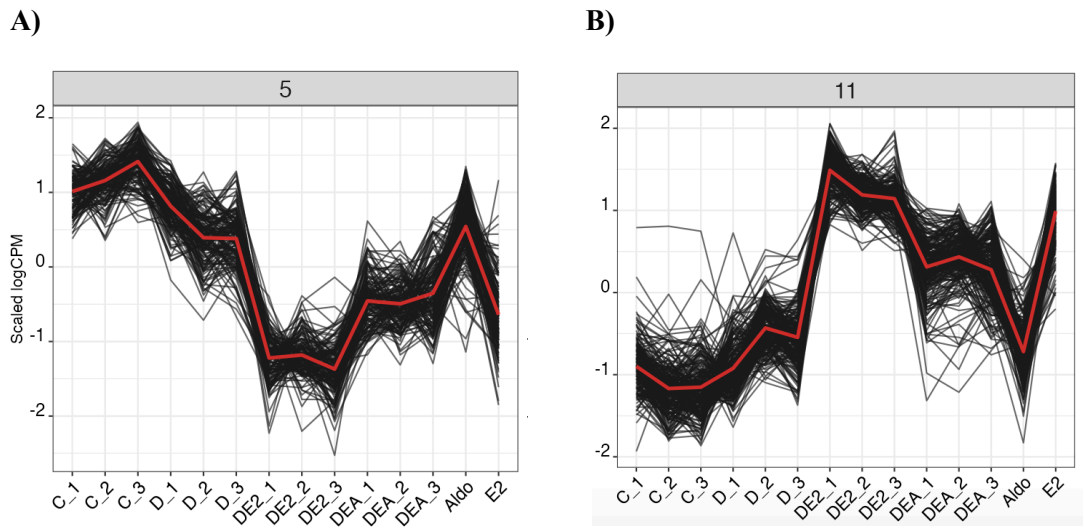


Figure 36. Expression patterns of mirror clusters 5 (A) and 11 (B).

In the MSigDB analysis using Hallmark and REACTOME pathways, genes in cluster 5 were enriched in NR1H2 and NR1H3 mediated signaling, heme metabolism and long-term potentiation (Figure 37A). NR1H2 (Liver X Receptor alpha) and NR1H3 (Liver X Receptor beta) are involved in cholesterol homeostasis, inflammation, and metabolism, while heme metabolism suggests an effect on iron homeostasis and oxidative stress, which are critical for the synthesis of cellular energy and may have an effect on the microenvironment and growth of tumors. The downregulation in long-term potentiation, a process linked to synaptic strength and memory, may suggest interesting functions in signal transduction and cancer cells' ability to adapt to the changing environment.

On the contrary, genes in cluster 11 were significantly enriched in MYC and E2F targets, formation of actin and tubulin folding, the resolution of sister chromatids, RHO GTPases activate formins, MTORC1 signaling and mitotic prometaphase (Figure 37A). The important roles played by MYC and E2F targets in cell cycle control and proliferation indicate the critical function that the increased upregulation effect by DE2 in cluster 11 plays in the development and division of cancer cells. These synergistically increased

genes in cluster 11 also affect cytoskeletal elements, which are essential for the invasion and spread of malignancy, and sister chromatids which have importance during cell division, indicating these genes can preserve genomic integrity. The increased upregulation effect in cluster 11 is associated with the cell cycle, cytoskeleton, and signaling.

In the essentiality metrics, scatter plots presenting the log2 of average CPM in the x-axis and coefficient of variation of CPM in the y-axis were drawn for the clusters 5 and 11. The size of the dots shows the log2 of range CPM. It was observed that cluster 11 contains more essential genes (n=121) than cluster 5 (n=33), and these essential genes displayed moderate to high expression in both clusters (Figure 37B, C).

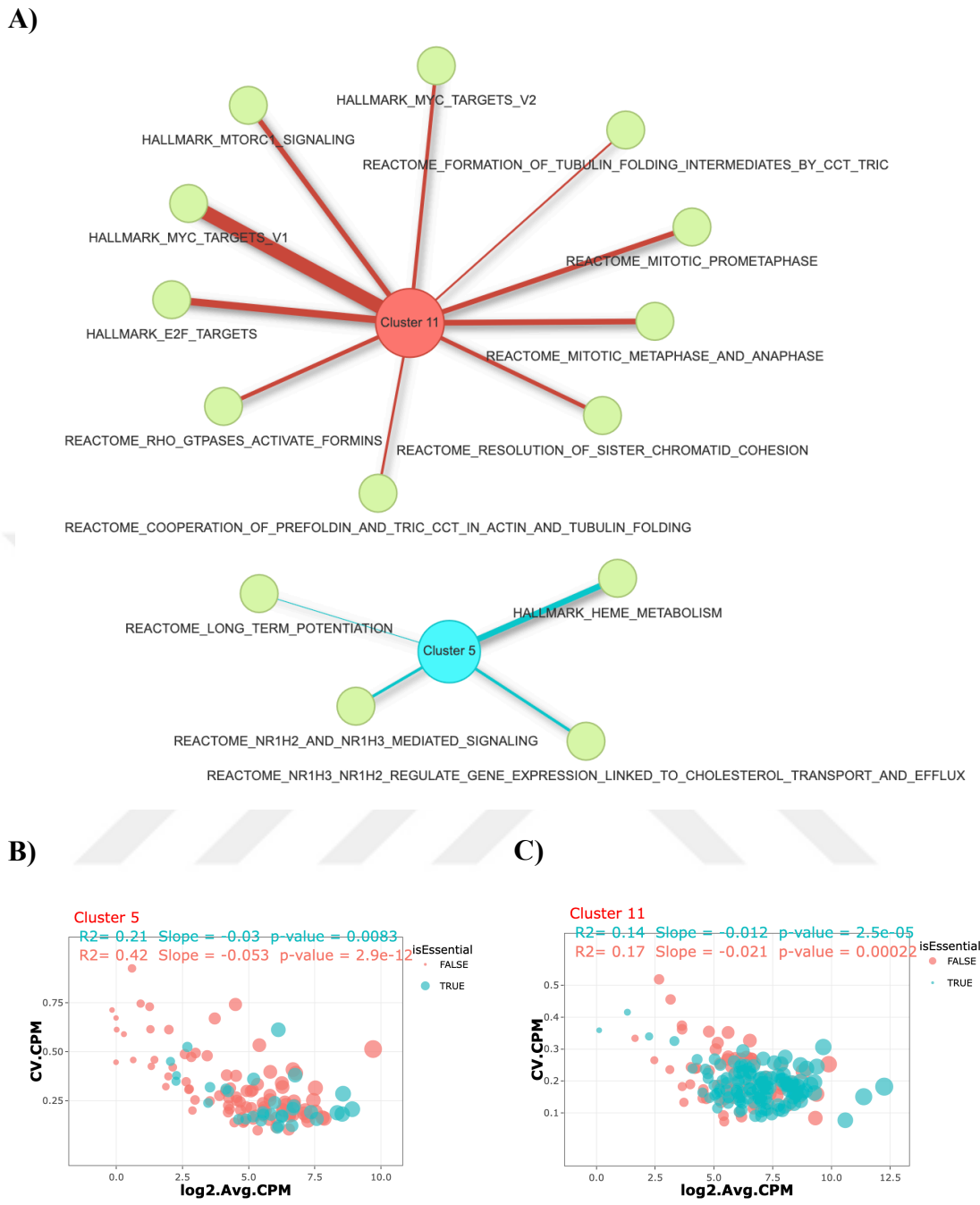


Figure 37. Analysis of clusters 5 and 11. A) MSigDB analysis using Hallmark and REACTOME gene sets with q-value< 0.1 for top 10 terms. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between D_E2 and control samples. B) Scatter plots for cluster 5 and C) cluster 11.

Among the essential genes showing a high correlation between mRNA and protein levels, FOXA1 was prioritized due to its decreased expression in D_E2 and high protein levels in breast cancers, while it was not found in normal breast tissues (Figure 38A, B, C). Its expression was significantly reduced by D_E2 samples, while the others had little effect. In the opposite cluster, cluster 11, the FKBP4 gene stood out because of its high expression levels and more increased in DE2 samples, whereas Drsp alone did not affect its expression but E2 alone increased it (Figure 38D). In HPA data, it was found at high levels only in glandular cells of breast tissue, not in other normal breast tissues, but was detected at high levels in all breast cancer patients (Figure 38E, F).

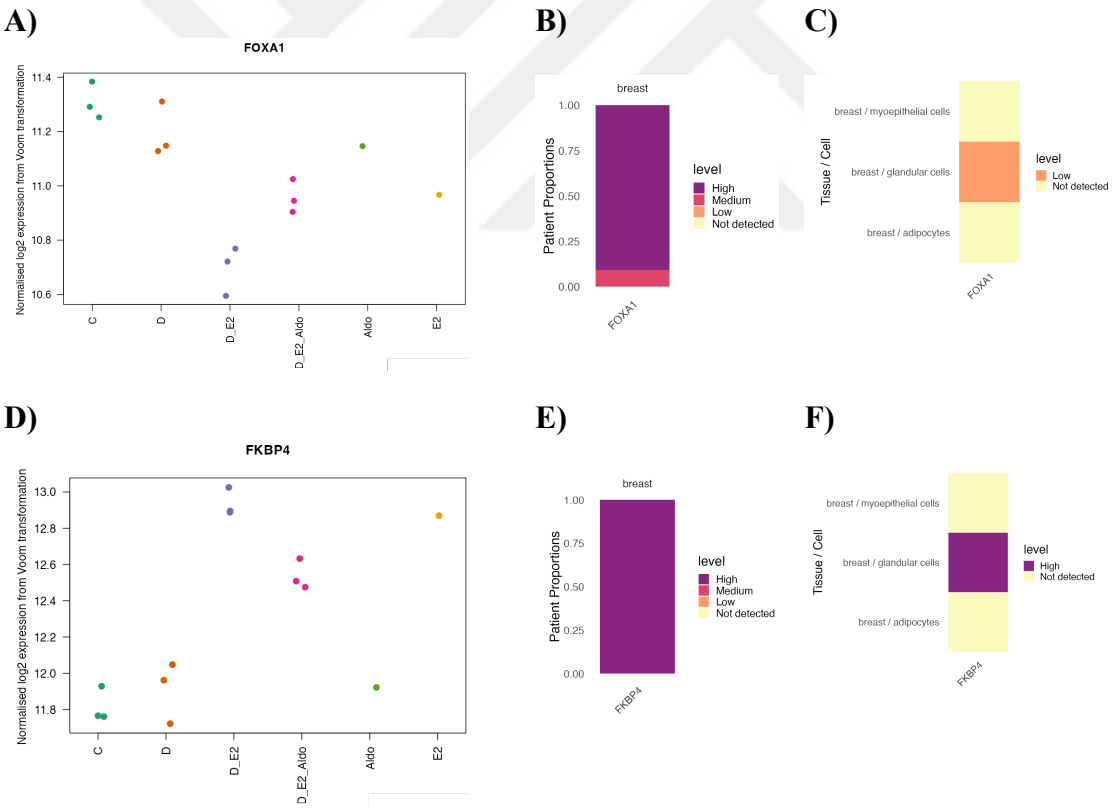


Figure 38. Log2 expression graphs and HPA results of FOXA1 (A-B-C), and FKBP4 (D-E-F) genes.

3.4.7 E2 Related Genes in Mirror Cluster 2 and 6

The genes affected by only E2 regardless of addition of other hormones were found in cluster 2 and 6. In the cluster 2, only E2 containing samples and E2 alone showed increased in the expressions of gene, while Drsp alone and Aldo alone had no impact on the gene expression (Figure 39). The opposite pattern of this was found in cluster 6. While genes in cluster 6 were enriched in only PTK6 expression of REACTOME pathway, genes in cluster 2 were enriched in estrogen responses, E2F targets, DNA repair, G2M checkpoint and telomer synthesis (Figure 40).

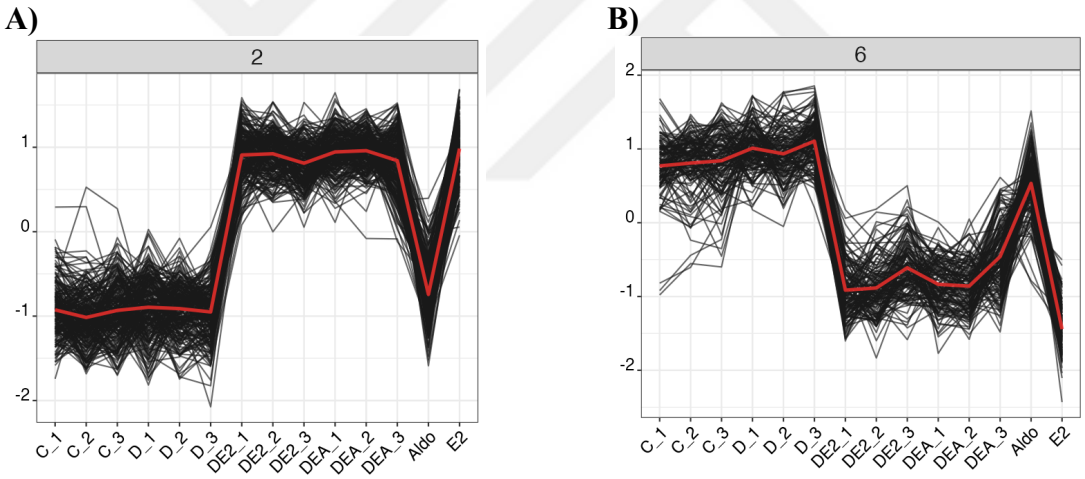


Figure 39. Expression patterns of mirror clusters 2 (A) and 6 (B).

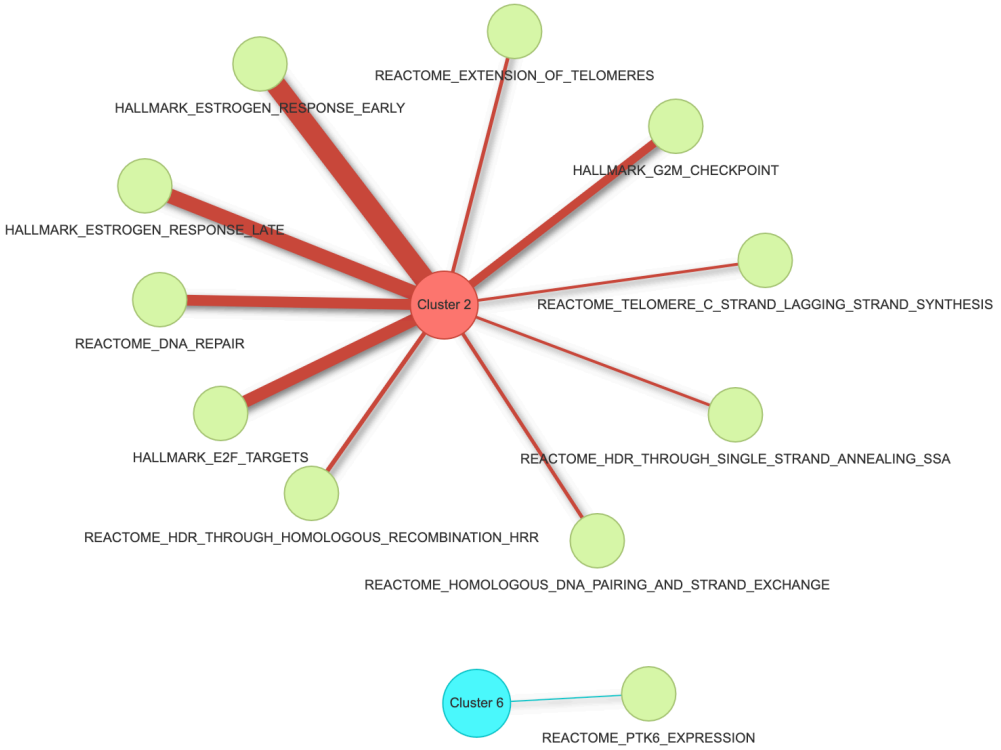


Figure 40. MSigDB analysis of clusters 2 and 6 using Hallmark and REACTOME gene sets with q-value< 0.1 for top 10 terms. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between D_E2 and control samples.

From the list generated by applying DESeq2 method on the genes of cluster 2, PDZK1 gene was identified as a candidate of interest due to its high expression in E2-containing samples and high correlation between mRNA and protein levels (Figure 41A). In HPA data, it was not found in normal breast tissues but found in low to medium levels in a subset of breast cancer patients (Figure 41B, C).

From the genes of cluster 6 where E2 showed decreasing effect, NDRG1 was selected. NDRG1 was also slightly affected by Drsp with an upward trend, while E2 decreased its expression but reduction with D_E2 was not as much as with E2 alone (Figure 41D), suggesting that Drsp can interact with E2 and reduces its effect. More than half of the breast cancer patients displayed medium to high levels of protein expression, while it was generally low in normal breast tissues (Figure 41E, F).

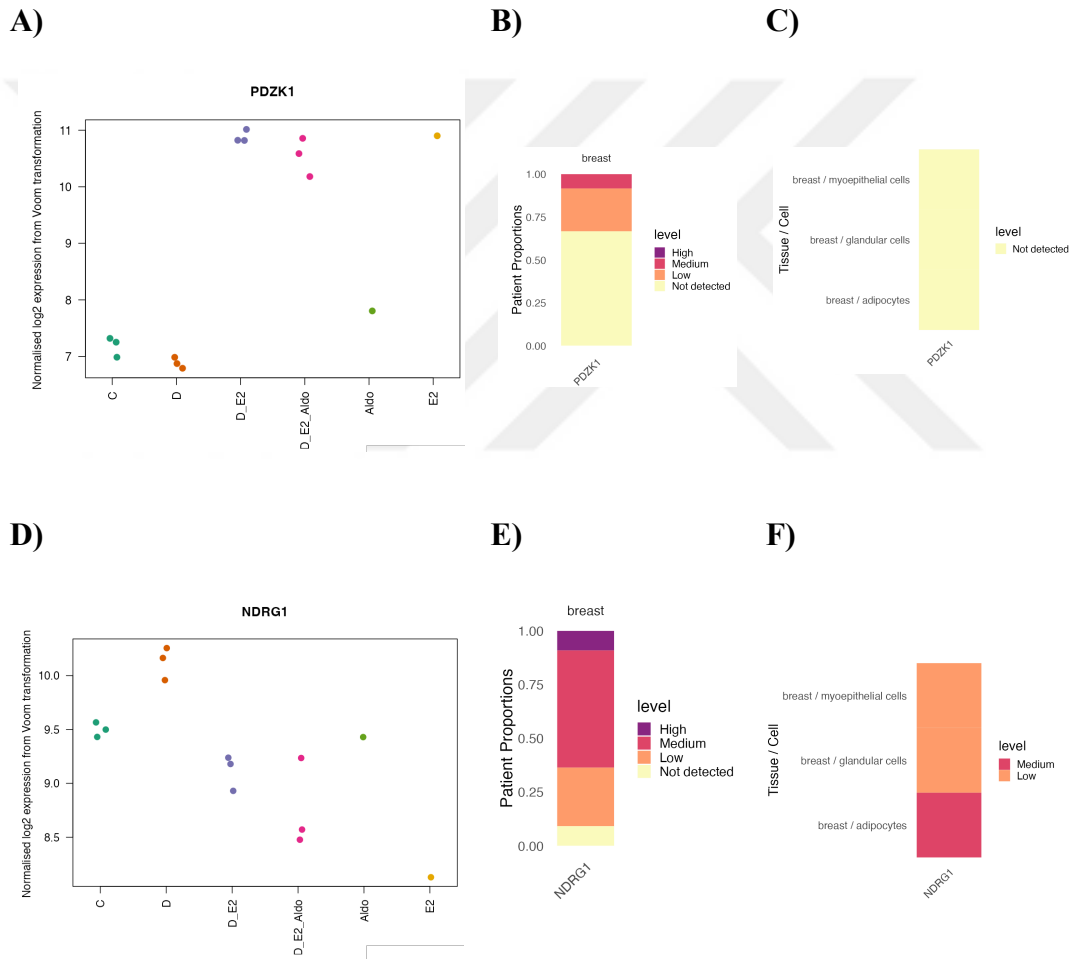


Figure 41. Log2 expression graph and HPA results of PDZK1 (A-B-C) and NDRG1 (D-E-F).

3.4.8 Essential Genes Downregulated by E2 Treatment

The cluster 3 contained 31 essential genes that were downregulated by the treatments containing estrogens (E2, D_E2 and DEA samples) (Figure 42A). Based on this pattern, it appears that these essential genes were responsive to estrogen either alone or together with the other hormones. In contrast, Aldo treatment did not significantly affect the expression of genes. However, treatment of Drsp alone caused slight downregulation, indicating a minor inhibitory action of this hormone on these essential genes. These essential genes also showed higher selectivity compared to genes in other clusters (Figure 42B). A subset of essential genes had selectivity close to 0.5, which was not observed in the majority of clusters.

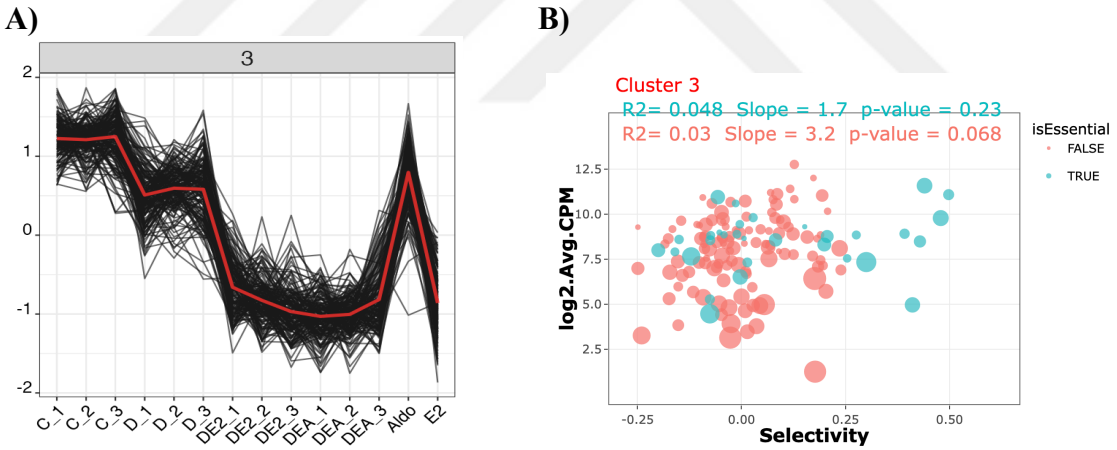


Figure 42. Analysis of cluster 3. A) Visualization of the expression patterns of cluster 3. B) Scatter plot showing the Selectivity values in the x-axis and log2 of average CPM in the y-axis of genes in cluster 3. The size of the dots shows the coefficient of variation of CPM.

In the MSigDB analysis of these essential genes using REACTOME and GO gene sets, ubiquitin-dependent protein catabolic process was appeared many times in the first ten terms (Table 25). Other appeared terms were positive regulation of proteolysis, SHC1 events in ERBB2 signaling, insulin receptor signaling pathway, negative regulation of cell adhesion, and SME4D related cell migration and semaphorin signaling.

Table 25. REACTOME and GO enrichment results of essential genes in cluster 3.

ID	Gene Ratio	BgRatio	p value	p adjust	q value	geneID	Count
GOBP_POSITIVE_REGULATION_OF_PROTEOLYSIS	5/28	373/19174	0	0.75	0.56	ZER1/TNFSF10/HERPUD1/FGFR4/CEBPA	5
GOBP_PROTEASOME_MEDIATED_UBIQUITIN_DEPENDENT_PROTEIN_CATABOLIC_PROCESS	5/28	413/19174	0	0.75	0.56	ZER1/WWTR1/HERPUD1/FBXW4/CEBPA	5
GOBP_POSITIVE_REGULATION_OF_PROTEASOMAL_UBIQUITIN_DEPENDENT_PROTEIN_CATABOLIC_PROCESS	3/28	93/19174	0	0.75	0.56	ZER1/HERPUD1/CEBPA	3
REACTOME_SEMA4D_INDUCED_CELL_MIGRATION_AND_GROWTH_CONE_COLLAPSE	2/28	20/19174	0	0.75	0.56	RND1/ERBB2	2
REACTOME_SHC1_EVENTS_IN_ERBB2_SIGNALING	2/28	22/19174	0	0.75	0.56	PRKCD/ERBB2	2
GOBP_POSITIVE_REGULATION_OF_UBIQUITIN_DEPENDENT_PROTEIN_CATABOLIC_PROCESS	3/28	110/19174	0.001	0.75	0.56	ZER1/HERPUD1/CEBPA	3
REACTOME_SEMA4D_IN_SEMAPHORIN_SIGNALING	2/28	24/19174	0.001	0.75	0.56	RND1/ERBB2	2
GOBP_PROTEASOMAL_PROTEIN_CATABOLIC_PROCESS	5/28	494/19174	0.001	0.75	0.56	ZER1/WWTR1/HERPUD1/FBXW4/CEBPA	5
GOBP_INSULIN_RECEPTOR_SIGNALING_PATHWAY	3/28	120/19174	0.001	0.75	0.56	PRKCD/PDK2/BCAR3	3
GOBP_NEGATIVE_REGULATION_OF_CELL_ADHESION	4/28	305/19174	0.001	0.77	0.58	RND1/PRKCD/ERBB2/BCAS3	4

Within this set of essential genes, ERBB2 gene was prioritized (Figure 43). While there was a slight decrease of its expression in Drsp alone samples, it was highly downregulated in other samples except Aldo. Protein expression was also found at moderate levels in

myoepithelial cells and glandular cells of normal breast tissue, while more than half of breast cancer patients showed moderate to high expression levels.

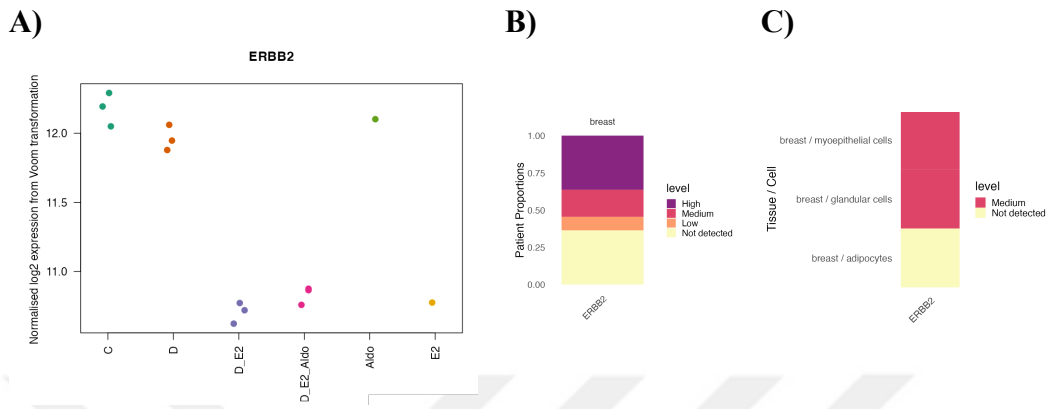


Figure 43. Log2 expression graph and HPA results of ERBB2 (A-B-C).

3.4.9 Genes Affected by DEA But Not Aldo Alone in Mirror Clusters 1 and 8

Genes in cluster 8 were significantly downregulated in DEA samples. While these genes were not affected by Aldo treatment alone, the addition of Aldo to D_E2 (DEA) resulted in increased downregulation compared to DE2 treatment (Figure 44). Generally, the amount of downregulation with E2 alone and D_E2 was the same for most of the genes, suggesting that the decrease in D_E2 may be due to the E2 effect alone, but Drsp alone also had some downregulating effect in the expression of some genes. Moreover, in some of the genes, the decrease by DEA and E2 alone was same, indicating that the addition of Aldo to D_E2 samples returns the effect to that of E2 alone. The GO enrichment of these genes revealed the pathways related to negative regulation of intrinsic apoptotic signaling, indicating the role of genes in preventing the apoptosis, protein localization which is required to ensure proteins reach their correct cellular destinations, and autophagosome assembly (Table 25).

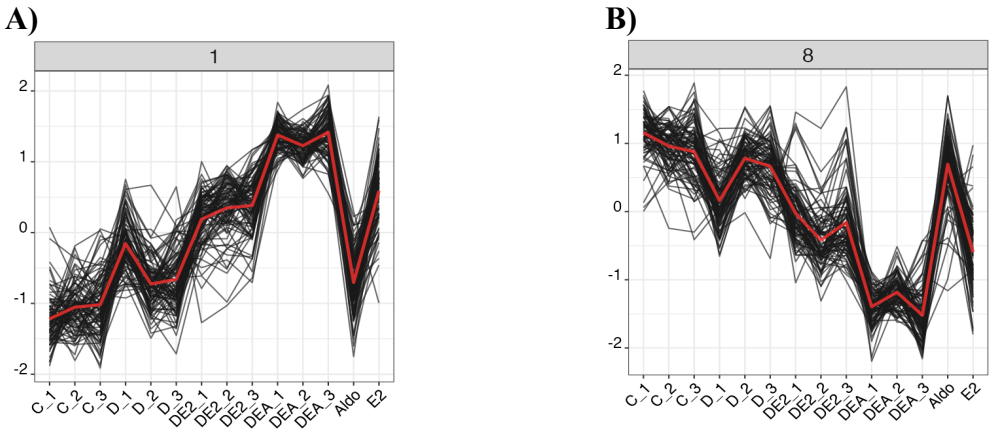


Figure 44. Expression patterns of mirror clusters 1 (A) and 8 (B).

Table 26. GO enrichment result of cluster 8.

ID	Gene Ratio	BgRatio	p value	p adjust	q value	geneID	Count
GOBP_NEGATIVE_REGULATION_OF_INTRINSIC_APOPTOTIC_SIGNALING_PATHWAY	4/79	102/18448	0.001	0.356	0.336	BCL2L1/GRINA/PINK1/PTTG1IP	4
GOBP_PROTEIN_LOCALIZATION_TO_PLASMA_MEMBRANE	6/79	285/18448	0.001	0.356	0.336	BCL2L1/FLOT2/INPP5K/RAB13/SLC9A3R2/ZDHHC3	6
GOBP_NEGATIVE_REGULATION_OF_PROTEIN_LOCALIZATION	5/79	199/18448	0.002	0.356	0.336	BCL2L1/CTNNA1/INPP5K/LEPROT/TRIM29	5
GOBP_NEGATIVE_REGULATION_OF_AUTOPHAGOSOME_ASSEMBLY	2/79	15/18448	0.002	0.356	0.336	PHF23/PINK1	2
GOBP_NEGATIVE_REGULATION_OF_APOPTOTIC_SIGNALING_PATHWAY	5/79	227/18448	0.003	0.356	0.336	BCL2L1/CTNNA1/GRINA/PINK1/PTTG1IP	5
GOBP_NEGATIVE_REGULATION_OF_ENDOPLASMIC_RETICULUM_STRESS_INDUCED_INTRINSIC_APOPTOTIC_SIGNALING_PATHWAY	2/79	20/18448	0.003	0.356	0.336	BCL2L1/GRINA	2
GOBP_PROTEIN_LOCALIZATION_TO_CELL_PERIPHERY	6/79	339/18448	0.003	0.356	0.336	BCL2L1/FLOT2/INPP5K/RAB13/SLC9A3R2/ZDHHC3	6
GOBP_RESPONSE_TO_ALCOHOL	5/79	236/18448	0.003	0.356	0.336	BCL2L1/CTNNA1/GLB1/GNB1/SDF4	5
GOBP_BRANCHED_CHAIN_AMINO_ACID_CATABOLIC_PROCESS	2/79	21/18448	0.004	0.356	0.336	BCKDHB/HIBADH	2
GOBP_APICAL_JUNCTION_ASSEMBLY	3/79	77/18448	0.004	0.356	0.336	CLDN7/CTNNA1/RAB13	3

Among the essential and highly correlated genes between mRNA and protein, BCL2L1 was identified. The expression of BCL2L1 gene remained stable by the treatments of Drsp and Aldo alone, but significantly decreased by D_E2, DEA and E2 alone treatments. However, the decrease in the expression was the same with DEA and E2. Furthermore, in HPA data, the protein expression of BCL2L1 was not detected in normal breast tissues, with only low amount observed in myoepithelial cells of breast. However, moderate levels of BCL2L1 protein expression were detected in over half of the breast cancer cases.

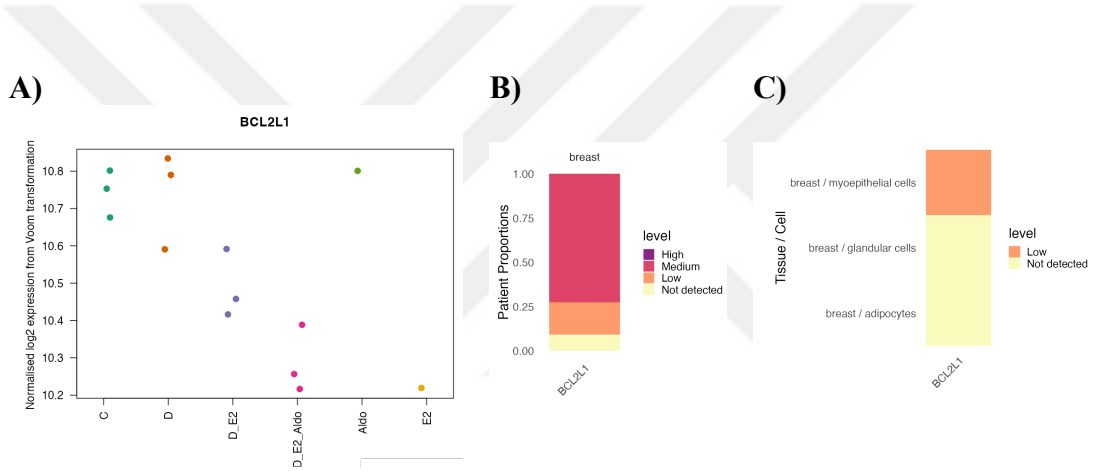


Figure 45. Log2 expression graphs and HPA results of BCL2L1 (A-B-C).

Cluster 1 which was the opposite cluster of 8, contained genes whose expression was increased by DE, DEA and E2. The enrichment of these genes by GO and REACTOME pathway, revealed the terms related to microtubule organization, centrosome duplication, chromosome and sister chromatid segregation and mitotic events. Among this list, centrosomal proteins were appear in most of the pathways.

Table 27. GO and REACTOME enrichment result of cluster 1.

ID	Gene Ratio	BgRatio	p value	p adjust	q value	geneID	Count
GOBP_MICROTUBULE_ORGANIZING_CENTER_ORGANIZATION	7/85	146/18448	0	0.008	0.008	CEP120/CEP135/CEP152/CEP85/CKAP5/ODF2/WDR62	7
GOBP_CENTROSOME_DUPLICATION	5/85	75/18448	0	0.012	0.011	CEP120/CEP135/CEP152/CKAP5/WDR62	5
GOBP_CHROMOSOME_SEGREGATION	9/85	350/18448	0	0.012	0.011	CEP85/DDX11/EME1/ESPL1/KIF14/KIF18B/KNTC1/NCAPG2/SMC4	9
REACTOME_RECRUITMENT_OF_MITOTIC_CENTROSOME_PROTEINS_AND_COMPLEXES	5/85	81/18448	0	0.012	0.011	CEP135/CEP152/CKAP5/ODF2/TUBGCP3	5
REACTOME_MITOTIC_PROMETAPHASE	7/85	203/18448	0	0.012	0.011	CEP135/CEP152/CKAP5/KNTC1/ODF2/SMC4/TUBGCP3	7
GOBP_SISTER_CHROMATID_SEGREGATION	7/85	207/18448	0	0.012	0.011	DDX11/ESPL1/KIF14/KIF18B/KNTC1/NCAPG2/SMC4	7
GOBP_NUCLEAR_CHROMOSOME_SEGREGATION	8/85	289/18448	0	0.012	0.011	DDX11/EME1/ESPL1/KIF14/KIF18B/KNTC1/NCAPG2/SMC4	8
GOBP_CENTRIOLE_ASSEMBLY	4/85	47/18448	0	0.012	0.011	CEP120/CEP135/CEP152/WDR62	4
GOBP_MITOTIC_NUCLEAR_DIVISION	8/85	296/18448	0	0.012	0.011	CEP85/ESPL1/KIF14/KIF15/KIF18B/KNTC1/NCAPG2/SMC4	8
REACTOME_RECRUITMENT_OF_NUMA_TO_MITOTIC_CENTROSOMES	5/85	95/18448	0	0.013	0.012	CEP135/CEP152/CKAP5/ODF2/TUBGCP3	5

Since the cluster 1 contained the genes for centrosomal proteins, one of them was selected for further investigation (Figure 46). CEP85, which was an essential gene, showed increased expression by DEA while the expression level by the treatments of DE2 and E2 alone was similar and less than DEA. In HPA data, CEP85 expression was not found in all normal breast tissues but medium to high expression of it was observed in all breast cancer patients.

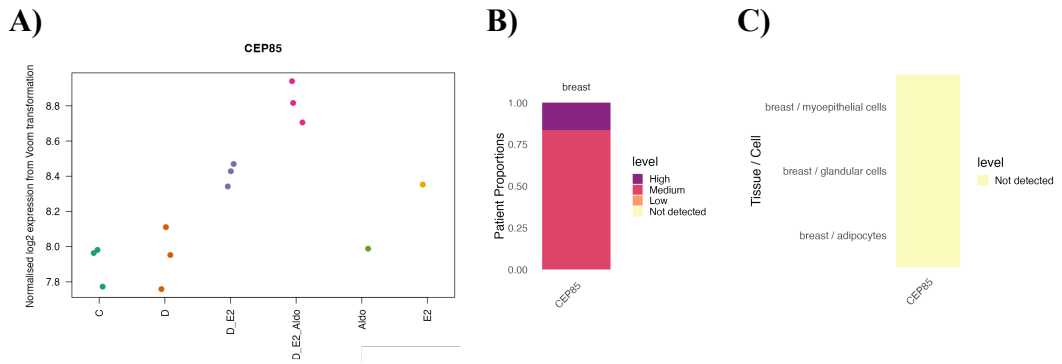


Figure 46. Log2 expression graph and HPA results of CEP85 (A-B-C).

3.4.10 Genes Affected by D_E2 and DEA But Not Solo Treatments in the Mirror Cluster 7 and 10

The genes of cluster 10 were increased by D_E2 and DEA treatments while same increase was not observed by any solo treatment. This indicates that the combination of drugs either as Drsp_ E2 or Drsp_E2_Aldo showed an increased effect. The same effect was also observed in cluster 7 in opposite manner (Figure 47).

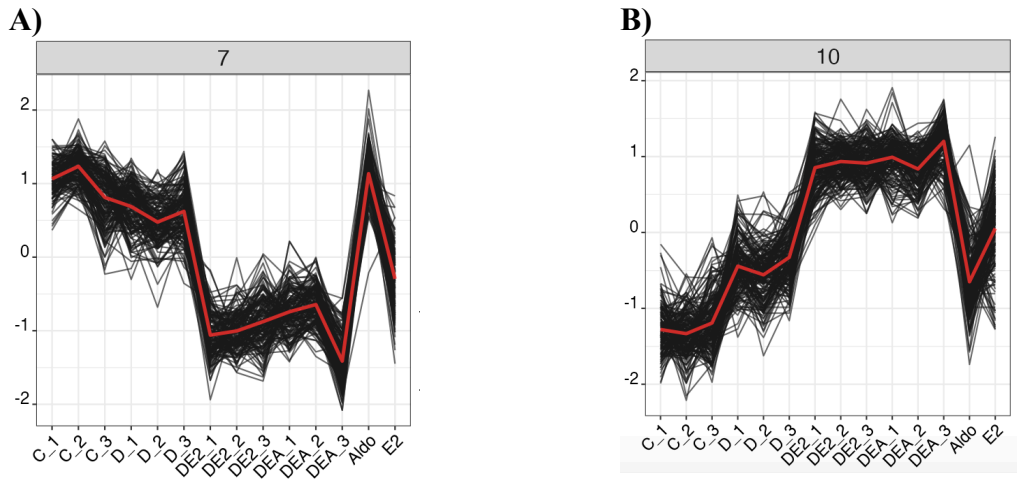


Figure 47. Expression patterns of mirror clusters 7 (A) and 10 (B).

The enrichment analysis of these two clusters using Hallmark and REACTOME pathways revealed the terms related to estrogen response, E2F targets, mitotic spindle, G2M checkpoint and glycolysis for the genes of cluster 10, and sphingolipid metabolism and synthesis, G protein events and UV response for the genes of cluster 7.

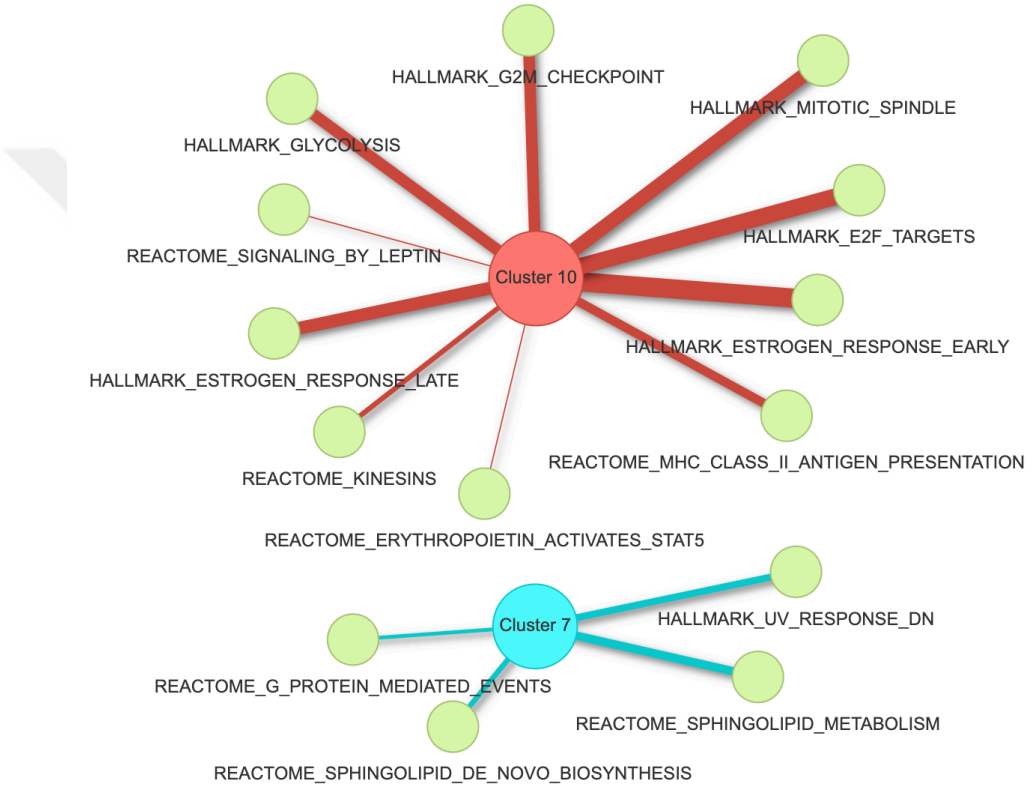


Figure 48. MSigDB analysis cluster 7 and 10 using Hallmark and REACTOME gene sets with q-value< 0.1 for top 10 terms. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between DEA and control samples.

From the genes of cluster 10, several essential genes were identified based on their high expression and correlation. Among them, RUNX1 showed increased expression by DE2 and DEA treatments, while other treatments could not increase it as much, while E2 alone caused slight increase (Figure 49A). It was also observed at medium levels in myoepithelial cells and glandular cells of breast tissue (Figure 49B). However, only a quarter exhibited medium levels of RUNX1 protein expression, with the remaining majority displaying low levels (Figure 49C).

Table 28. DESeq2 results of selected genes from cluster 10 and 7 in comparison of D_E2 versus Control samples.

Genes	Correlation	Log FC	baseMean	p value	p adj. value	Efficacy	Selectivity	isEssential	r
RUNX1	0.974	1.85	3127.44	0	0	-0.732	0.77	true	0.7374
ARF6	0.975	0.66	15008.93	0	0	-0.642	0.39	true	0.5661
CITED2	0.983	-2.18	3358.39	0	0	-0.498	0.419	true	0.654
TPM4	0.933	-2.75	69.31	0	0	-0.559	-0.054	true	0.5314

Another selected gene was ARF6 which also increased the expression by DE2 and DEA, while others have not much effect (Figure 49D). In HPA, it was found at medium levels only in glandular cells, and low levels in myoepithelial cells (Figure 49E). However, in all breast cancer patients, the protein expression of it was observed at medium levels (Figure 49F).

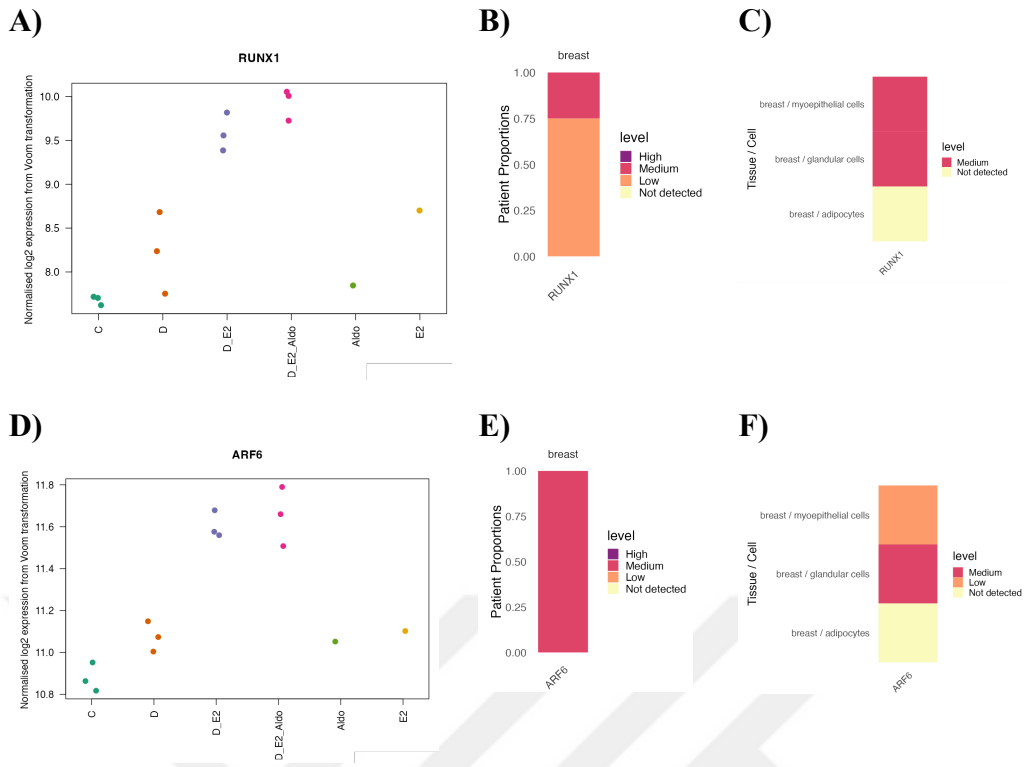


Figure 49. Log2 expression graphs and HPA results of RUNX1 (A-B-C) and ARF6 (D-E-F) genes.

In the opposite cluster, cluster 7, CITED2 and TPM4 were selected (Figure 50). Both showed decreased expression by DE2 and DEA samples, while E2 alone had little effect on it (Figure 50A). In HPA, protein expression of CITED2 was observed medium to high levels in normal breast tissues and high levels in all breast cancer patients (Figure 50B, C). The protein expression of TPM4, on the other hand, was high in both glandular cells and myoepithelial cells, but only half of the patients had high level while the rest of them had low to medium levels.

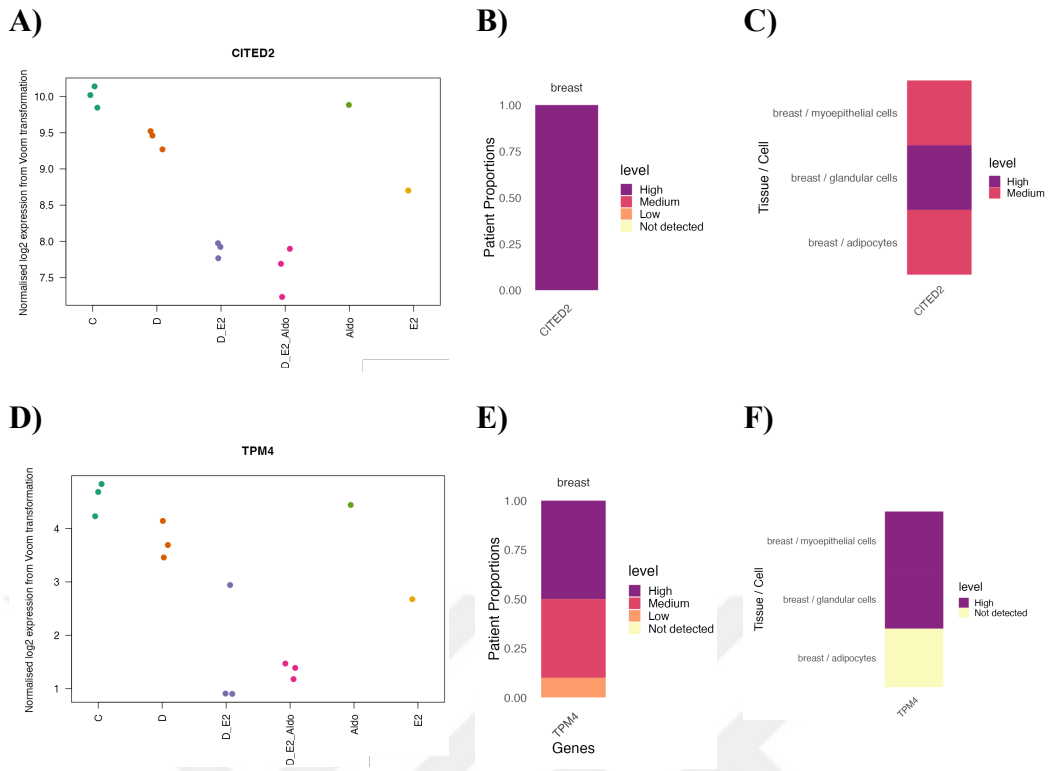


Figure 50. Log2 expression graphs and HPA results of CITED2 (A-B-C), and TPM4 (D-E-F) genes.

3.5 Hormone Driven Expression Changes in MCF-7 Cells

Another breast cancer RNA-seq data obtained from MCF-7 cell line was analyzed in CAP-RNAseq.

3.5.1 RNA-seq Quality

Following RNA isolation, samples were sent to the company for sequencing. Table 29 lists each sample's RNA Integrity Number (RIN) values. RNA sequencing for MCF-7 cell line was done for 6 different groups (Control, Drospirenone, Drospirenone: Estrogen, Drospirenone: Aldo, Estrogen:Aldo and Drospirenone: Estrogen: Aldo) with 2 biological replicates from each.

Table 29. RIN values of samples from MCF-7 cells.

Sample Name	Integrity value
Control_1	9.90
Control_2	9.80
Drsp_1	9.70
Drsp_2	9.90
Drsp_E2_1	8.60
Drsp_E2_2	10.00
Drsp_Aldo_1	9.70
Drsp_Aldo_2	9.70
E2_Aldo_1	10.00
E2_Aldo_2	10.00
Drsp_E2_Aldo_1	10.00
Drsp_E2_Aldo_2	10.00

After removing zero and lowly expressed genes from MCF-7 RNA-seq data in R, PCA was drawn to see the distributions of samples. It was observed that replicates did cluster

well, and Drsp_E2 and Drsp_Aldo groups were well separated, but some groups (e.g., control and Drsp, and Drsp_E2_Aldo and E2_Aldo) clustered very close to each other.

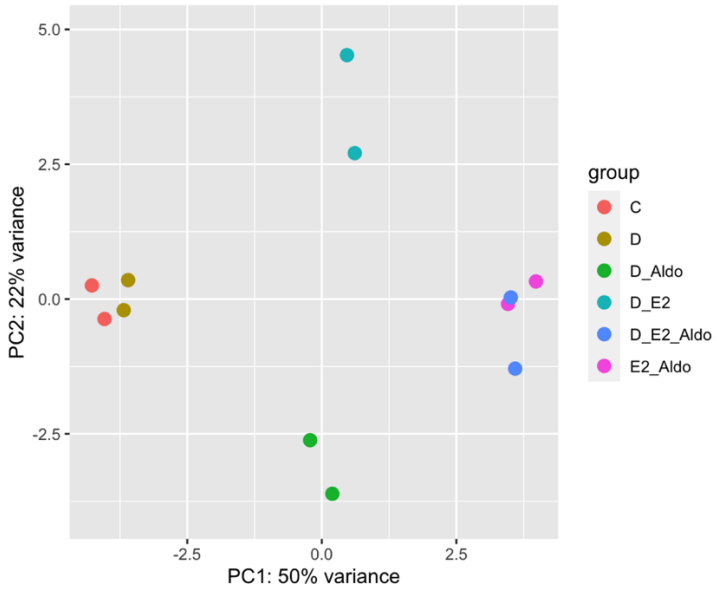


Figure 51. PCA result of MCF-7 RNA-seq data.

3.5.2 Analysis of MCF-7 Data in CAP-RNAseq

Following the removal of genes with zero or lowly expression levels in R, remaining genes were uploaded to CAP-RNAseq and underwent CPM filtering upon being with a minimum threshold of 0.1 CPM in at least 2 samples. Subsequent filtering was applied using vst+ANOVA with 0.05 cutoff for p-value.This resulted in 1,994 genes having notable differential expression across groups. These genes were then analyzed through cluster analysis, resulting in the identification of 16 distinct expression clusters (Figure 52). The mirror clusters were also identified using the dissimilarity heatmap and used in the analysis (Figure 53). However, due to the structure of the data, some clusters did not form a good mirror cluster.

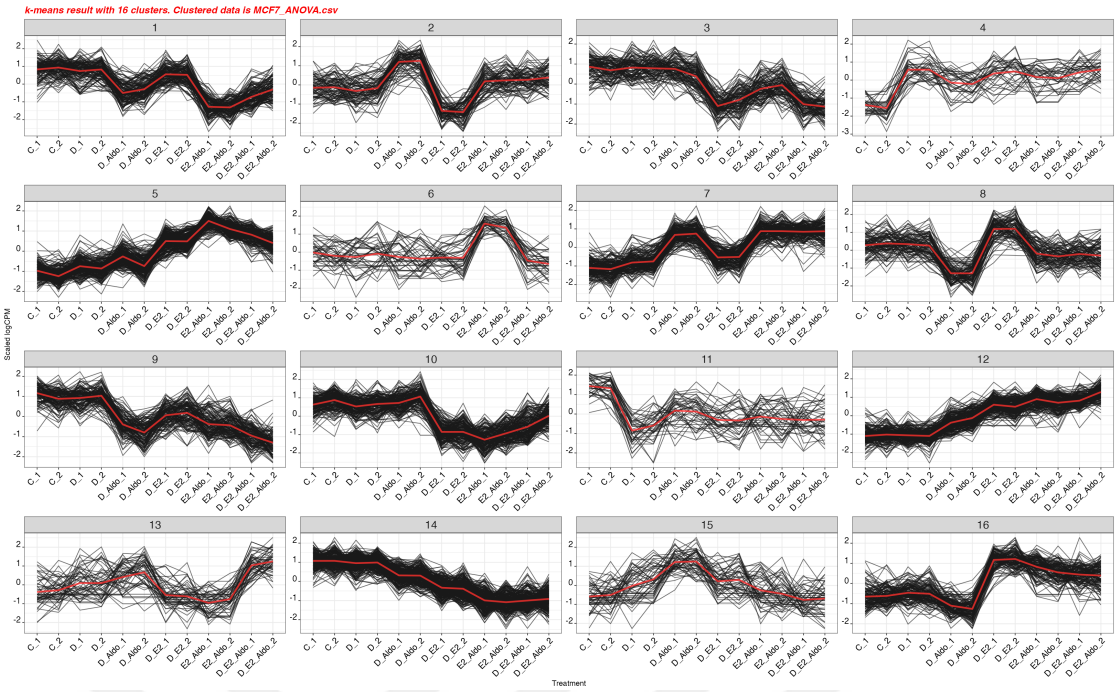
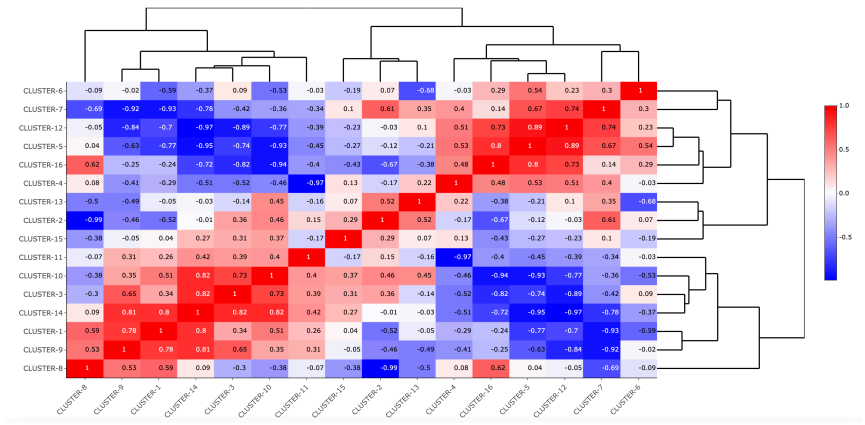


Figure 52. k-means clustering result of 1,994 genes in MCF-7 RNA-seq data with 16 clusters. Red lines indicating the cluster centroids show the cluster pattern.

A)



B)

Cluster	Mirror
CLUSTER-1	CLUSTER-7
CLUSTER-2	CLUSTER-8
CLUSTER-3	CLUSTER-12
CLUSTER-4	CLUSTER-11
CLUSTER-5	CLUSTER-14
CLUSTER-6	CLUSTER-13
CLUSTER-7	CLUSTER-1
CLUSTER-8	CLUSTER-2
CLUSTER-9	CLUSTER-7
CLUSTER-10	CLUSTER-16
CLUSTER-11	CLUSTER-4
CLUSTER-12	CLUSTER-14
CLUSTER-13	CLUSTER-6
CLUSTER-14	CLUSTER-12
CLUSTER-15	CLUSTER-16
CLUSTER-16	CLUSTER-10

Figure 53. Dissimilarity heatmap for 16 clusters to find the mirror clusters. A) Heatmap visualization. B) Table visualization of mirror clusters.

Most of the essential genes were accumulated in the cluster 5 followed by cluster 12, which both show increasing expression pattern from the first to the last sample in the

graphs (Table 30). Therefore, their mean of efficacy values was also higher compared to other clusters. On the other hand, less essential genes were found in cluster 11, followed by cluster 15 and 4. These clusters also contain a smaller number of genes and show variations between the replicates of samples. Genes in cluster 11 was generally downregulated by all samples and more by Drsp treatment. While cluster 4 is the opposite cluster of cluster 11, genes in cluster 15 was upregulated by only D_Aldo treatment. Moreover, most selective genes were found in cluster 8 and then cluster 9, which contained the genes downregulated by D_Aldo treatment.

Table 30. Essentiality information of clusters for MCF-7 dataset. SD: standard deviation; OR: Odds Ratio; FET: Fisher's Exact Test; p.val: p-value.

Clusters	Mean of Efficacy	SD of Efficacy	Mean of Selectivity	SD of Selectivity	Number of Essential Genes	Total Number of Genes	Percentage of Essential Genes	OR of FET	p.value of FET
Cluster 5	-1.05	0.44	0.04	0.18	121	211	57.35	3.97	0
Cluster 12	-1.03	0.47	0	0.16	77	177	43.5	2.27	0
Cluster 14	-0.66	0.27	0.19	0.25	41	240	17.08	0.61	0
Cluster 16	-0.77	0.36	0.05	0.12	34	117	29.06	1.21	0.38
Cluster 10	-0.62	0.4	0.04	0.23	27	150	18	0.65	0.05
Cluster 7	-0.7	0.36	0.07	0.24	26	154	16.88	0.6	0.02
Cluster 1	-0.57	0.19	0.14	0.24	24	118	20.34	0.75	0.27
Cluster 8	-0.61	0.2	0.35	0.96	15	78	19.23	0.7	0.28
Cluster 9	-0.71	0.23	0.27	0.32	15	118	12.71	0.43	0
Cluster 3	-0.66	0.28	0.15	0.27	14	89	15.73	0.55	0.04
Cluster 2	-0.6	0.2	-0.01	0.13	8	74	10.81	0.36	0
Cluster 13	-0.5	0.12	-0.02	0.1	8	38	21.05	0.79	0.71
Cluster 6	-0.71	0.33	0.01	0.13	7	34	20.59	0.77	0.69
Cluster 4	-0.69	0.46	0.09	0.14	5	26	19.23	0.7	0.65
Cluster 15	-0.64	0.18	0.2	0.19	3	36	8.33	0.27	0.02
Cluster 11	-0.69	0.34	0.15	0.4	2	29	6.9	0.22	0.03

3.5.3 Gene Set Enrichment Analysis of All Clusters

Due to variations in replicates of samples, some clusters did not show a great pattern such as clusters 5, 11 and 13. Therefore, the enrichment analysis using Hallmark was performed for all clusters at the same time before moving on to the analysis of individual clusters. This analysis was performed for essential and non-essential genes, separately (Figure 54).

Clusters 5 and 12 have a similar pattern of expression, both showing increasing expression from the first to the last sample in the graph. Therefore, the enrichment of essential genes in these clusters revealed in more common terms. This implies that the genes in clusters 5 and 12 may be functionally related or involved in similar regulatory networks, which are unfolded protein response, MYC targets and E2F targets. Distinctly, genes in cluster 12 was also enriched in genes linked to early and late estrogen response, emphasizing its participation in hormone-regulated cellular processes, whereas cluster 5 exhibits enrichment in the mTORC1 signaling pathway, suggesting a function in cell growth and metabolism. Moreover, 3 specific biological terms were found to be common across clusters 5, 12 and 16. Cluster 16, which contains genes that were increased in all samples containing E2, shared enrichment in MYC targets, E2F targets, and the G2/M checkpoint with clusters 5 and 12. These common terms suggests that the genes in these three clusters may be important for DNA repair, cellular proliferation, and regulation of cell cycle, which are all affected by MYC and E2F activity and are essential for the G2/M phase of the cell cycle. Moreover, genes in cluster 16 were also enriched in mitotic spindle. These clusters, in general, show an increase in D_E2 samples compared to the control. However, clusters 1, 9 and 11 exhibited decreasing of the genes' expressions in D_E2 samples compared to the control and did not share any common term. Genes in

cluster 1 were not affected by Drsp but showed decreased expression in all samples containing Aldosterone. While a slight decrease was also observed in D_E2 samples, possibly due to E2, higher decrease was shown in E2_Aldo samples and addition of Drsp to these two slightly reversed this decrease. Cluster 9 showed a similar pattern of decreased gene expression, but, in this case, the addition of Drsp to E2_Aldo did not reverse the decrease. Moreover, in cluster 11, while there was greater variability between replicates, the overall pattern was a similar reduction in all samples compared to control and a pronounced reduction in Drsp alone.

While genes in cluster 1 enriched in interferon alpha and gamma responses, suggesting a role of genes affected by Aldo in immune signaling and potentially antiviral responses, genes in cluster 9 enriched in TGF beta signaling indicating involvement in processes like cell growth, differentiation, and apoptosis. Finally, genes in cluster 11 were enriched in reactive oxygen species pathways, which are important for cell signaling and stress responses, suggesting the role of genes affected by Drsp.

Moreover, the enrichment of non-essential genes revealed that late response to estrogen was shared by clusters 12 and 16, while early response to estrogen was shared by clusters 7, 8, 12 and 16. Cluster 7 showed an overall increase in gene expression in all samples, while significant increases were observed in the samples containing Aldo. Cluster 8, on the other hand, showed decreased expression of genes in D_Aldo samples while showing increased expression in D_E2 samples. In Cluster 12, there is a general trend of ascending expression levels, peaking in the last sample, with the exception of samples treated with Drsp only.

The response to interferon alpha was shared by clusters 8 and 1, since genes in both clusters were downregulated by the samples containing Aldo. The mirror cluster of cluster 8, cluster 2 demonstrated an enrichment in hypoxia, indicating a response to low oxygen conditions. While cluster 9 showed an enrichment in interferon gamma response, cluster 3 showed the enrichments in UV response and coagulation, suggesting a role in DNA damage repair mechanisms and blood clotting processes, respectively.

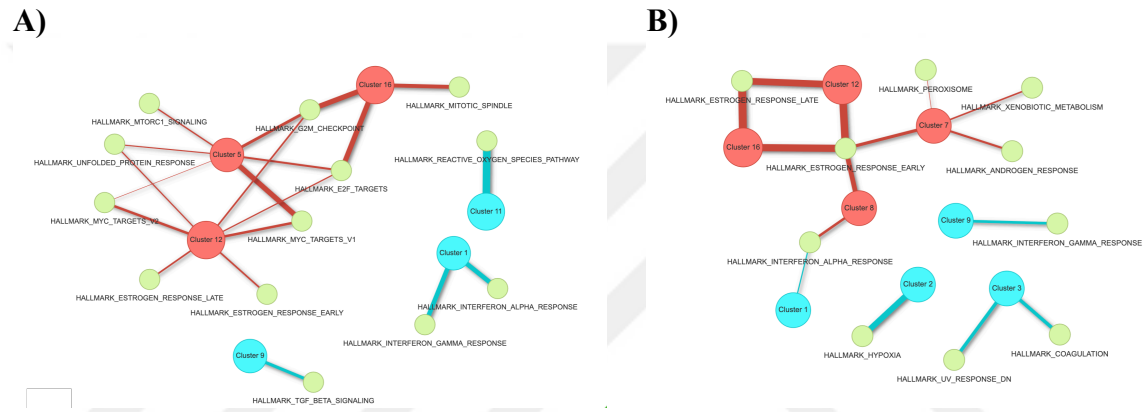


Figure 54. MSigDB analysis using Hallmark gene sets with q-value< 0.05 for top 10 terms. A) Only essential genes were used. B) Only non-essential genes were used. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between D_E2 and control samples.

3.5.4 Genes Affected by Drsp

Drsp treatment mainly affected genes in clusters 4 and 11, which were characterized as mirror clusters (Figure 55). These clusters had the lowest total number of genes and also contained the fewest essential genes. However, some of these essential genes had higher expression compared to other genes within the clusters.

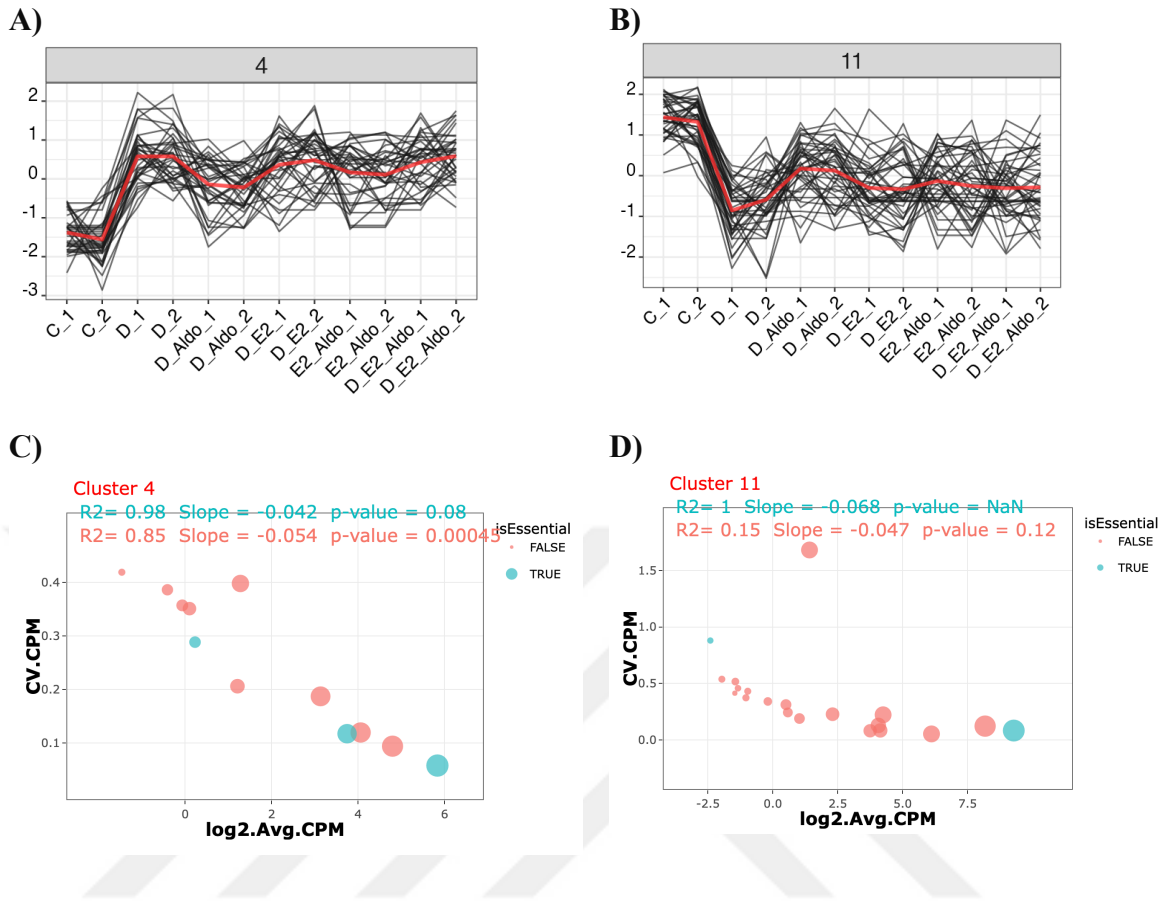


Figure 55. Analysis of clusters 4 and 11. Visualizations of cluster 4 (A) and cluster 11 (B). Scatter plots displaying the log2 of average CPM values in the x-axis and CV of CPM in the y-axis of genes in cluster 4 (C) and cluster 11 (D). The size of each dot displays a gene's log2 of range CPM value.

Among these genes, CHI3L2 was the essential genes that was upregulated by Drsp treatment in the cluster 4, while other treatments had no such effect (Figure 56A). The protein expression of this gene was also found at all amounts in breast cancer patients but not much found in normal breast tissues (Figure 56B, C).

ENO3 was another gene markedly upregulated only in response to Drsp treatment, with no such increase observed in any other treatments (Figure 56D). However, its protein expression was not detected in HPA dataset (Figure 56E, F).

From the cluster 11, CYP1B1 gene was prioritized since its expression was reduced by Drsp and D_Aldo treatments, while the reduction in the other samples was not so much (Figure 56G). This may simply be due to the effect of Drsp, which disappears when Drsp is given in combination with E2. In normal breast tissues, the protein level was found at low to moderate levels, while in breast cancer patients it was found at all levels (Figure 56H, I).

Another significant gene selected from cluster 11 is TXNRD1 (Figure 56J). It was highly expressed, essential, and selective in the cells and had moderate correlation between mRNA and protein levels (Figure 56K, L). TXNRD1 expression decreased in all samples, while a significant decrease was observed in the D_E2 sample. Protein expression was found at low levels in al normal breast tissues and medium levels in all breast cancer patients. In Hallmark gene set collection, it is found in reactive oxygen species pathway and MTORC1 signaling.

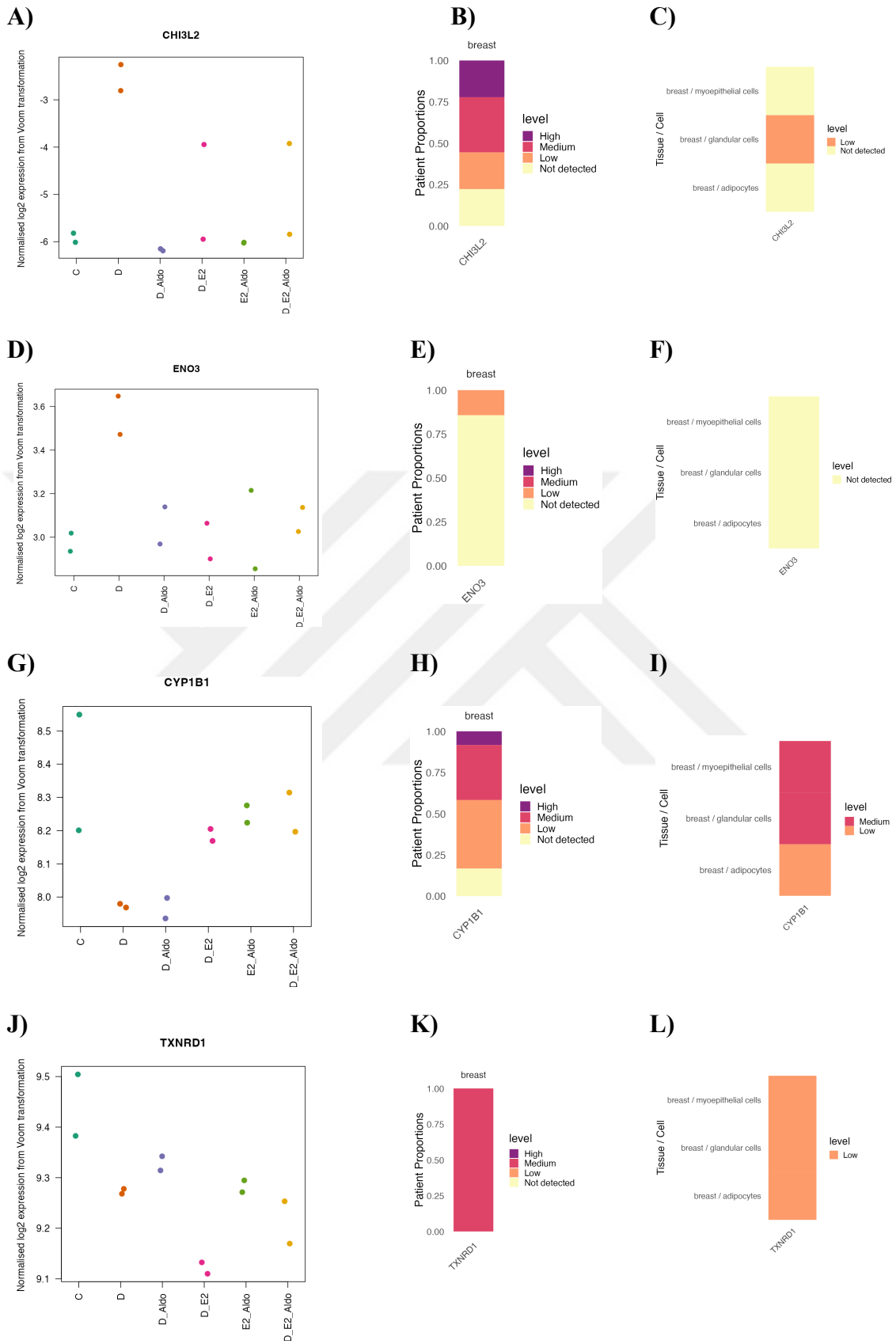


Figure 56. Log2 expression graphs and HPA results of CHI3L2 (A-B-C), ENO3 (D-E-F), CYP1B1 (G-H-I), and TXNRD1 (J-K-L) genes.

3.5.5 Genes Affected by Aldosterone

Genes in cluster 1 were downregulated in D_Aldo and E2_Aldo samples compared to control but most of them were slightly reversed in D_E2_Aldo samples. Based on MSigDB analysis, these genes were responsible in immune signaling. Among them, RELB was identified since it was essential and highly selective and correlated between RNA and protein expression levels (r=0.62). RELB appeared to be unaffected by Drsp and D_E2 alone but showed a significant decrease in D_Aldo and E2_Aldo. However, this reduction was slightly reversed when all hormones were given (Figure 57A). Interestingly, while this gene was found at high levels in all normal breast tissues, it was found at low levels in only half of breast cancer patients, while it was not detected in the other half (Figure 57B, C).

Another gene, PREX1, showed a decrease in samples containing Aldo (Figure 57D). However, more decrease was observed in the samples containing both E2 and Aldo. In HPA analysis, this gene was not detected in normal breast tissues, while low to medium expression levels were detected in cancer patients (Figure 57D, F).

On the other hand, cluster 9 contained genes that are upregulated by Aldo-containing samples. From this cluster BCL6 was prioritized which is an essential gene. It only affected by Aldo in MCF-7 cells showing upregulation (Figure 57G). In HPA data using breast cancer patients and normal tissues, its protein expression was not found in any (Figure 57H, I).

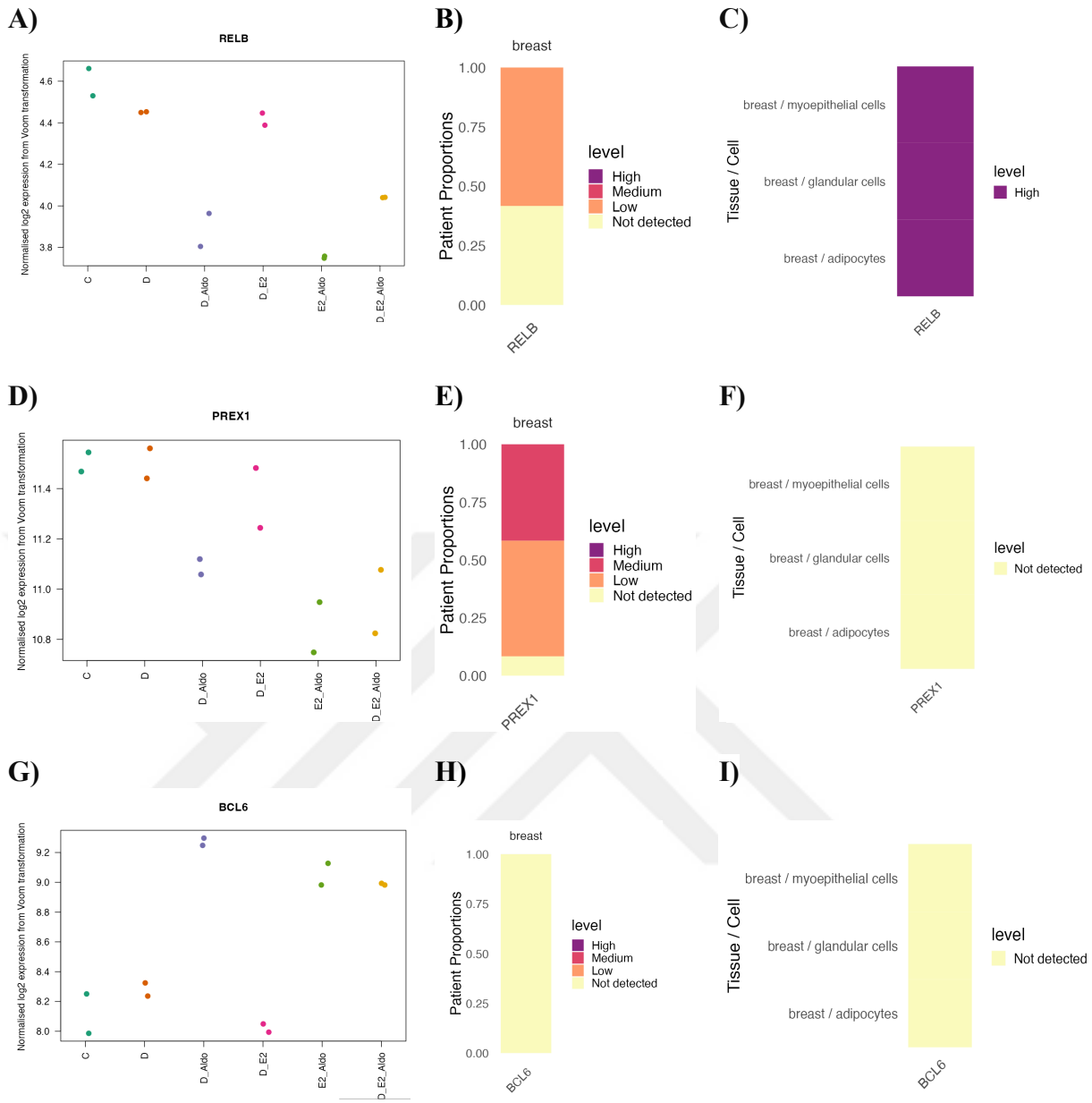


Figure 57. Log2 expression graphs and HPA results of RELB (A-B-C), PREX1 (D-E-F) and BCL6 (G-H-I) genes.

3.5.6 Genes Affected by E2

Most of the genes that were affected by the samples containing E2 were found in cluster 10 and 16. While genes that were downregulated in the samples containing E2 were placed in cluster 10, the upregulated genes were in the cluster 16. The enrichment analysis of these genes using Hallmark, REACTOME and GO revealed that genes of cluster 10 were mostly enriched in phosphatidylinositol and phospholipid binding pathways, which shows the role of these genes in lipid signaling and cellular membrane dynamics (Figure 58). In contrast, genes in cluster 16 were primarily enriched in pathways associated with estrogen response, Additionally, these genes were linked to the processes of chromosome localization, metaphase plate congression, mitotic nuclear division, and mitotic sister chromatid segregation, indicating their roles in cell division and chromosome segregation during mitosis.

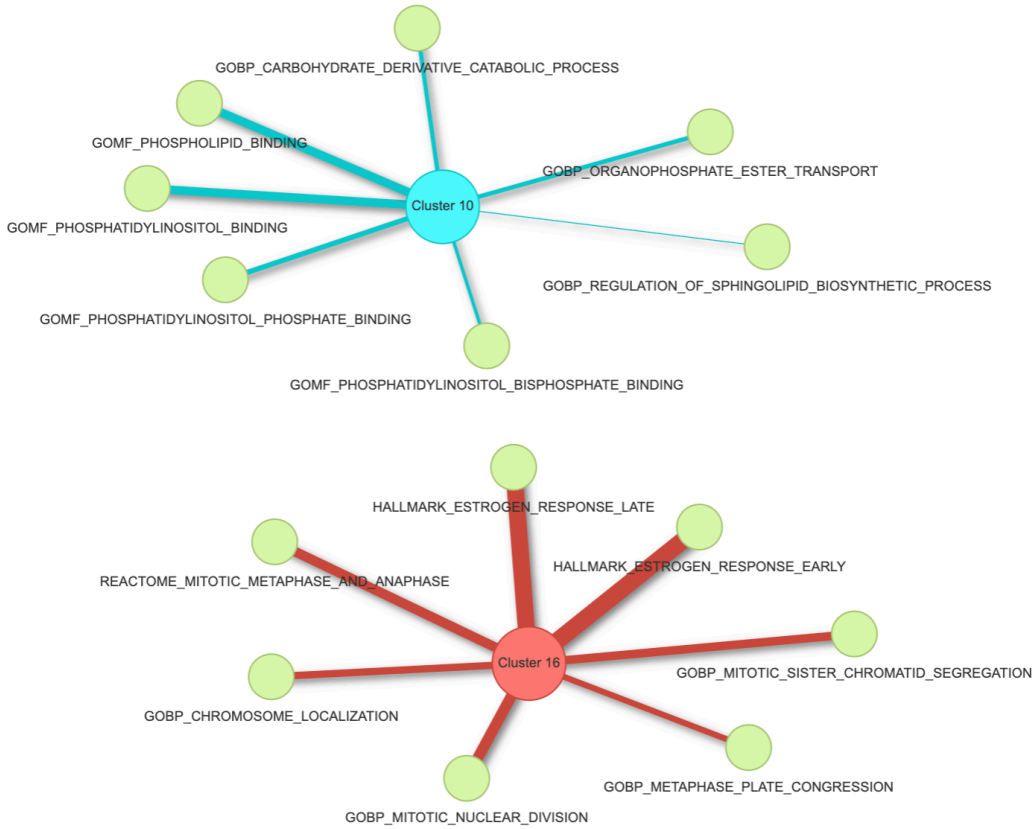


Figure 58. MSigDB analysis using Hallmark, REACTOME and GO gene sets with q-value< 0.1 for top 7 terms. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between D_E2 and control samples.

In cluster 16, CD44 exhibiting a strong correlation between mRNA and protein levels (r=0.85), displayed an increase in expression in E2 containing samples (Figure 59A). Its expression was also increased by Drsp alone treatment while no change was observed in D_Aldo treatment. Its protein expression was found at high levels in only glandular cells and but at different levels in breast cancer patients (Figure 59B, C).

PGR in cluster 16, which has high correlation between mRNA and protein levels (r=0.9), showed an increased expression in E2-containing samples, while it was slightly

downregulated by D_Aldo treatment (Figure 59D). Protein expression was found at moderate levels only in glandular cells, but at different levels in half of breast cancer patients (Figure 59E, F).

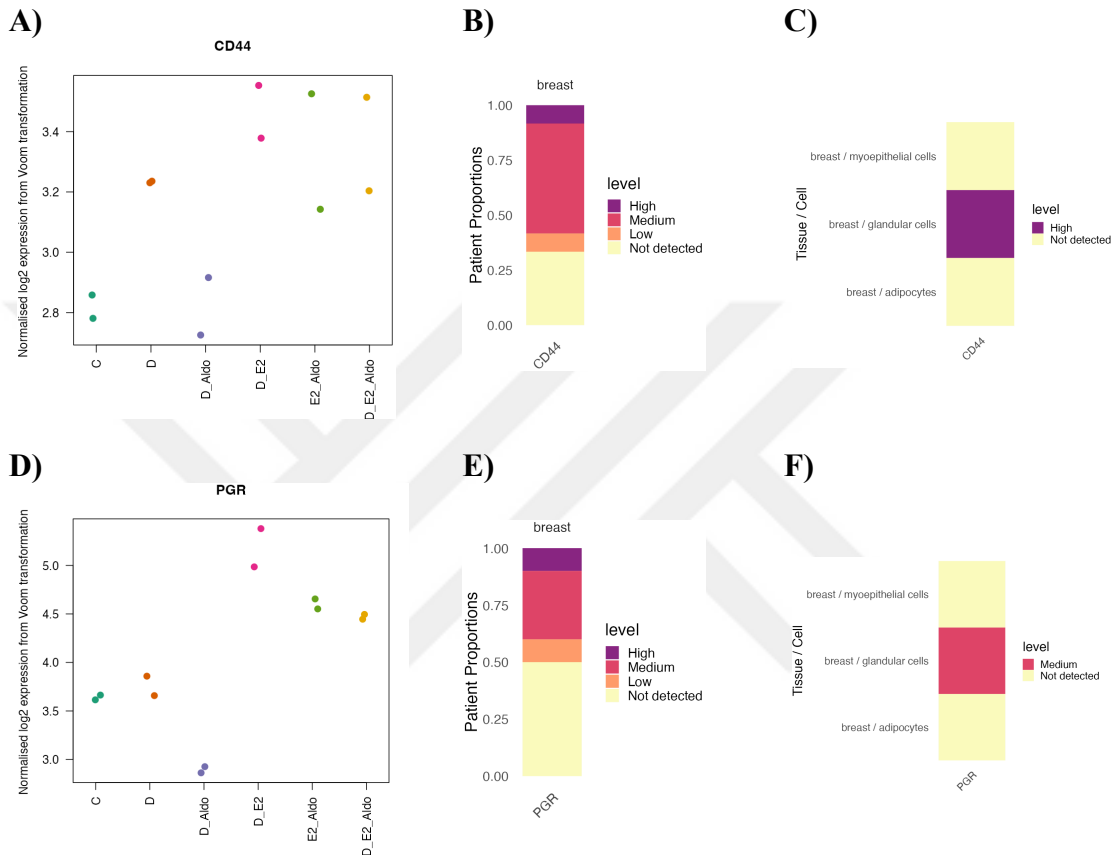


Figure 59. Log2 expression graphs and HPA results of CD44 (A-B-C) and PGR (D-E-F) genes.

3.5.7 Aldosterone effect on D_E2 Treatments

The analysis of cluster 11 revelaed that D_E2 treatment caused the most increase in the expression of genes. However, introduction of Aldosterone into this mix (DEA) reduced the effect of observed by D_E2 treatment (Figure 60A). The opposite pattern was seen in cluster 5, which was its mirror clusters (Figure 60B). In the enrichment analysis of these genes using essential genes by MSigDB, essential genes in cluster 11 were found to be involved in E2F targets, MYC targets, MTORC1 signaling and mitotic phases, which were related to cell cycle regulation, proliferation, and DNA replication (Figure 60C). In the opposite cluster, essential genes were enriched in Heme metabolism and cholesterol homeostasis.

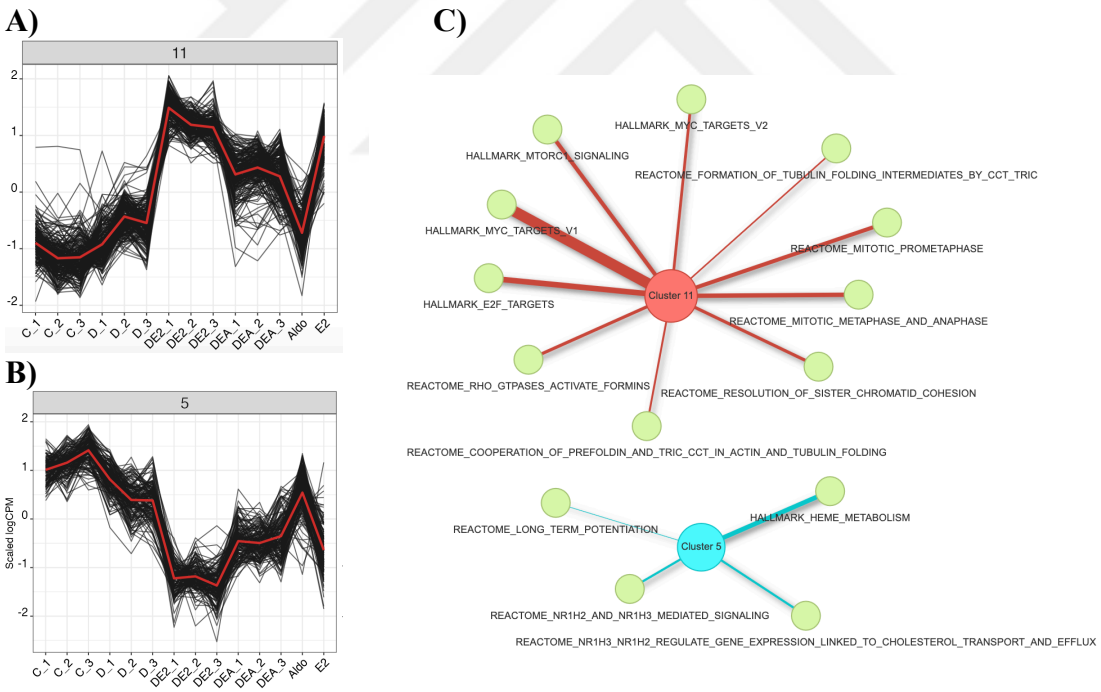


Figure 60. Analysis of mirror clusters, 11 and 5. A) The expression pattern of clusters 11. B) The expression pattern of clusters 5. C) MSigDB analysis for essential genes using Hallmark gene sets with q-value< 0.05 for top 10 terms. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between D_E2 and control samples.

3.5.8 Cross-Dataset Mapping for the Genes Increased by Drsp in T47D Data

There were 33 genes in cluster 9 of T47D data and those were upregulated in response to treatments containing Drsp, while no such effect was observed under the treatments either E2 or Aldo alone (Figure 61A). Therefore, these genes were searched for in the MCF-7 data and 16 of them were successfully mapped to cluster in the MCF-7 data, with 11 of these genes located in cluster 7 (Figure 61C). In contrast to their response in the T47D data, the expression of genes in cluster 7 of the MCF-7 dataset was not significantly affected by Drsp treatments (Figure 61B). Instead, these genes were predominantly upregulated in response to Aldo treatment in all samples containing Aldo, suggesting a differential response to hormonal treatments between the two cell lines. This also suggests that both Drsp and Aldo may have an effect on the expression of these genes, yet the impact varies between cell lines due to the diversity and abundance of receptors present in each cell type.

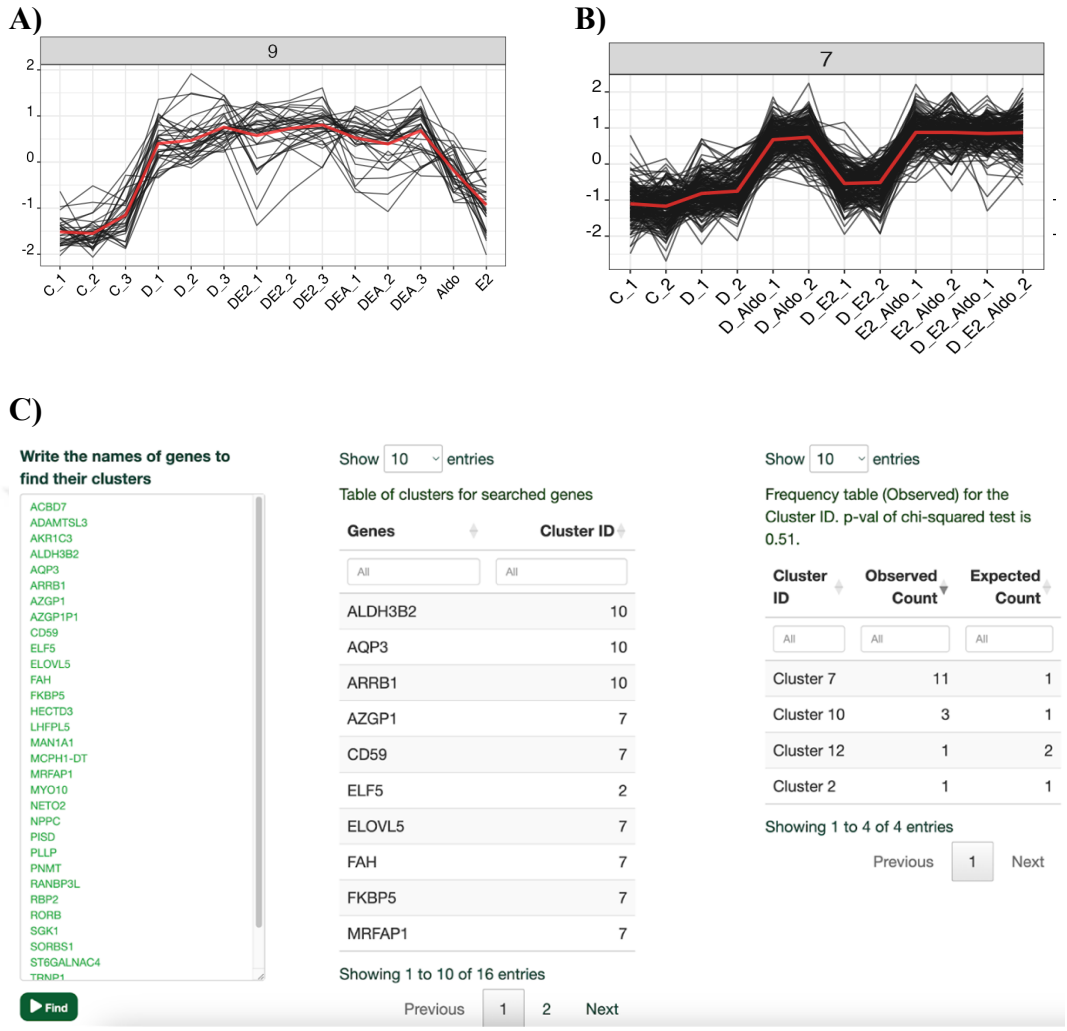


Figure 61. Cross dataset mapping. A) Visualization of cluster 9 of T47D data. B) Visualization of cluster 11 of MCF-7 data. C) The result of the cross-mapping when the genes of cluster 9 in the T47D data were searched in the MCF-7 data.

3.6 Hormone Studies in Breast Cancer Cell Lines

Transfection of MCF-7 cells was performed by transient transfection. Since the whole experiment lasted for 3 days, it was investigated how long the effect of transient transfection lasted. Even on the fourth day, there was still MR protein expression after serum deprivation was performed (Appendix A). Drsp and E2 are used in combination or alone in contraception as well as menopause. In order to use them in our wet lab studies, several preliminary studies were conducted to determine cell numbers and drug doses (Appendix B).

MTT assays were performed to determine the number of cells seeded in the plates that showed the best effect on cell viability when 10 nM and 20 nM E2 were used, and significant differences in cell viability were observed only in groups seeded with 5000 cells. Also, both E2 doses showed similar results on OV transfected MCF-7 cells. Furthermore, different (high and low) doses of Drsp were also investigated on the cell viability, alone or in combination with other hormones; E2 and Aldo. As a result of these studies, no significant difference in cell survival between low and high doses of Drsp alone was found on the OV transfected cells. In comparison to the control (EV), there was a noticeable increase in cell viability of OV transfected cells when 20 nM of Drsp and 100 nM of Aldo were used. This was not the case with a higher dose of Drsp (100 nM) and Aldo (100 nM); in fact, there was a modest decrease in the viability seen in this combination. Also, addition of E2 to this group (100 nM Drsp and Aldo) caused a significant increase in OV cells compared to EV cells, while no such effect was observed when E2 was used with 20 nM Drsp and Aldo.

Another set of preliminary study was done to investigate whether high-dose Drsp could counteract the effects of Aldo in both the presence and absence of MR expression (Appendix C). In this set, several genes were investigated by qRT-PCR. These findings showed that low-dose Drsp may show a more effective ESR1-E2 synergy or antagonism, and MR-expressing cells respond more effectively than EV-transfected cells. Therefore, 20 nM Drsp with 10 nM E2 concentration were chosen for further studies to mimic the hormone therapies used in the literature, as other doses did not cause significant and large changes.

3.6.1 Investigation of the Effect of Drsp in T47D Cells

Since T47D cells endogenously express MR, the effect of Drsp was also investigated directly in T47D cells. In order to examine the effect of Drsp dose, the expression of selected genes was analyzed with RNAs from T47D cells treated with 20 nM and 40 nM Drsp (Figure 62).

These samples were analyzed for SGK1, S100P, MR and CXCL12 genes. A statistically significant increase in the expression of SGK1 was observed and a higher expression bias was found at higher Drsp dose. One-way ANOVA results for SGK1 were as follows: $F(2,3)= 27.03$, $p=0.0121$. The expression of MR gene was also increased but due to the variation it was not found significantly ($F(2,3)=1.279$, $p=0.397$).

It has been reported that Drsp affects the expression of S100P and CXCL12 genes (Bray et al., 2005). The highly upregulated gene S100P encodes a calcium-binding protein associated with breast cancer and has been previously identified as a gene whose

expression is regulated by progesterone in T47D cells. Drsp and other progesterones increase the mRNA expression of this gene. CXCL12, a growth stimulatory cytokine, is known to be down-regulated by Drsp and other progesterones in T47D cells but induced by estrogen in MCF-7 breast cancer cells (Boudot et al., 2011). The difference between 20 nM and 40 nM Drsp doses was significant only in the S100P gene, suggesting that 40 nM dose may be more effective for this gene ($F(2,3)=32.67$, $p=0.0092$). The expected decrease trend in the expression of CXCL12 was observed in T47D cells, but it was not significant due to the variation ($F(2,3)=0.094$, $p=0.913$). There were also other genes analyzed by qRT-PCR in this set (Appendix D). However, they do not show significant changes between two doses of Drsp.

In conclusion, studies with 20 nM and 40 nM Drsp doses showed that there was not much difference between the effects of the two doses. Therefore, the remaining experiments were continued with the 20 nM Drsp dose.

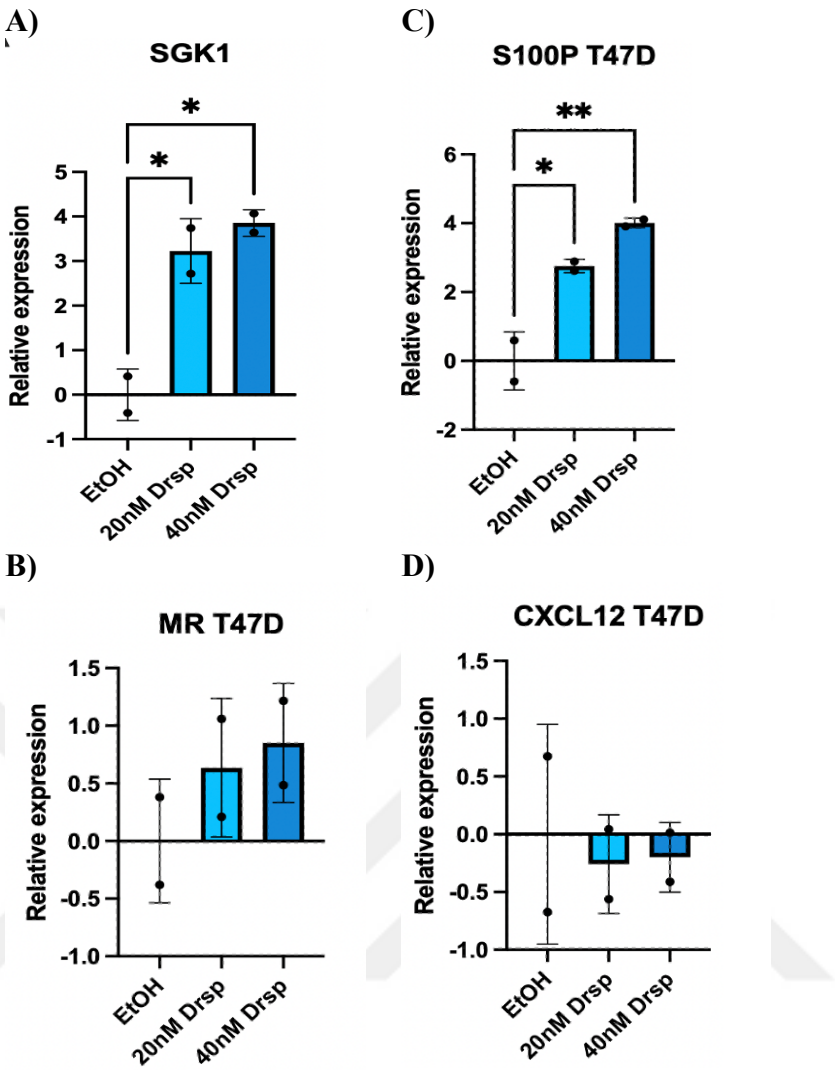


Figure 62. qRT-PCR results with RNAs from T47D cells. A) SGK1, B) MR, C) S100P and D) CXCL12 genes. One-way ANOVA with Tukey's multiple comparisons test was used (*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$, ****:$p<0.0001$).

In T47D cell line samples obtained by using Drsp, Aldo, E2, Drsp _E2, Drsp_Aldo, Drsp_Aldo_E2, Aldo_E2 combinations, mRNA expressions of genes previously examined in MCF7 cells were also examined by qRT-PCR (Figure 63). The PGR gene was analyzed with samples in which all combinations were used. Although the variation was high, a trend of increased expression was observed in all groups containing E2 ($F(7,9)= 2.933$, $p=0.0678$).

ZBTB16 and S100P genes were also analyzed to confirm Aldo and Drsp function, respectively. Significant increases were found in both genes, confirming that the given drugs were working. ANOVA results for these genes were as follows: F(2,3)=34.97, p=0.00293 for ZBTB16 and F(2,3)=35.48, p=0.00817 for S100P.

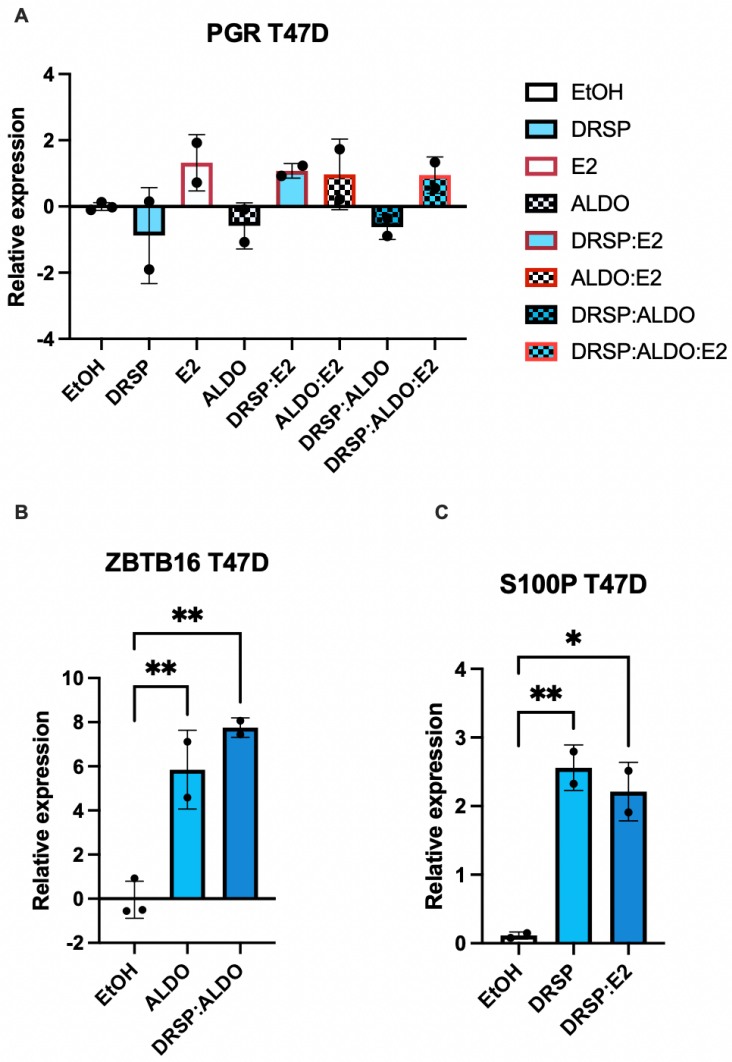


Figure 63. Checking the expression of PGR, ZBTB16 and S100P genes by qRT-PCR with RNAs from T47D cells. One-way ANOVA with Tukey's multiple comparisons test was used(*: p<0.05, **: p<0.01, ***: p<0.001, ****:p<0.0001).

3.6.2 Validation of Genes

Previous findings showed that 20 nM Drsp can be as effective as 40 nM, and OV-expressing cells respond more effectively than EV cells. Therefore, 20 nM Drsp was continued to be used with 10 nM E2. Validations of genes by qRT-PCR were conducted in MCF-7 cells, which were also sequenced. The results were normalized to the OV-EtOH control group and then normalized to the TPT1 gene used as a reference gene within the scope of delta-delta Ct and tested by one-way ANOVA (followed by Tukey's multiple comparisons tests).

The changes in some of the genes were compared in both MCF-7 and T47D RNA-seq data in R. Table 31 shows the logFC and p-values of these genes after DESeq2 analysis. Except for SGK1, ZBTB16, and HSD11B2 genes, the other genes were not affected by Drsp alone in MCF-7 cells, but the number of affected genes was higher in T47D cells. Similarly, the effect of the Drsp_E2 combination is higher in T47D than in MCF-7 cells.

In Figure 61D, the decrease in the CXCL12 gene observed with 20 nM and 40 nM Drsp in T47D cells was also observed with Drsp alone in T47D cells in RNA-seq. However, this decrease was increased with Drsp_E2 in T47D. A significant and large increase in the SGK1 gene (Figure 61A) and non-significant increases in the NEDD4L gene and HSDB12B2 genes (Figure 87) were observed with 20 nM and 40 nM Drsp in T47D cells. In the RNA-seq results, the SGK1 gene showed increases with Drsp alone and Drsp_E2 in both MCF-7 and T47D cells. NEDD4L gene showed a similar level of increase with Drsp and Drsp_E2 alone in the RNA-seq results of T47D cells. In the HSD11B2 gene, while Drsp alone showed an increase, it increased even more with Drsp_E2 in T47D cells. In the RNA-seq of MCF-7 cells, the increase in the HSD11B2 gene caused by Drsp alone

and Drsp_E2 is at the same level. This suggests that E2 has no effect on this gene in MCF-7 cells.

Table 31. LogFC and p-values of Drsp and Drsp_E2 samples of MCF-7 and T47D RNA-seq data using compared to control.

	MCF-7 data				**T47D data**			
	Drsp vs. Control		**Drsp_E2 vs. Control**		**Drsp vs. Control**		**Drsp_E2 vs. Control**	
Genes	**logFC**	**p-value**	**logFC**	**p-value**	**logFC**	**p-value**	**logFC**	**p value**
AREG	-0.001	0.997	1.218	0	0.65	0	1.704	0
BAX	0.128	0.352	0.158	0.252	0.048	0.681	0.182	0.117
BCL2	-0.077	0.461	0.081	0.437	-0.688	0.009	0.486	0.055
CCND1	-0.089	0.344	0.073	0.439	-0.205	0.041	0.147	0.143
CCNE2	-0.022	0.887	-0.17	0.264	-0.538	0	-0.254	0.003
CHEK1	-0.115	0.355	0.099	0.425	-0.06	0.443	0.333	0
CLDN1	-0.025	0.82	-0.148	0.171	-0.292	0.003	-1.042	0
CXCL12	0.154	0.106	0.598	0	-1.134	0	1.528	0
GATA3	0.035	0.682	-0.138	0.111	0.016	0.818	-0.304	0
HSD11B2	0.668	0	0.676	0	0.442	0.007	1.632	0
NEDD4L	-0.09	0.491	-0.416	0.002	0.508	0	0.696	0
NR3C1 (GR)	-0.078	0.485	-0.058	0.606	0.109	0.404	-1.024	0
NR3C2 (MR)	0.252	0.034	0.17	0.154	-0.521	0.002	-0.546	0.002
PER1	-0.057	0.732	-0.132	0.431	0.21	0.184	-0.111	0.486
PGR	0.125	0.442	1.56	0	-0.5	0	1.019	0
RPL37	0.076	0.621	0.056	0.717	-0.15	0.216	0.294	0.015
RRM2	-0.047	0.688	0.101	0.39	-0.184	0.002	0.451	0
SGK1	1.572	0	1.444	0	3.412	0	4.062	0
TFF1	0.066	0.632	0.778	0	-5.389	0.001	1.405	0.234
ZBTB16	4.853	0.002	4.875	0.002	10.812	0	10.127	0

3.6.2.1 qRT-PCR Results in MCF-7 Cells

In this set of experiments, the increase in PGR was tested as a positive control of the E2 effect, and the increase in ZBTB16 was tested as a positive control of Aldo treatment whereas MR expression was not expected to differ between groups since only OV transfected cells were used (Figure 64). It was observed that Drsp alone did not affect PGR, but Drsp_E2 increased PGR expression, and this was also observed in RNA-seq results. In all samples containing E2 treatment, PGR expression was found to be increased compared to EtOH control. In between-group comparisons, significant increases were observed in the Drsp_E2 sample vs. EtOH control ($p=0.0453$), Drsp_Aldo vs Drsp_E2 samples ($p=0.0230$), and Drsp_Aldo vs Aldo_E2 samples ($p=0.0297$) (Figure 64A). However, in RNA-seq result, Drsp alone caused a very small decrease in PGR in T47D cells, which was observed in Figure 63 (not significant).

For the MR gene, no significant differences were observed between groups (Figure 64B), indicating that MR gene expression at the mRNA level was not altered by the doses of Drsp_E2 administered and in the presence or absence of Aldo. This result suggests that MR expression may be altered at the protein level but not at the mRNA level.

The ZBTB16 gene, which is the target of MR and GR, showed significant increases in the presence of Aldo (Figure 64C). ZBTB16 expression was similar in Drsp alone and Drsp_E2 samples, which is also observed in the RNA-seq results of both cell lines and increased with the addition of Aldo or Aldo_E2 to Drsp ($p<0.0001$). It can be said that the increase in ZBTB16 was significantly increased by Aldo but to a lesser extent by Drsp or E2 without Aldo.

Moreover, no significant change was observed in TFF1 gene expression which is ESR1-target (Figure 64D). It was thought that the E2 treatment was effective but not sufficient to increase TFF1, although it increased PGR.

Glucocorticoid receptor (GR) activation has previously been associated with ER-negative breast cancer (West et al., 2018). Aldosterone acts through MRs and GRs, which are structurally similar to each other. Accordingly, changes in GR expression were also examined. Drsp caused a decrease in the absence of Aldo and neutralization in the presence of Aldo. This indicates an antagonistic relationship between Drsp and Aldo and supports the idea that GR expression is increased by Aldo and decreased by Drsp. Accordingly, a significant increase was observed between Drsp and Drsp_Aldo, Drsp and Drsp_Aldo_E2 samples ($p=0.0254$ and $p=0.0197$, respectively) (Figure 64E). In conclusion, the presence of Aldo in GR increased its expression.

Another analyzed gene, PER1, which showed a non-significant decrease only in the presence of Drsp and E2, showed significant increases in the presence of Aldo (Figure 64F). However, the fact that it did not increase in the Aldo_E2 sample may suggest that E2 has a reducing effect. The differential interaction of Drsp with Aldo and E2 can be observed by changes in PER1 gene expression. Accordingly, a significant difference was observed between Drsp_Aldo and Drsp_E2, and between Drsp_Aldo_E2 and Drsp_E2 samples ($p=0.0302$ and $p=0.0211$, respectively). These findings support that there is an interaction between MR, GR, and ESR1 and that Aldo and E2 can interact with each other, indicating that the 20 nM Drsp:10 nM E2 doses have activity at MR and GR target genes.

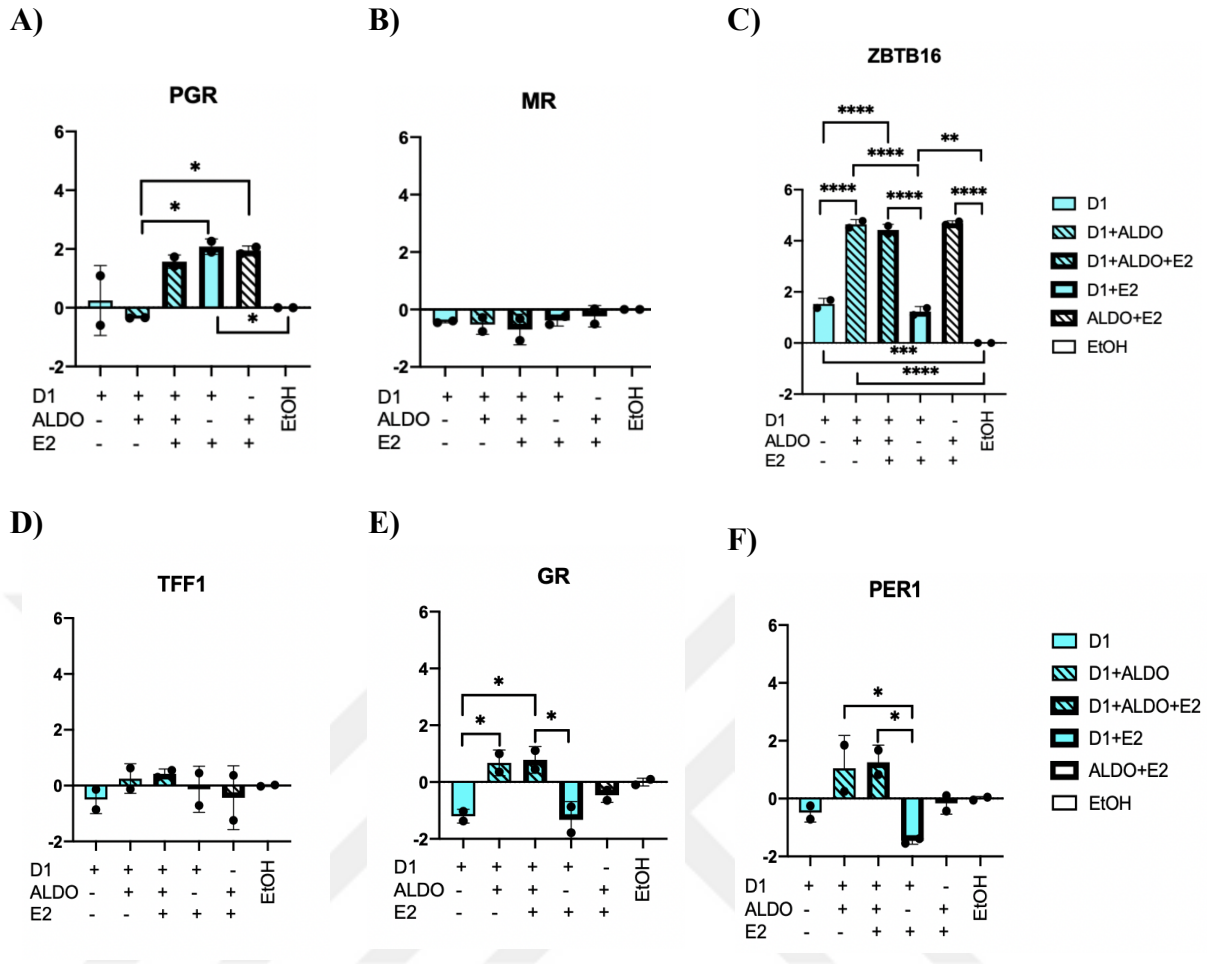


Figure 64. qRT-PCR results with RNAs collected from MCF-7 cells transfected with OV alone and exposed to 5 different drug combinations. (*: p<0.05, **: p<0.01, ***: p<0.001, ****:p<0.0001). ANOVA results are as follows: F(5,6)= 8.77, p=0. 00994 for PGR; F(5,6)= 1.184, p=0. 415 for MR; F(5,6)= 286.9, p= 4.72e-07 for ZBTB16; F(5,6)=0.633, p= 0.684 for TFF1; F(5,6)=10.01, p= 0.0071 for GR; and F(5,6)=6.408, p= 0.0213 for PER1.

Although PGR expression increased with E2, no effect was observed in ESR1 or other E2 target genes (CCND1, CCNE2, and RRM2). There was no significant difference in the qRT-PCR results of these genes (data not shown). There was a trend towards an increase in primary targets such as CCND1, but it was not statistically significant. Furthermore, the effects on genes related to DNA damage and apoptosis (CHEK1, BAX,

and BLC2) were also examined, but no significant changes were observed (data not shown).

3.6.2.2 Western Blot Results in MCF-7 Cells

Moreover, in the experimental setup in which qRT-PCR studies were performed, the pellet was taken from the cells with the same passage and same treatment, and protein was extracted for the Western Blot studies to be performed. The proliferative and apoptotic effects of 20 nM Drsp:10 nM E2 treatment were examined at the protein level, and changes in MR and GR levels were also tested (by Burçin Irem Arıcı). One-way ANOVA with Dunnett's multiple comparisons test was performed and significant effect was observed in only MR gene ($F(5,6)=17.64$, $p=0.00158$). The presence of MR expression in the control group indicates that it is due to OV treatment. MR expression increased, especially with Drsp_E2 treatment (Figure 65). Accordingly, the unique finding that E2 and Drsp can synergistically increase MR protein expression in MR overexpressing cells was revealed. In contrast to this increase, Aldo treatment decreased MR expression, indicating for the first time in the literature that Drsp and E2 may have opposite effects on MR expression in an Aldo-dependent manner (Figure 65A, B). In the case of GR, the paralogous protein of MR, the negative effect of Aldo was observed only in combination with E2, while Drsp alone or in combination with Aldo tended to increase the amount of GR but not to statistical significance. Furthermore, there was a downward trend in CHEK1 and an upward trend in BCL2/BAX proportional proteins compared to the EtOH group, but they were not statistically significant (Figure 65B, C).

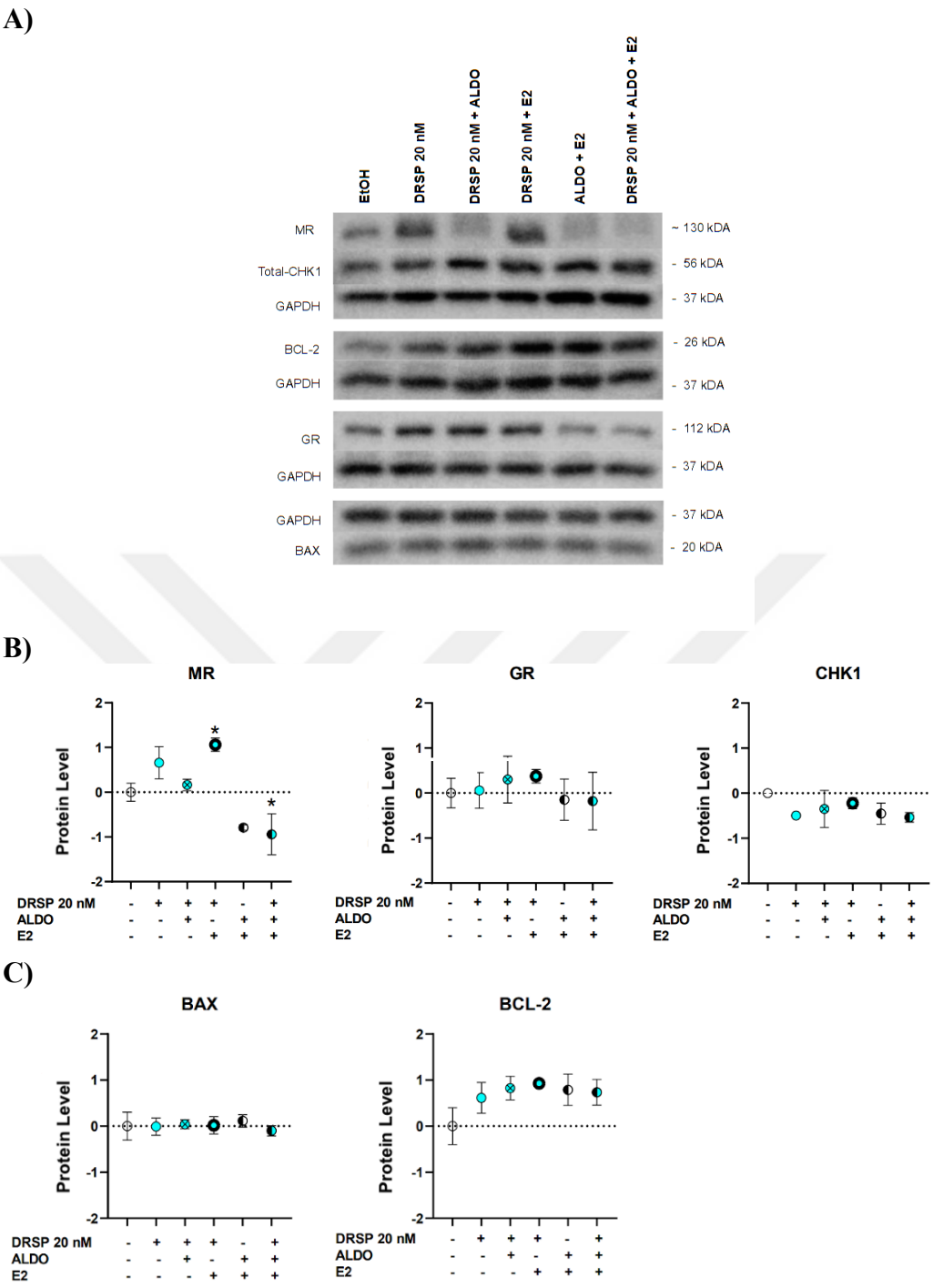


Figure 65. Effect of Drsp, Aldo, E2, and combination treatment on protein levels of MCF-7 cells transfected with MR expressing vector (OV) for 24 hours. A) Western blot images of the protein levels of the groups after the application and GAPDH protein used as loading control and B) MR, GR, CHK1, C) BAX, and BCL-2 protein levels. (* : $p<0.05$)

3.6.3 Comparison of Genes in both MCF-7 and T47D Cell Lines

Some of the genes prioritized in CAP-RNAseq were selected to validate with qRT-PCR in both cell lines (Figure 66). FKBP5 gene was significantly upregulated by Aldo in MCF-7 RNA-seq data, while Drsp had little effect on it. A similar result was observed in qRT-PCR validation, but only the E2_Aldo sample failed to show the same increase in qRT-PCR as in RNA-seq. However, FKBP5 was significantly affected by Drsp in T47D cells, which showed similar results in both qRT-PCR and RNA-seq. In T47D cells, E2 and Aldo alone samples had also increased effect on the gene, but Drsp caused more increase.

The NDRG1 gene appears to be affected by Aldo in MCF-7 RNA-seq data but increases its expression in qRT-PCR only with the Drsp_Aldo sample, while it appears unaffected by DEA and E2_Aldo (Figure 67). In T47D RNA-seq data, its expression was increased with Drsp alone and decreased with E2 alone, while slight decreases were observed in DEA and D_E2 sample and no effect was observed by Aldo alone. However, its qRT-PCR result showed that Aldo also decreased it, while D_E2 and DEA samples had no effect. The increase caused by Drsp alone and decreased by E2 alone were also confirmed with qRT-PCR.

Lastly, BCL6 was identified in MCF-7 cells as upregulated by Aldo, which was also confirmed by qRT-PCR (Figure 68). However, in T47D cells, Drsp caused an increase in its expression, while Aldo seems to be not affected.

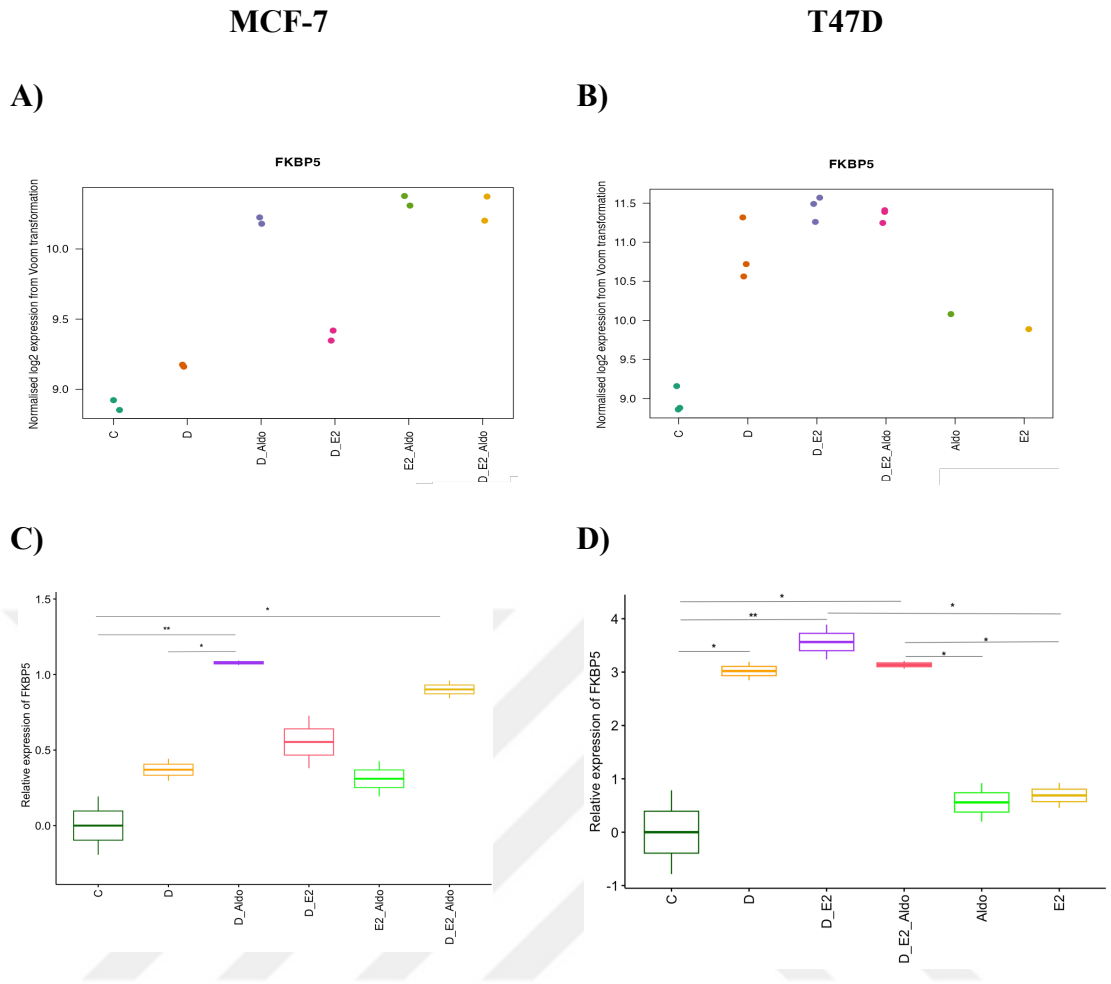


Figure 66. Expression pattern of FKBP5 and its validation results in qRT-PCR. One-way ANOVA with Tukey's multiple comparisons test was used. ANOVA results in MCF-7 cells are as follows: : $F(5,6)=10.58$, $p=0.00616$. ANOVA results in T47D cells are as follows: $F(5,6)=15.78$, $p=0.00213$. (*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$, ****:$p<0.0001$).

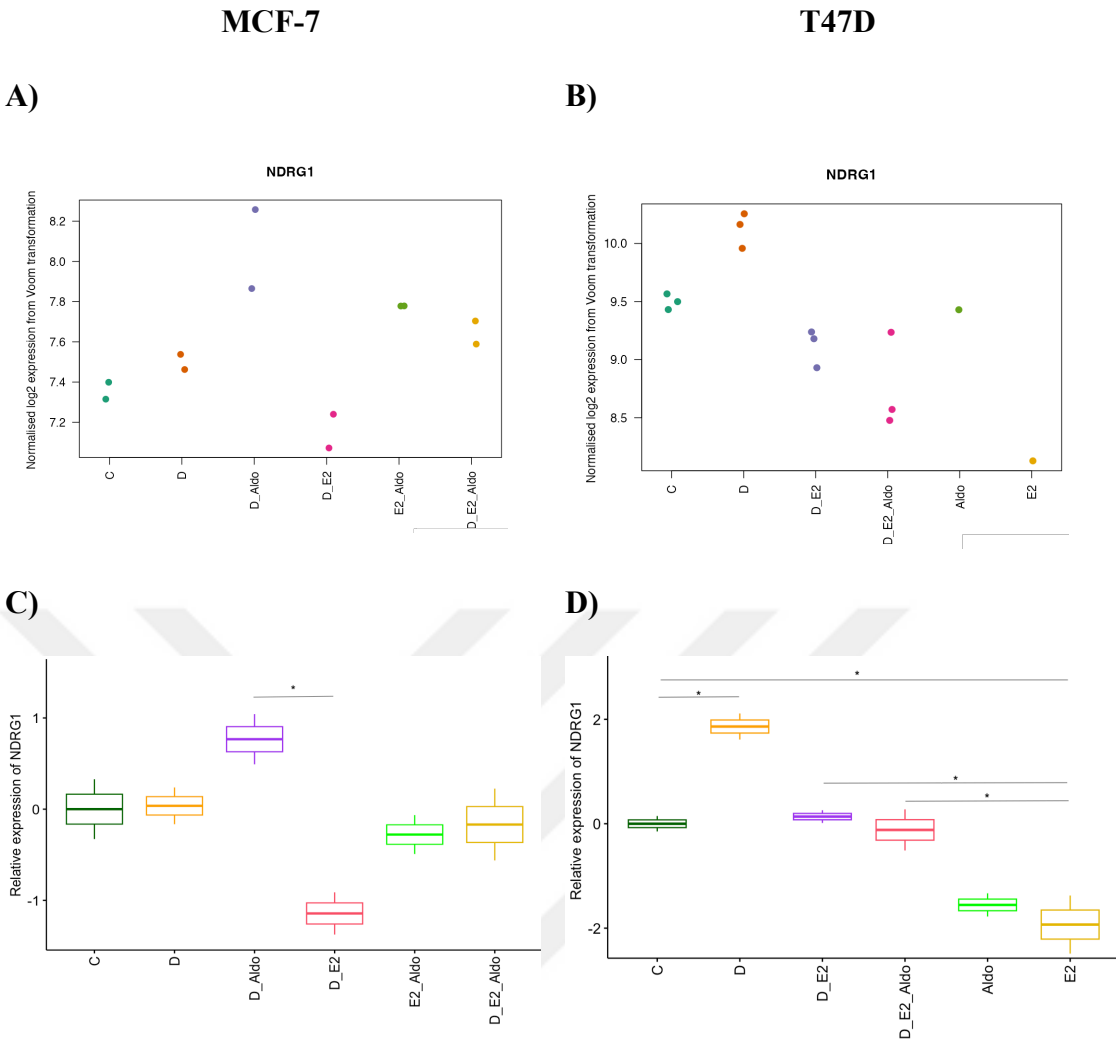


Figure 67. Expression pattern of NDRG1 and its validation results in qRT-PCR. One-way ANOVA with Tukey's multiple comparisons test was used. ANOVA results in MCF-7 cells are as follows: : F(5,6)=4.771, p=0.0417. ANOVA results in T47D cells are as follows: : F(5,6)=18.03, p=0.00148. (*: p<0.05, **: p<0.01, ***: p<0.001, ****:p<0.0001).

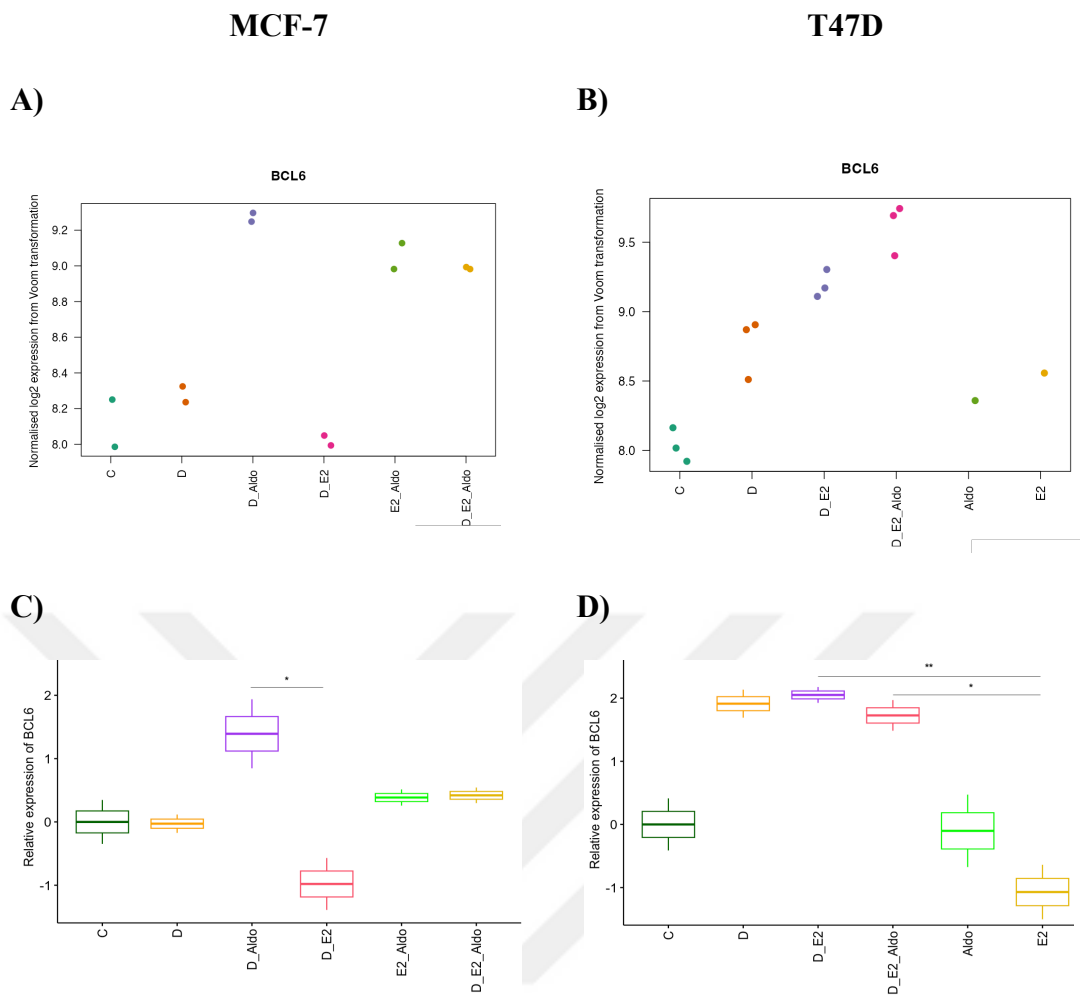


Figure 68. Expression pattern of BCL6 and its validation results in qRT-PCR. One-way ANOVA with Tukey's multiple comparisons test was used. ANOVA results in MCF-7 cells are as follows: : F(5,6)=5.619, p=0.0289. ANOVA results in T47D cells are as follows: : F(5,6)=12.75, p=0.00378. (*: p<0.05, **: p<0.01, ***: p<0.001, ****:p<0.0001).

3.6.4 ChIPseq Analysis

The ChIP assay was performed in collaboration with Dr. Serap Erkek at IBG Izmir, and sequencing was performed at Macrogen (see Methods in Appendix E). In the ChIP experiment, it was focused on how estrogen receptor binding sites (EREs) are altered in the presence of E2 and Drsp with the ESR1 antibody.

After ChIP sonication, uncut chromatin samples and input samples were run on an agarose gel to check whether the chromatin was sonicated as 200-500bp fragments. As shown in Figure 69, after the fragments were found to be the desired size, the experiment continued, and ChIP DNAs were obtained.

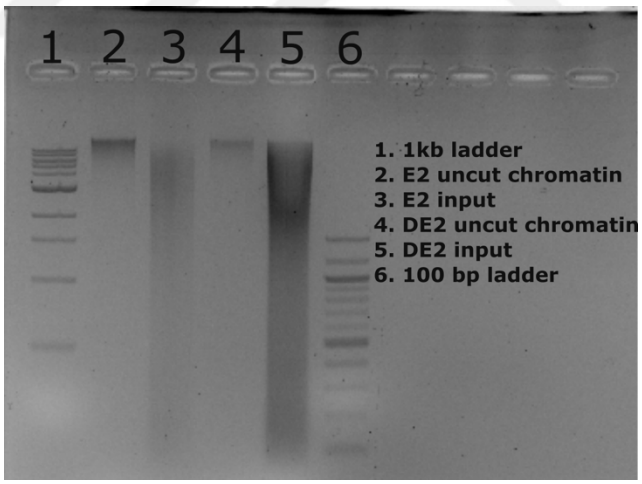


Figure 69. Uncut and input chromatin samples run on the agarose gel.

Before submission for sequencing, ChIP experiments were checked by qPCR using GREB1 and PGR genes. The ChIP-qPCR result was analyzed by both the percent input method (Figure 70A, B, C) and fold-enrichment method (Figure 70D). As expected, E2-treated samples showed an increase in both GREB1 and PGR. TPRG1, which was used

as a positive control, showed a high increase in all samples. In this way, we can see that the ChIP experiments were performed correctly.

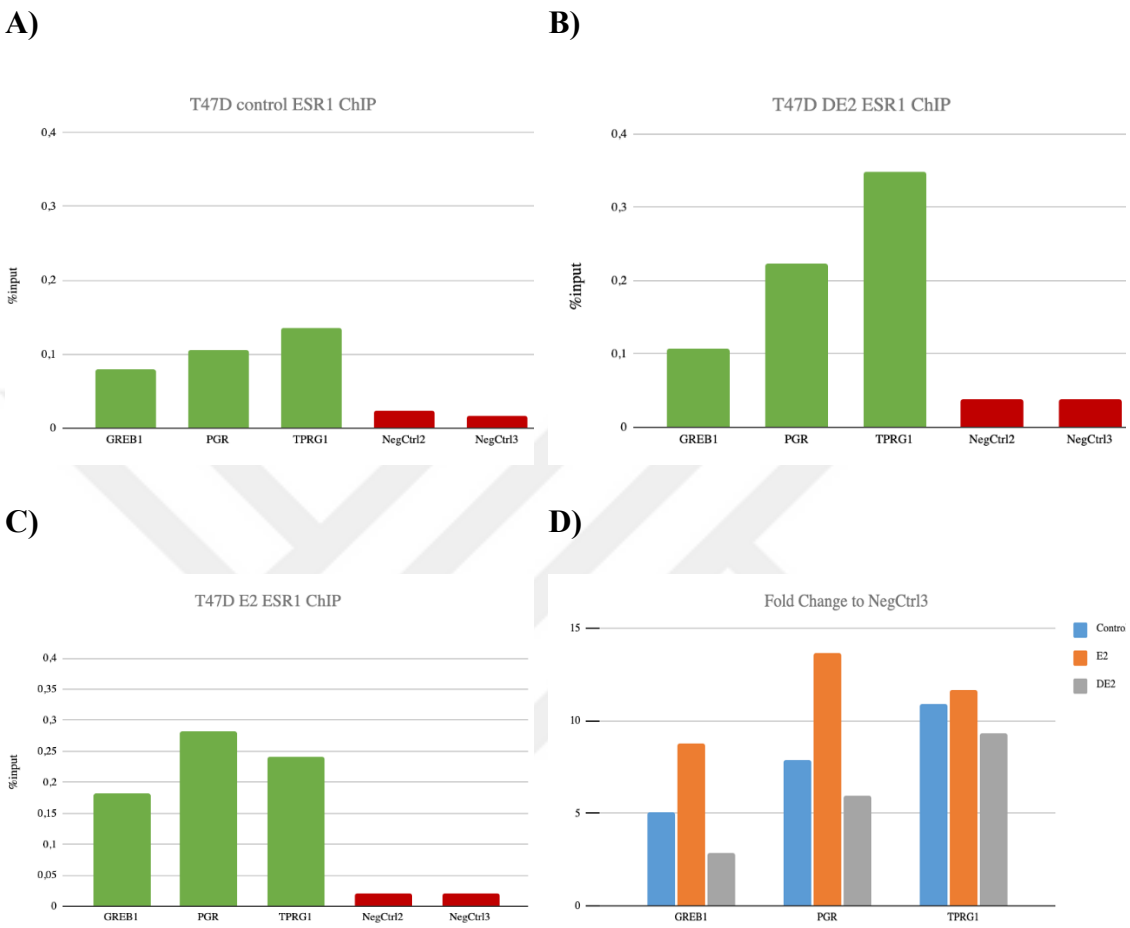


Figure 70. qPCR analysis by Percent input method for A) Control sample, B) DE2 sample, C) E2 sample. D) by Fold Enrichment method normalized to negative control 3.

After we received the raw data, a quality check of the fastq files was performed in Galaxy. Figure 71 shows the results of the examples.



Figure 71. Quality control results of all samples checked at Galaxy.

Peak calling was performed as both narrow and broad peaks using input normalization. When the narrow peaks were analyzed in IGV, the desired peaks were not seen in the expected genes (such as PGR, GREB1), so the analysis was continued with broad peaks. After the peak call, the covplot function of the ChIPseeker package (G. Yu et al., 2015) was used to see the peak positions over the whole genome. Thus, the extent of peak regions on chromosomes is calculated and a figure is generated to visualize it (Figure 72A, B, C). Although the control and E2 peaks can be seen more clearly, the peaks appear to be more intense in the DE2 sample due to the scarcity of high-scoring peaks. However, the high scores found in control and E2 are absent in DE2. The peaks of these three samples were plotted on a Venn Diagram and the overlapping peaks were analyzed. The E2 sample (49437) contains more peaks than DE2 (21435) and the control (8480) (Figure 72D). In the overlapping peak analysis, the number of common peaks was lowest between the control and DE2 samples (1834) and highest between E2 and DE2 (10388).

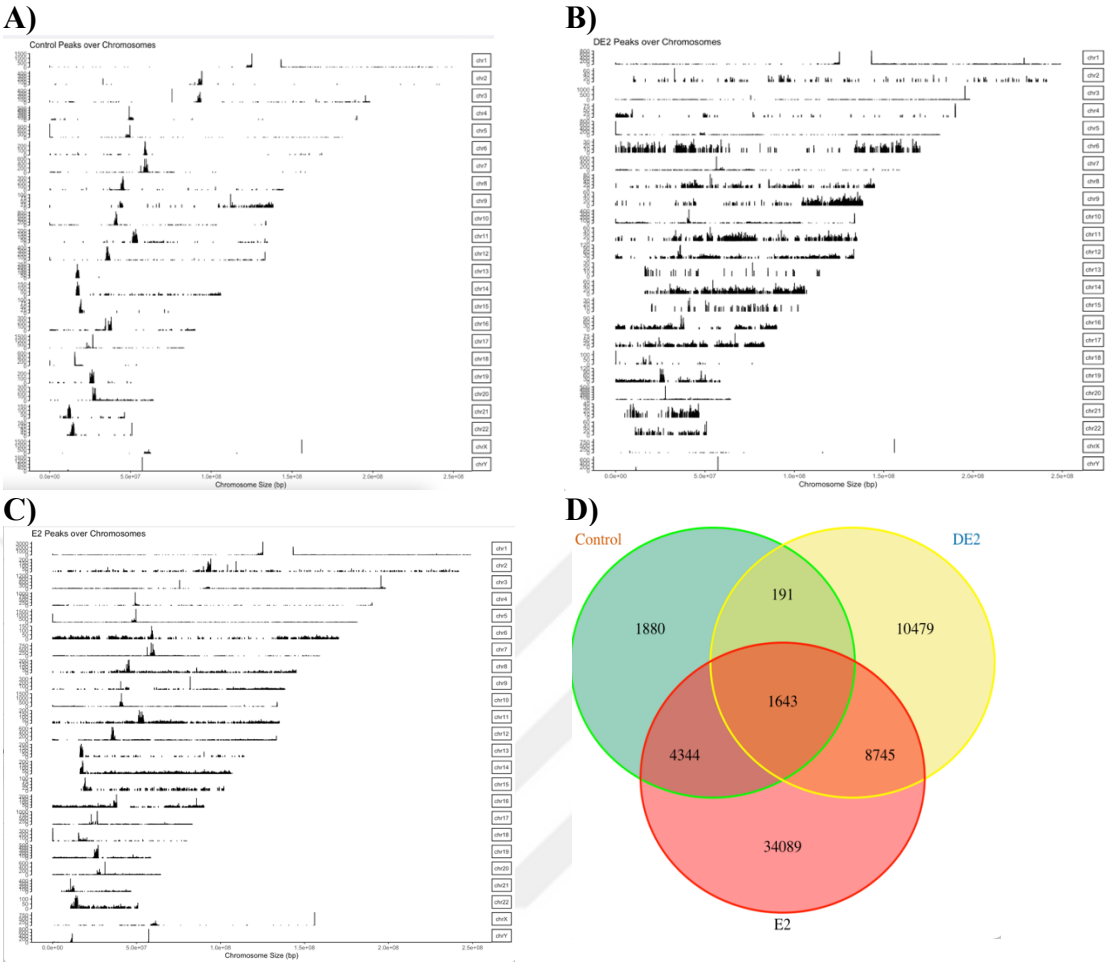


Figure 72. Location and intensity of peaks over the whole genome and Venn diagram A) Control peaks; B) D_E2 peaks and C) E2 peaks. D) Venn diagram representation for common and individual peaks of samples after removal of blacklist regions.

Using a window of +/- 3000bp around the TSS of the genes, the density of read count frequency is visualized to see where binding is relative to the TSS or to each individual sample. These plots estimate the confidence interval by bootstrap (500 iterations) and is shown in the gray shading following each curve. The peak frequency in the TSS and its immediate vicinity was lower in all samples than in other parts (Figure 73A). In addition, the E2 peaks show a nice peak at the TSS with small confidence intervals, while the control peaks showed a slightly wider and less flattened peak around the TSS with the

widest confidence intervals. In D_E2, they were between the other two samples. Another method of visualizing the frequency of read counts by TSS, the heatmap, was also plotted and is shown in Figure 73B. The density and distribution of read counts close to TSS revealed patterns of binding in relation to the regulation of gene expression. Samples treated with E2 showed more prominent peaks in this area than did control and D_E2 samples, suggesting that E2 may have a more effect on the genes through ERE binding near TSS.

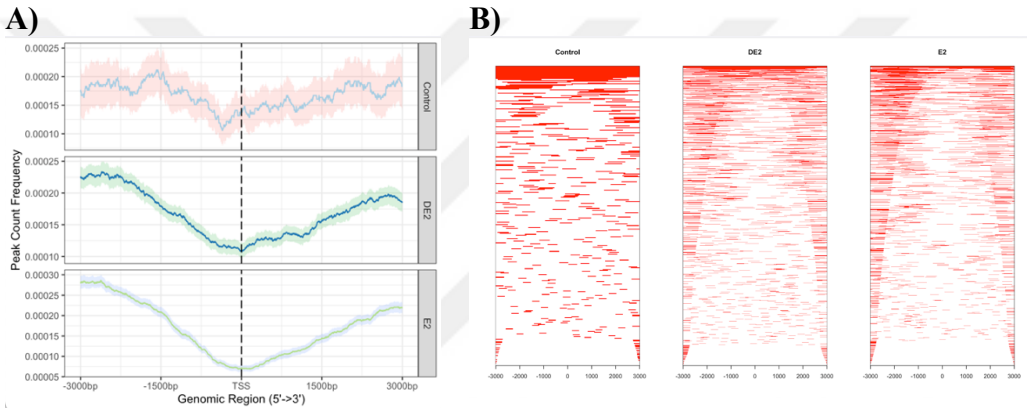


Figure 73. Peak profile plots and heatmap. A) Read count frequency of control, D_E2 and E2 peaks by TSS region, B) heatmap image of the information in A.

The "annotatePeak" function of the ChIPseeker package was used to annotate the position of a particular peak in terms of genomic features. Genomic annotations were found, including whether a peak was in the TSS, Exon, 5' UTR, 3' UTR, Intronic or Intergenic. The feature distribution of these samples, for which the TSS region was defined between -3kb and +3kb, was shown as a bar graph in Figure 74A. Furthermore, the distance from the peak of the nearest gene to the TSS region was calculated and the distribution of binding sites upstream and downstream from the TSS of the nearest genes was visualized with the plotDistToTSS function. Figure 74B compares the distance of the samples to the TSS profiles. Accordingly, the binding rate to regions farther than 100kb was very high

in the control, but this rate was lowest in D_E2. The binding rate to the regions closest to the TSS was also high in D_E2.

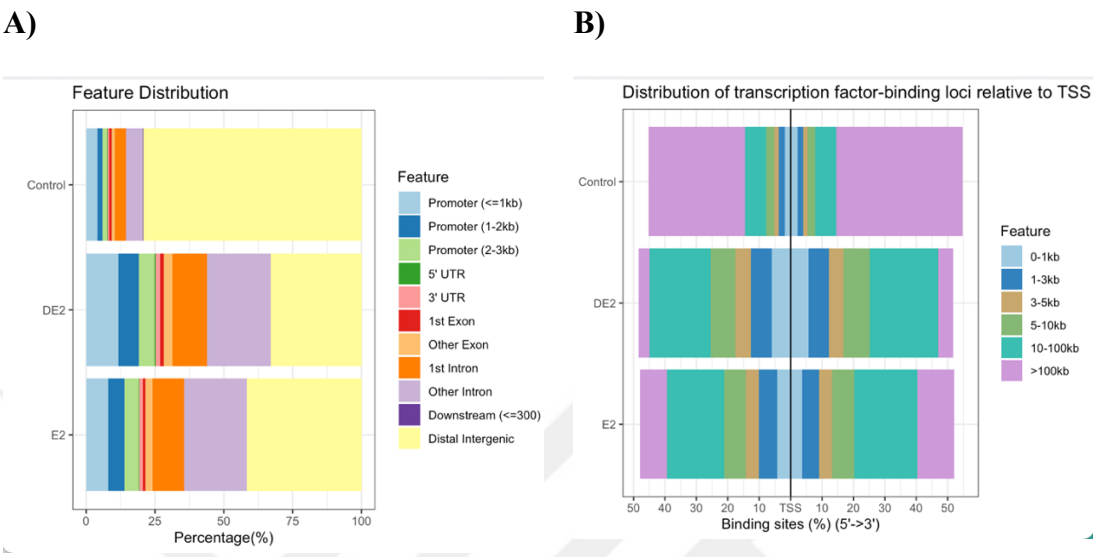


Figure 74. Genomic annotation of the samples. A) Distribution of genomic features in control, DE2 and E2 samples. B) Distribution of binding portions of control, DE2 and E2 peaks according to their distance from the TSS.

On the other hand, the ClusterProfiler package includes a CompareCluster function to compare biological themes between gene clusters. This function was also used to compare peaks from D_E2 and E2, after removing control peaks. The results obtained were analyzed by juxtaposing groups of KEGG pathways (Figure 75). Through KEGG pathway enrichment, a comparative examination of biological themes across gene clusters from D_E2 and E2 samples displayed certain pathways that were enriched differently in each sample. This implies that distinct signaling pathways linked to cancer and other physiological processes are regulated differently by E2 and D_E2 therapies. Enrichment of pathways such as choline metabolism, insulin signaling, AMPK, ErbB, phospholipase D, and sphingolipid signaling in the E2 sample demonstrates the extensive

regulatory activities of estrogen in pathways relevant to cancer metabolism, cell proliferation, and survival. Conversely, the distinct pathways like the cGMP-PKG signaling pathway, insulin secretion, parathyroid hormone synthesis, and aldosterone synthesis and secretion in the D_E2 sample suggests that the addition of Drsp to E2 treatment introduces particular changes in hormone synthesis, secretion, and signaling processes.

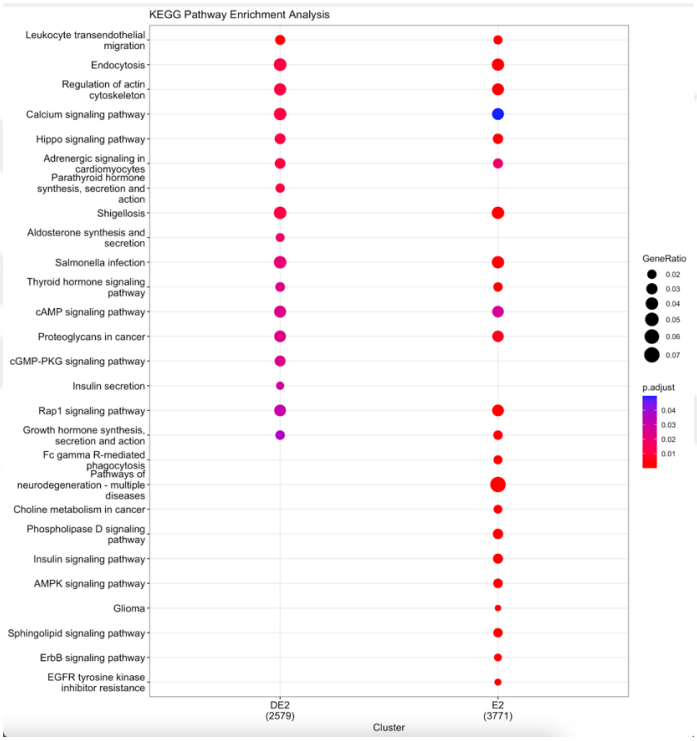
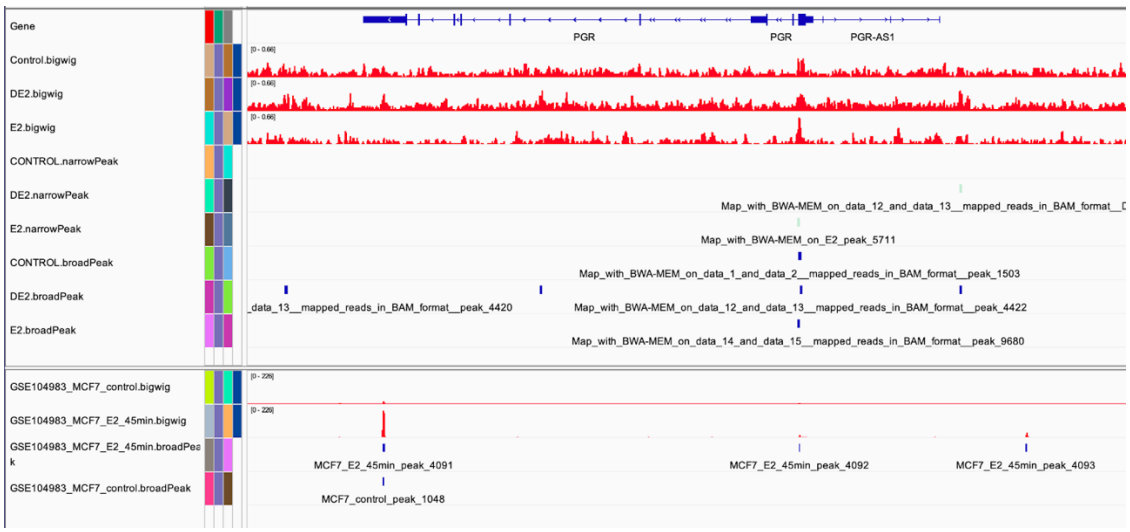


Figure 75. Comparison of KEGG pathways in DE2 and E2 samples.

Narrow peaks were obtained for the analysis of transcription factors, while broad peaks were obtained from histone modifications and chromatin-binding proteins. Therefore, both narrow and broad peaks were generated in Galaxy and loaded into IGV (J. T. Robinson et al., 2011) to look for genes that were already analyzed, such as PGR. Looking as broad peaks made it easier to see the peaks. Figure 76 shows the peak and overlap images of PGR and GREB1 region in IGV. Also, another ChIPseq data (Hinohara et al.,

2018) was obtained from the literature, which was performed in MCF-7 cells using the same antibody as ours, and analyzed together with the results of our experiment. The last 4 tracks in Figure 76 shows this MCF-7 data. There were significant disparities between our results and the MCF-7 data that were analyzed together in the IGV. In particular, coverage in our data was different, as seen by the bigwig tracks (lack of prominent peaks), suggesting that antibody binding in our ChIP assay was present but to a lesser extent and gave high background noises.

A)



B)

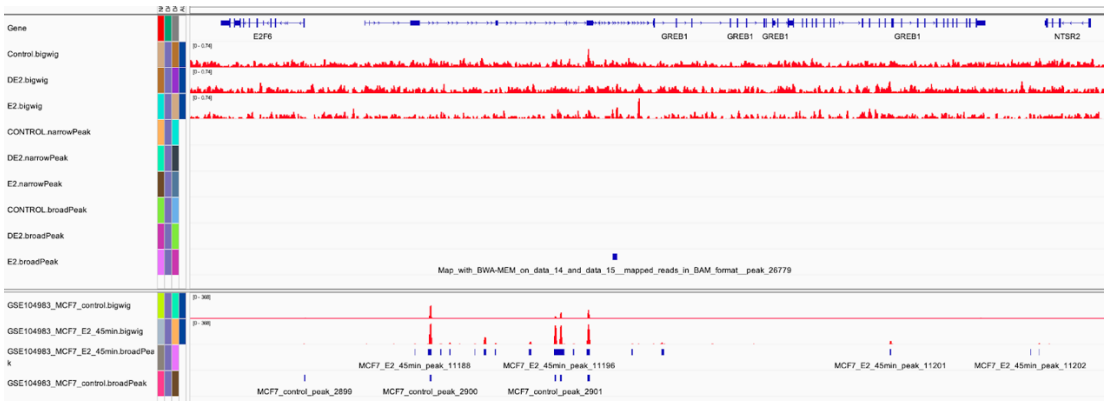


Figure 76. Peak and coverage result of our own ChIP data and GSE104983 data in the PGR (A) and GREB1 (B) region in IGV.

3.7 Ongoing Studies

3.7.1 Investigation of Associations Between Senescence and Liver Fibrosis and Finding Markers

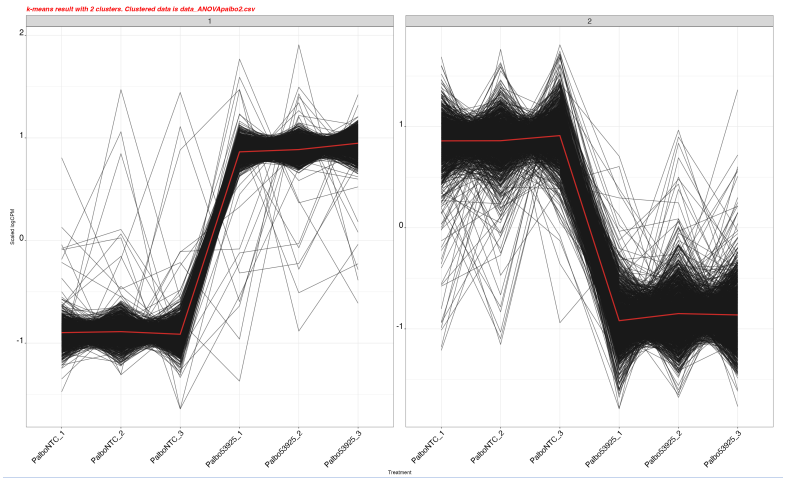
Cellular senescence has been implicated in the aging-related loss of homeostatic and regenerative capability in liver. Under pathological circumstances, senescence can have a negative impact on tissue regeneration, cellular viability, and liver functions (Ferreira-Gonzalez et al., 2018). Alcoholic liver disease, non-alcoholic fatty liver disease, and acute and chronic liver damage have all been related to the existence of senescent cells in the liver (Huda et al., 2019; Moustakas et al., 2021; Wan et al., 2017). Senescent cells are known to have an adverse effect on liver regeneration after hepatectomy, highlighting the significance of senescence on liver regeneration (Duan et al., 2022; Song et al., 2023). Furthermore, a connection has been demonstrated between hepatocyte senescence and liver fibrosis, suggesting that hepatocyte senescence contributes to the development of hepatic fibrosis (Wijayasiri et al., 2022). Accumulated senescent cells and elevated senescence genes were seen at the beginning of hepatic fibrosis regression (Song et al., 2023). However, these were not addressed by focusing on “essential genes”.

In this context, I have reanalyzed part of an existing public study from GSE228941 dataset (Pan et al., 2023) in which senescent liver cells (Huh7) were treated with siRNA against thrombomodulin mRNA (THBD) which has led to senolytic activity to find out which essential genes were affected (downregulated or upregulated) by using CAP-RNAseq (Vural-Ozdeniz et al., 2024). I used 0.2 CPM as the threshold in at least three samples for filtering the raw count data and performed a vst+ANOVA using FDR< 0.001 and obtained 2785 genes with a significant difference of mean expression in at least one group

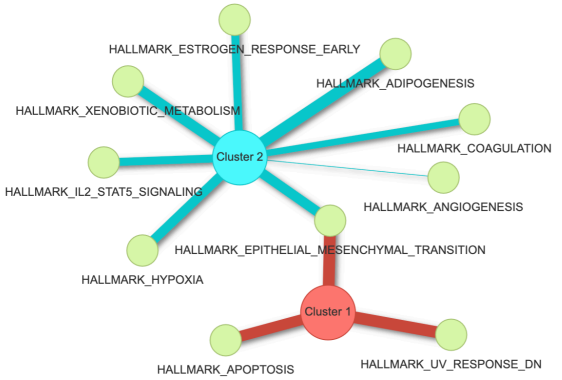
between the senescent cells treated with non-targeting (NT) siRNA and THBD-siRNA (Pan et al., 2023). Cluster analysis of these genes into two clusters, i.e., upregulated and downregulated in response to THBD-siRNA when compared with NT siRNA (Figure 77A), were then annotated for non-essential and essential genes using HallMarks from MSigDB (Figure 77B-C). The results showed that apoptosis was among the highly significant terms for the upregulated genes, but these were not essential (Figure 77B) and cell cycle and growth genes along with cholesterol homeostasis were among the downregulated essential genes in the presence of THBD-siRNA treatment (Figure 77C).

Previous research has shown that senescent macrophages prevent the efflux of cholesterol and may cause disorders associated with aging (Sene et al., 2013). The hepatic signaling molecules oxysterols, which function through the liver X receptor (LXR) that regulates host lipid metabolism and cholesterol homeostasis, are mostly dependent on cholesterol as their precursor (Lorbek et al., 2015). It was also shown that cholesterol synthesis was impaired in senescent cells causing the accumulation of cholesterol (Ziegler et al., 2024). Our findings may suggest that senescent cell clearance in fibrosis by the downregulation of THBD signaling maybe associated with reduced cholesterol homeostasis genes, and this can have interactions with high fat diet. We are further working on the determination of novel candidate genes that can predict the state of clearance and its association with diet by analyzing other public datasets as well as translate this knowledge to the zebrafish model.

A)



B)



C)

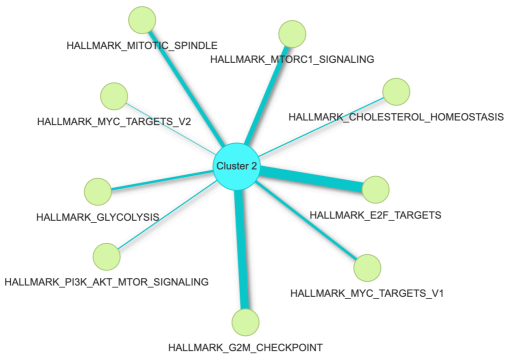


Figure 77. Analysis of GSE228941 data using two clusters. A) Expression patterns of genes in clusters. B) MSigDB analysis using Hallmark for non-essential genes. C) MSigDB analysis using Hallmark for essential genes. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between treatment and control samples.

3.7.2 Understanding the Involvement of Essential Genes in TP53 Mutant Breast Cancer Cells

Previously in our lab Burcin Arıcı has developed wildtype (WT) or TP53 mutant copy overexpressing stable isogenic MCF7 cell lines (Arıcı, 2022; Bilkent University). mRNA RNA-seq of these cell lines have been performed and fastq files were processed via our main pipeline to obtain the raw count data and were also incorporated in TP53-META shiny app developed and explained in Moiz Aftab's Master's Thesis (Aftab, 2024; Bilkent University). I have used these data from EV or WT TP53 overexpressing homozygous as well as R175 and R273 mutant TP53 overexpressing heterozygous MCF7 cells (each with two replicates except EV) using CAP-RNAseq. I used 0.1 CPM as the threshold in at least one sample for filtering the raw count data and performed a vst+ANOVA using FDR< 0.05 and obtained 2,310 genes. Expression patterns of these genes were visualized into 8 clusters (Figure 78A). Clusters were annotated for essential and non-essential genes using HallMarks from MsigDB. The analysis of essential gene enrichment prioritized three co-expression clusters, Clusters 1, 4 and 8. Interestingly focusing on the essential genes only, Cluster 1 marked the upregulated genes particularly in cells with R273H mutation which are related to glycolysis, Myc-targets_2, hypoxia, while Cluster 4 comprising of genes upregulated in WT overexpressing MCF7 cells had mTORC1 and unfolded protein response pathways, distinctly. But most interestingly E2F targets, and G2M checkpoint pathways were downregulated in WT overexpressing cells but some of the genes in these two pathways were upregulated in the R173H mutant expressing isogenic cells (Figure 78B). Moreover, WT overexpressing cells (Cluster 8) had also lower expression of MYC targets and DNA repair genes specifically. These finding suggested that essential genes belonging to cell cycle and DNA repair and oncogenic pathways were downregulated in

a WT overexpression rather than heterozygous mutant P53 overexpressing cells (Figure 78B).

On the other hand, functional annotation of non-essential genes brought up the hallmark interferon gamma response in Cluster 5, where cells expressing R175H had upregulated and R273H had downregulated genes. The non-essential genes upregulated in WT TP53 cells while being downregulated in the mutant cells from clusters 4 and 7 were enriched in estrogen response pathways (Figure 78C). Our findings showed that MCF7 cells expressing mutant TP53 in the wildtype background did not downregulate E2F1 targets and G2M checkpoint genes as the WT TP53 overexpressing genes did and also exhibited lower downregulated early ER signaling genes. On the other hand these cells had specific functional enrichments, e.g., TNF alpha and interferon gamma responses. However, mTORC1 response seemed to be commonly enriched in all isogenic cell lines while in WT cells mostly the essential genes of mTORC1 were upregulated.

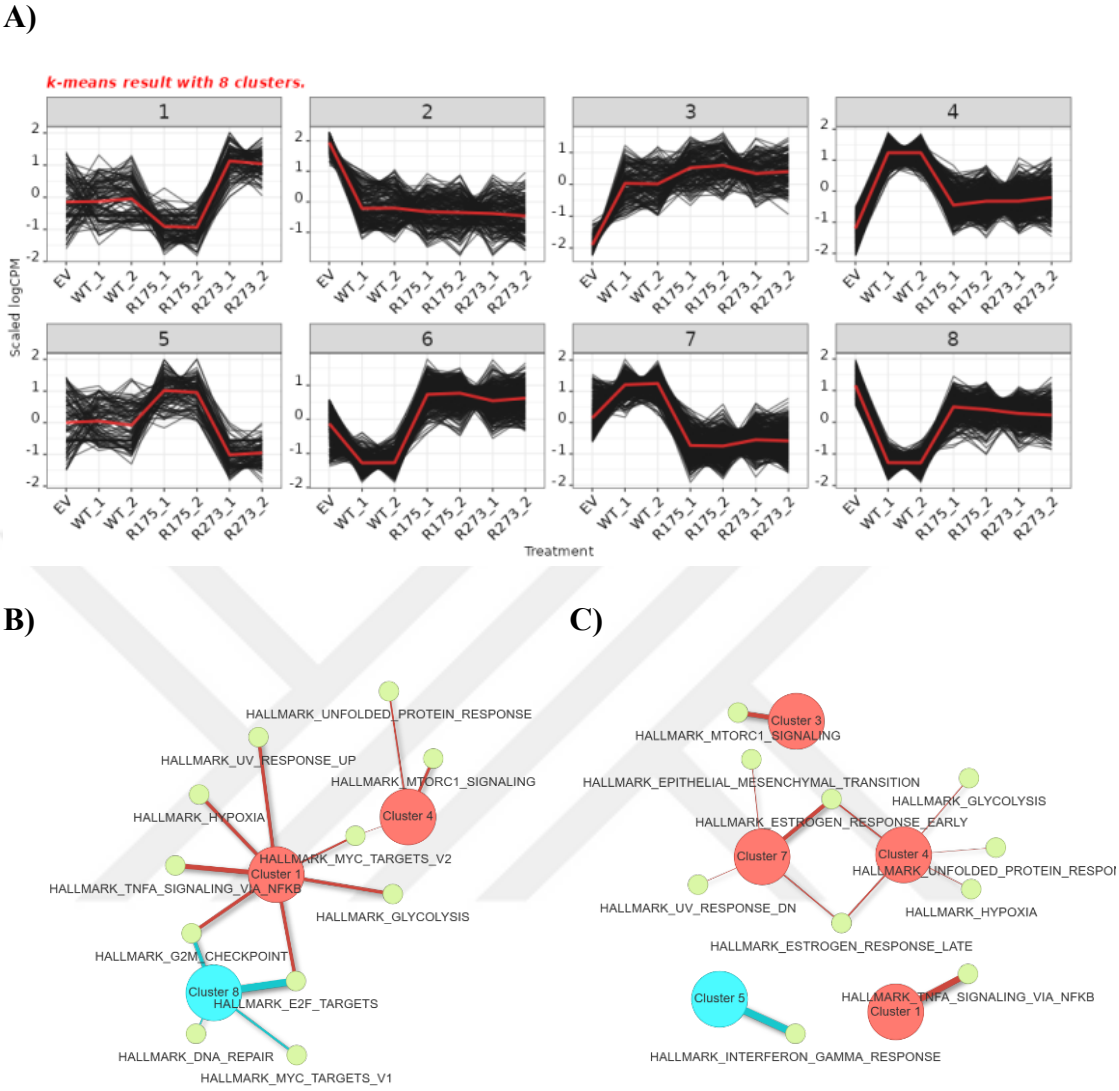


Figure 78. Cluster and functional annotation analysis of empty vector (EV), wildtype (WT), R175H and R273H TP53 mutant expressing MCF7 cells using CAP-RNAseq. A) Expression patterns of genes in 8 clusters. B) MSigDB analysis using Hallmark for essential genes. C) MSigDB analysis using Hallmark for non-essential genes. The colors of the cluster nodes showed whether the average expression of genes was increasing (red) or decreasing (blue) between WT and EV samples.

CHAPTER 4

DISCUSSION, CONCLUSION AND FUTURE PERSPECTIVES

4.1 A Summary of Novelties Brought by CAP-RNAseq

RNA-seq has become a fundamental element in biological research, generating large amounts of data that offer unprecedented insights into patterns of gene expression across different concepts. However, the large volume of such high-throughput data poses a significant challenge, especially for researchers without extensive programming or statistical skills. This emphasizes the critical need for RNA-seq analysis tools that are not only powerful but also accessible and user-friendly so that complicated sequencing data can be evaluated and interpreted without the need for extensive bioinformatics knowledge. In this thesis, after evaluation of existing online existing RNA-seq analysis tools in the literature (please see introduction), a novel all-in-one pipeline has been offered, i.e., CAP-RNAseq. This pipeline starts from the filtering of raw data and ends with prioritization of co-expression clusters/genes using different approaches and primer design of selected genes. The main theme in CAP-RNAseq is the clustering of genes based on their expression patterns across two or more experimental groups followed by identification of 'mirror clusters' and their simultaneous annotation, so that the functions of genes that behave oppositely can be easily visualized using cluster networks differently than other tools in the literature, e.g., iDEP (Ge et al., 2018). Mirror clusters have been defined in the literature (Baldessari et al., 2005; Lewis et al., 2013; Yaslianifard et al., 2023) yet there was no tool that extracts and annotates available data before CAP-RNAseq.

The use of different datasets such as essentiality of genes from DepMap and protein expression level data in tumors and normal samples from HPA is another distinguishing feature of CAP-RNAseq. Accordingly, the clusters and genes can now be prioritized based on gene essentiality and congruency of mRNA-protein expression levels, which is not yet found in any other RNA-seq tools. The congruence between mRNA and protein levels is of great importance in understanding gene expression and its regulation. It allows researchers to assess how transcriptional changes correlate with changes in protein abundance, providing insights into the regulation and function of genes (Vogel & Marcotte, 2012). Furthermore, the congruence between mRNA and protein levels is crucial for accurately identifying essential genes. The expression of essential genes at both the mRNA and protein levels ensures that the necessary proteins are being produced in sufficient quantities to carry out critical cellular functions (Li et al., 2014). Understanding this congruence can help in identifying and validating essential genes, as discrepancies between mRNA and protein levels may indicate regulatory mechanisms or post-transcriptional processing that affect protein abundance (Wang, 2008).

Furthermore, CAP-RNAseq allows the comparison of essential genes across different datasets and treatments, revealing their interesting relationships. Essential genes are required for the survival of organisms. Cells from various lineages often rely on different essential genes in comparison to cells within the same lineage (Shimada et al., 2021). They are described as lineage-dependent essential. The difference between common essential and lineage-dependent essential genes is that common essential genes are required for the basic cellular functions and are essential for the survival of all organisms, regardless of their lineage (Zhan & Boutros, 2016). On the other hand, lineage dependent essential genes may only be essential for specific lineages or species and may not be

present or necessary in other organisms. Identification of candidate drug targets can be achieved by focusing on selectively essential genes in each lineage (Behan et al., 2019). Most genes that exhibited high selectivity scores were also recognized as targets for one or more lineages in their analysis. Conversely, targets suggested for multiple lineages tended to display moderate efficacy scores, rather than high selectivity scores necessarily (Shimada et al., 2021). This indicates that there may be a trade-off between selectivity and efficacy when it comes to identifying lineage-dependent essential genes as potential targets. CAP-RNAseq includes this lineage-dependent essentiality for the selected gene for detailed analysis of the genes.

The enrichment analyses of these genes can also provide valuable insights into the biological pathways and processes, revealing the roles of essential genes in health and disease. For gene-wise prioritization, CAP-RNAseq facilitates the identification of new genes/biomarkers in different contexts based on different approaches such as the degree of gene essentiality, selectivity, expression level, change in expression levels (LogFC) as well as the congruency of mRNA and protein expression amounts and assessment of protein levels in normal and cancerous tissues. There are many tools that use DepMap or HPA yet CAP-RNAseq is the first tool that makes use of these in the context of cluster and gene annotation and prioritization.

CAP-RNAseq also provides a range of visualization options to meet different types of analytical requirements, which improves the interpretation of RNA sequencing data. It displays clustering results as line or box plots, providing a clear representation of gene expression patterns in different clusters. The enrichment results can be also visualized as a network so that the interactions between the clusters and biological terms can be

interpreted easily. In addition, CAP-RNAseq offers scatter plots where the genes can be analyzed based on their essentiality in the different context such as variation in the form of CV, average CPM levels, range of CPM levels, efficacy/selectivity values, etc. Genes are colored based on their essentiality (as essential or non-essential) while size of the plotted dots/genes further represents a third dimension for a gene's properties (e.g., expression, range of logFC) where the axes show two other axes for the genes, such as CV, CPM, etc. This type of plotting is particularly valuable for comparing different parameters within the clusters such as essentiality vs expression, or CV vs expression. Moreover, it offers advantages over MA plots which are commonly used to compare differences in gene expression (Robinson et al., 2010). Although MA plots are helpful in identifying differentially expressed genes across expression levels, they are used for two group analyses and do not annotate for more than two axes. On the other hand, CAP-RNAseq's MA like plots are applicable to data with more than two groups and provides additional and user-dependent visualization options that enable a more sophisticated analysis of gene clusters that contain more than two groups and two variable plot options on the graph.

In conclusion, CAP-RNAseq emerges as a valuable tool to annotate and prioritize the co-expression clusters and genes, as it makes easier for the user to comprehend patterns and connections within and between different datasets on a deeper level.

4.2 Novel Associations Discovered by CAP-RNAseq for NTRK2 Signaling in Neural Stem Cells and Fibroblasts

NTRK2 has been linked with gliomas (Gatalica et al., 2019) and has a role in brain tumor-initiating cell survival (Lawn et al., 2015), and NTRK2 was found to be crucial for the formation of gliomas in a mouse glioma model using a library that targets kinases (Ni et al., 2017). Neurotrophins bind to Trk (tropomyosin receptor kinase) receptors with known roles in nerve cell survival and NTRK2 gene coding for a protein known as TrkB, is one them that is important in differentiation of neurons (Chao et al., 2006). The full length and isoforms of TrkB, TrkB-T1 and TrkB-T-Shc have been described (Klein et al., 1990; Stoilov et al., 2002) where TrkB.T1 lacks a kinase domain (Brodeur et al., 2009) along with a conserved C-terminal 11 aa addition (Fenner, 2012; Luberg et al., 2010). TrkB.T1 is also often detected in gliomas (Pattwell et al., 2020) and associated with different tumorigenic functions.

CAP-RNAseq was applied to two public datasets that involved NTRK2 overexpression and silencing, and hence was used as a comparative transcriptomics tool in this thesis to discover gene sets that are modulated inversely in two independent studies. Moreover, it allowed to discover mirror clusters with associated hallmark terms that can give insight into pathways behaving in different directions across more than two experimental groups. These analyses revealed the novel relationships in overexpression/depletion models of NTRK2. While it was shown that TrkB.T1 isoform promoted the PDGF-driven gliomagenesis by increasing the PDGF-induced Akt and STAT3 signaling pathways in the original study (Pattwell et al., 2020), CAP-RNAseq demonstrated additional mechanisms and relationships. Accordingly, the genes upregulated by the overexpression

of NTRK2 were responsible in pathways of cell division, underscoring the role of TrkB in promoting cellular proliferation, which was shown in the literature before (Yuan et al., 2018). By analyzing the mirror cluster acting in the opposite direction, however, the downregulated genes by the overexpression of NTRK2 indicated that TrkB overexpression may inversely affect pathways related to differentiation and cell migration. Cells undergoing division due to overexpression of NTRK2 might downregulate genes critical for cell migration, immune responses, and p53 pathway. This could only be possible to visualize by using "mirror clusters" and their simultaneous pathway association.

On the other hand, CAP-RNAseq analysis also indicated there were differential effects of overexpression of the isoform TrkB.T1 and full length TrkB.FL on the transcriptome such that genes related to EMT and downregulated by UV response were downregulated by TrkB.T1 while this downregulation was more by overexpression of TrkB.FL, which may have a greater influence on cellular systems linked to UV response and EMT. On the contrary, genes upregulated more by TrkB.FL were related to energy production, growth signaling, lipid metabolism, and fat cell development. While CAP-RNAseq showed that overexpression of both NTRK2 isoforms caused increase in the expression of genes related to cell cycle regulation and division, TrkB.T1 also affected another subset of genes in other cluster related to cell cycle regulation and division by increasing their expression, which possibly promoted more the proliferation. TrkB.T1-dependent increase in the proliferation was found to be mediated by PDGF signaling in the original study (Pattwell et al., 2020) and was also observed through neural progenitors in another study (Tervonen et al., 2006).

CAP-RNAseq also has allowed analysis of essential genes and how they were affected by overexpression of the two forms of NTRK2. Similar to the existing studies (Freischel et al., 2022), essential genes were highly expressed in the demo dataset 1 where NTRK2 was overexpressed in neural stem cells. The novelty CAP-RNAseq brings is that each RNA-seq dataset may have a different set of essential genes modulated and CAP-RNAseq makes possible testing whether this is always true. Moreover, while log2 range CPM values were significantly higher, CV of CPM was lower for essential genes indicating when standard deviation normalized to the mean expression of a gene the relative deviation was smaller although the absolute number of transcriptomic changes (i.e., reads) was more for the essential genes. It was also observed that most of the essential genes were found in the clusters where TrkB.FL increased the cluster members' gene expressions.

CAP-RNAseq highlighted the importance of CCND2, one of the essential genes, in its interaction with NTRK2 expression across the two different datasets I analyzed. In GBM, CCND2 is known to be essential for the progress of the cell cycle, tumorigenicity, and radio-resistance, and it was shown that CCND2 expression was markedly elevated in GBM and rarely occurs in normal brain tissue or low-grade gliomas (Koyama-Nasu et al., 2013; X. Zhang et al., 2005). In demo dataset 1, the overexpression of NTRK2 reduced the expression of CCND2, with the greatest effect seen by that of TrkB.T1, the predominant isoform in gliomas. This showed the negative correlation between NTRK2 and CCND2, and further validated in TCGA LGG dataset from cbioportal (Cerami et al., 2012; De Bruijn et al., 2023; Gao et al., 2013). In demo 2 dataset analyzed in CAP-RNAseq, CCND2 was shown to be increased in senescent cells and NTRK2 silenced senescent cells. In a study using cells undergoing various cellular growth arrest states

induced by contact inhibition, serum starvation (quiescent), or cellular senescence, CCND2 mRNA and protein expressions were significantly increased and the overexpression of CCND2 prevented the DNA synthesis and progression of cell cycle, inhibiting the entry into S phase (Meyyappan et al., 1998). In the same study, CCND2 was also found in the nucleus in quiescent cells during G0 phase, indicating a potential nuclear role, and interacts with CDK2 and CDK4 but not CDK6 when cells are growth arrested. Therefore, CCND2 can act as a suppressor of growth by promoting cell cycle exit and assisting in the maintenance of cells in a non-proliferative condition.

This novel association discovered by comparative transcriptomics analysis performed by CAP-RNAseq could point to a dynamic regulatory mechanism in which NTRK2 functions as a suppressor of CCND2 expression and suggests that the targeting the NTRK2-CCND2 axis as a possible therapeutic strategy in the glioma therapies can be used in further research to explore the biological significance and therapeutic potential.

4.3 Senescence and Essential Gene Relationship

Senescence differs from quiescence, which is defined as the arrest of cell cycle, such that it irreversibly halts cell division. Previous studies have identified genes essential for senescence by CRISPR screens (X. Liu et al., 2019) as well as shown that conserved essential genes to be associated with increased genetic stability and regulation of life span (Curran & Ruvkun, 2007; Jazwinski, 2000; Oz et al., 2022). In addition essential genes in protein degradation and PI3K-Akt pathways were associated with signatures of long-lived animal and being ancient (Tyshkovskiy et al., 2023). However, the association between essential genes and senescence is not yet fully understood yet hence CAP-

RNAseq's application on fibroblast senescence data has been one of the most important outcomes of this thesis.

In association with this, I found that most of the essential genes were predominantly located within the clusters that contained genes significantly downregulated by the senescence (demo 2 dataset). Since essential genes are needed for cell survival their downregulation makes sense (Behan et al., 2019; T. Wang et al., 2015). These genes were responsible for DNA repair, G2M checkpoint, and were members of MYC and E2F targets, and MTORC1 signaling, necessary for cell survival and division (Clay & Fox, 2021; García-Gutiérrez et al., 2019; Hung et al., 2012). During senescence, their downregulation implies a diminished ability to repair DNA which might result in the formation of genetic mutations and the stability of the senescent state, arrest of cell cycle progression and cessation of cell proliferation which are the hallmarks of senescence (Roger et al., 2021). However, similar pathways are also downregulated in quiescence hence it is important to test whether the magnitude of downregulation in the proliferation related essential genes are lower in senescence than quiescence. In order to test such a hypothesis existing RNA-seq datasets in the literature can be tested in CAP-RNAseq so that these two conditions could be compared with each other. One such dataset is GSE74620 in which both serum starvation and different senescence models have been profiled in dedifferentiated liposarcoma cell line, LS8817 (Kovatcheva et al., 2017). To demonstrate I have used this additional dataset on CAP-RNAseq; generated 12 clusters at $p<0.0005$; and identified that Cluster 2 and Cluster 4 genes were downregulated in CDK4 inhibitor treated cells in comparison with serum starved cells (senescence vs. quiescence). As expected, the proliferation related essential genes were relatively more downregulated in senescence (Appendix F). In the original article, the authors used Venn

diagrams and identified 166 genes that were specific to senescent cells only and focused on those yet did not perform similar analyses to those performed here in CAP-RNAseq. This shows that CAP-RNAseq is a highly useful tool to test hypotheses generated by one dataset against additional datasets and identify novel findings that were not reported in the original articles.

Similarly, some of the essential genes downregulated by senescence the demo 2 dataset interestingly belonged to “cholesterol homeostasis”, which included genes important for cholesterol synthesis, i.e., FDFT1, SCD, SQLE, TMEM97, ALDOC, FADS2, ACTG1, PCYT2, SREBF2. However, serum starvation or quiescence has been shown to increase cholesterol synthesis (Angius et al., 2015; Chapman et al., 2020); and yet, recent literature has shown that the disruptions in lipid metabolism is one of the key features of senescence (B. Liu et al., 2023). Again, using the GSE74620 dataset, genes upregulated by senescent cells in comparison to serum starvation were found in clusters 11 containing upregulated cholesterol homeostasis genes in serum starvation (i.e., FDFT1, SCD, SQLE, HMGCS1, HMGCR, LDLR, FADS2, FDPS, MVD, FASN and SREBF2), many of which were the same essential genes that were downregulated in the senescent cells of demo 2 dataset. It is then intriguing to suggest that cholesterol synthesis as one of the main modulators of the senescence based on the recent literature complemented with reanalyzes of public datasets by CAP-RNAseq and this association now warrants further study in the laboratory.

Moreover, some essential genes upregulated in replicative senescence in fibroblasts belonged to the genes encoding the components necessary for the function and maintenance of peroxisomes, which might indicate the importance of maintaining the

lipid metabolism and the detoxification of reactive oxygen species in senescence. Peroxisomes are required for beta-oxidation and proper cholesterol synthesis because peroxisome deficiency reduces it (Charles et al., 2020), hence peroxisomal feedback might be involved in senescence as well. This association between cholesterol homeostasis genes having high expression in growth arrest but not in senescent cells could be important for liver senescence and fibrosis and relevant diseases, such as hepatosteatosis. More datasets dealing with senescence will be analyzed in the future in relevance to cholesterol biosynthesis.

Interestingly, non-essential genes in this dataset (demo 2) accumulated in the clusters containing other upregulated genes by senescence, which were enriched in KRAS signaling that plays role in cell growth, division, and survival although not being essential (Zhou et al., 2024). The enrichment of genes upregulated in this pathway during senescence has suggested potential changes in cell signaling that could promote survival and adaptation in a non-proliferative state, even though KRAS is frequently linked to oncogenic transformation in other contexts (Jančík et al., 2010). Moreover, KRAS has been involved in oncogene induced senescence (Muraki et al., 2022), hence its activity might be important in fibroblast senescence which was induced by etoposide in demo 2 data (Anerillas et al., 2022). On the other hand, downregulated non-essential genes were found to be linked to G2M checkpoint, MTORC1 signaling, interferon responses and in response to UV radiation. Downregulation of these genes might indicate a reduction in cell proliferation, aligning with the senescent state where cells exit the cell cycle. Moreover, the downregulation of these genes results in decreased activity in the MTORC1 pathway, which may indicate a change in senescent cells toward conservation and stress tolerance as well as an adaptation to reduced energy needs. However, these

pathways also contain essential genes downregulated by senescence and the interactions between these two sets of downregulated genes can be studied further. Moreover, the reduction in immune responses might indicate a weakened ability to fight viral infections in senescent cells. Indeed, in literature, increased dysfunction of immune responses with aging were found (Lee et al., 2022). These genes have also been found to be downregulated by UV radiation, which may indicate a reduced capacity to repair DNA and protect it from UV radiation, potentially contributing to the accumulation of DNA damage that is characteristic of senescent cells.

As a result, CAP-RNAseq's mirror cluster analysis has shown that essential genes, which were downregulated by senescence, had high max. fold changes and CV and as well as log range CPM values; this set of essential genes behaved relatively differently than the general essential gene properties, e.g., high levels of expression and low CV. Moreover, without comparing the senescent cell signature with a quiescent cell signature, it would be impossible to determine whether these genes were downregulated also in quiescence and if so, how differently from the senescent state. CAP-RNAseq has provided that opportunity to the users.

4.4 The Effect of NTRK2 Inhibition

It has been known that apoptosis can be suppressed in senescent cells and the inhibition of NTRK2 selectively increases apoptotic death of senescent cells and reduces the viability of senescent cells in the study whose data were used as demo 2 in CAP-RNAseq (Anerillas et al., 2022a). The clusters, where the apoptosis appeared as an enriched term, contained the genes that were upregulated in senescent cells, but this increase in

expression was partially reverted when NTRK2 expression was inhibited. While overexpression of some of these genes, e.g. TIMP3 (Qi & Anand-Apte, 2015), IGFBP6 (Bei et al., 2017) and RELA (Khandelwal et al., 2011), induces apoptosis, knockdown of others such as GPX1 (Lei et al., 2023), CD44 (P.-T. Wu et al., 2019), DNAJA1 (Roth et al., 2021) and ATF3 (Ma et al., 2015) induces apoptosis. Having apoptosis as a category modulated by senescence and NTRK2 inhibition in CAP-RNAseq allows prioritization of these genes in cancer studies using HPA as well as mRNA-protein correlation tables, in the future.

Moreover, the genes that were upregulated by the inhibition of NTRK2 (siNTRK2) when compared to the senescence control (siCtrl_S) displayed significant enrichment with the mitotic spindle suggesting that inhibition of NTRK2 can affect the efficiency of chromosome segregation and cell division. As is well known, cells stop dividing during a process of cellular senescence (Campisi, 2013); and the inhibition of NTRK2 in senescent cells has been shown to lessen senescence-induced downregulation of genes involved in cell division in our analysis. This indicates that the higher level of expression of NTRK2 in senescent cells might help sustain the decrease in cell division, while the inhibition of NTRK2 feedbacks and lessens the downregulation of cell division genes. In their original article the authors also showed that the inhibition of NTRK2 caused the reduced accumulation and viability of senescent fibroblast cells (Anerillas et al., 2022b). Future studies should test whether senolytics in general behave similarly as the inhibition of NTRK2.

Furthermore, NTRK2 may function as a regulatory factor in preserving the epithelial phenotype or preventing the mesenchymal transition under normal circumstances,

according to the finding that EMT-related genes were increased in response to NTRK2 inhibition by CAP-RNAseq, which is a unique finding revealed by the reanalysis of the data in CAP-RNAseq. The cells may be more likely to undergo EMT as a result of NTRK2 inhibition, which may enable alterations in cell behavior linked to enhanced motility and invasiveness. In the literature, overexpression of TrkB caused the changes in the expression of molecular mediators of EMT such as upregulating Twist and downregulating E-cadherin (Kupferman et al., 2010). This interaction between EMT and NTRK2 needs further investigation.

Moreover, UV radiation-responsive genes were differentially expressed in the clusters upon CAP-RNAseq analysis of demo 2. Since genes in these cluster where they were downregulated by the inhibition of NTRK2 in senescent cells were enriched with genes that are upregulated in response to UV radiation, they may be involved in cellular defense mechanisms against UV-induced damage. Conversely, another cluster which contained genes upregulated by the inhibition of NTRK2 in senescent cells, was found to be enriched in genes that were downregulated in response to UV radiation. This suggests that NTRK2 inhibition affects a subset of genes that may decrease the cell's sensitivity to UV-induced stress by interfering with pathways that may be involved in the detection, signaling, or repair processes of UV damage. While the relationship between UV damage and senescence is well established in the literature (Debacq-Chainiaux et al., 2012; Fitsiou et al., 2021; Yao et al., 2022), it was not mentioned in the original article and there is no research on the effect of NTRK2 in this context, indicating the novel relationship between them that needs to be investigated.

4.5 Genes Affecting the Recurrence of Cancer and the Identification of Biomarkers

Different gene expression patterns linked to various responses to neoadjuvant chemotherapy (NAC) were found by re-analyzing the GSE201085 dataset as the demo 3 in CAP-RNAseq to find new biomarkers for breast cancer detection and treatment. In their 2022 article, Axelrod et al. have performed RNAseq on peripheral blood mononuclear cells (PBMC) of 53 patients with different disease outcomes in breast cancer and found that Hallmarks, IFN gamma and alpha response as well as complement pathways were over-represented in the complete remission group (pCR) in comparison to those with recurrence (RD) (Axelrod et al., 2022). Moreover, they found that this Hallmark signature was inversely correlated with published cytotoxicity score. They also used CIBERSORTx to be able to deconvolute their RNA-seq data into specific cell populations in the PBMC and associate the disease outcome with specific blood cell populations. This dataset has provided an important resource for identifying novel biomarkers of pCR by reanalysis in CAP-RNAseq and studying the markers with respect to their breast cancer tissue and normal breast cells protein expression as well as further testing whether the biomarkers had high mRNA-protein correlation so that measuring by qPCR rather than at the protein level could be a possibility.

Reanalysis of GSE201085 in CAP-RNAseq has revealed that genes upregulated in pCR samples obtained from patients who did not have the disease after NAC when compared to RDR samples with recurrence after NAC were related to immune system, energy metabolism, apoptosis, inflammation, survival, and growth of cell based on the MSigDB analysis. Apart from immune response, oxidative phosphorylation, IL6/JAK/STAT3 signaling, and MTORC1 signaling terms were identified as enriched.

Moreover, the downregulated essential genes in pCR samples were found to be related to DNA replication and cell cycle regulation, while upregulated essential genes were enriched in pathways related to various activities such as cell growth and survival, immune response, energy metabolism, and hypoxia. Non-essential genes, on the other hand, were mostly enriched in the pathways related to immune signaling. CAP-RNAseq's ability to annotate essential and non-essential genes show that there is a reduction in tumor cell's ability to proliferate while changing its homeostatic and immune system. But most importantly, CAP-RNAseq has facilitated the prioritization of genes based on their essentiality, protein levels and the congruence between mRNA and protein levels, leading to identification of IL1R2 and ENPP5 genes as possible biomarkers.

IL1R2, found as highly expressed in breast cancer while no expression was found in normal breast tissues based on HPA, had higher expression in RDR samples. In breast cancer, IL1R2 overexpression was found to promote cell invasion and proliferation and linked to poor prognosis (L. Zhang et al., 2020). By blocking IL-1-mediated inflammatory signaling pathways, IL1R2 plays a crucial role in regulating the immune system and inflammatory responses (Peters et al., 2013). Therefore, IL1R2 can be a potential biomarker for the recurrence since it was highly expressed in RDR samples.

ENPP5, was also highly expressed in RDR samples compared to RDnR and pCR samples with increased expression in samples who didn't receive the NAC; and exhibited high mRNA and protein expression level correlation and high expression in breast cancer in HPA. In the literature, there are few studies on ENPP5 in cancer. In breast cancer cells, ENPP5, a target gene of miR-126, was found highly upregulated (Lee et al., 2023). In

TNBC, ENPP5 was also upregulated (W. Wu et al., 2021). These results suggest that ENPP5 is an ideal candidate for further study as a breast cancer biomarker.

4.6 Intricate Crosstalk between Hormones in T47D Cells

4.6.1 Pathway Enrichment of Essential and Non-essential Genes

Drsp is a PR agonist while being an analog of spironolactone, i.e., an MR antagonist, and its combination with E2 has been used in hormone replacement therapy and as a contraceptive for a long period of time (Archer, 2007; Palacios et al., 2006). In this thesis, I have tested whether 20nM of Drsp and 10nM of E2, the common proportion used in clinic, interacted with each other in terms of pathways/genes they modulated in the ER+ and PR+ breast cancer cell line, T47D. Since the effects of Drsp-E2 combination had not been studied with respect to the transcriptome of breast cancer cells in the presence or absence of aldosterone (an important hormone with equivalency to salt in diet), this thesis is highly novel. Moreover, the analysis of our RNA-seq data in this context by CAP-RNAseq has opened novel avenues and hypotheses for the association between hormone use and cancer before and after menopause with respect to modulation of essential genes.

Interestingly, my analyses showed that the expression cluster that contained most of the essential genes was the cluster with the highest increase in gene expression in response to E2. In this cluster of genes, the presence of E2, especially in combination with Drsp (D_E2), increased the expression of genes further, while other treatments did not have as much effect. Moreover, the other clusters with high numbers of essential genes were clusters also containing genes that showed an increased expression pattern in different E2-containing samples. On the other hand, only one cluster contained a single essential

gene, and the genes in this cluster exhibited increased expression in all samples except E2 alone. This cluster also included genes known to be influenced by Drsp and Aldo, such as SGK1 and ZBTB16 (Náray-Fejes-Tóth et al., 2008; Valinsky et al., 2018). In support of the positive association of increased essential gene expression with E2 exposure, the lowest number of essential genes were also found in the clusters which showed decreased expression pattern in response to E2 or other groups containing E2. This overall suggested that E2 effect was in general proliferative as has been already been well established (Fuentes & Silveyra, 2019); yet, this thesis also showed that many of these were the essential genes such that clusters displaying increased expression pattern in samples containing E2 were found to be enriched in the pathways related to DNA replication, cell division and growth and cell-cycle regulation.

Specifically, essential genes in the cluster which contained the highest number of essential genes had roles in cell growth. Pathways related to early and late estrogen responses (Hallmarks, MSigDB) were found in only one cluster that also displayed increased expression pattern with only the E2-containing samples. These both establish that E2 modulates essential gene expression levels. In the literature, while several NRs were identified as essential genes for the cancer cell growth, and several genes related to DNA replication and cell cycle progression were revealed as essential genes in breast cancer (Frigo et al., 2021; Sever & Glass, 2013), there is not any study testing the essential genes in response to estrogen treatment.

Conversely, essential genes that exhibited decreased expression patterns were characterized by distinct pathways. Genes in one of the clusters that were decreased in all samples except Aldo alone were involved in the processes crucial for lipid metabolism,

and responses to stress. Estrogens are known to regulate lipid metabolism and hence estrogen deficiency after menopause leads to the dysregulation of the lipid metabolism (Palmisano et al., 2018; Szafran & Smielak-Korombel, 1998). Progestins are also known to increase lipid storage when given alone or in combination with estrogen (Ward et al., 2022). Therefore, the genes downregulated by these hormones need to be further investigated in the context of lipid metabolism.

Genes in other clusters which showed decreases in expression more by D_E2 and E2 alone treatments were involved in UV response which might indicate the diminished capacity to deal with DNA damage caused by UV in response to hormone treatment. Indeed in the literature this has not been studied in the breast cancer but it was shown that E2 treatments affect the response of skin to UV damage (Şorop et al., 2020). On the other hand, non-essential genes in the clusters showing increases in expression in the E2-containing samples were found to be enriched in estrogen response pathways. This can be considered as a validation of our RNA-seq data, as it is expected. In the cluster containing only one essential gene, non-essential genes were found to be enriched in pathways of androgen response and xenobiotic metabolism. This suggested that at least some of the ESR1 and AR related genes may not be directly acting on cell survival/death. CAP-RNAseq can distinguish these genes from the essential ones and help identify which genes are the survival factors by testing the different Hallmark selections, e.g., transcription factors or CGP, other gene sets obtained from different expression studies.

Non-essential genes linked to adipogenesis were specifically downregulated in one cluster which displayed significantly reduced expression by all E2-containing samples but only slightly reduced expression by Drsp. This has never been shown in the literature

before implying that D_E2 therapy may have a side effect on the cell's ability to undergo adipogenic differentiation, which is the process of cells differentiating and accumulating lipid. However, there is evidence in the literature suggesting that E2 has both pro- and anti-adipogenic effects on adipogenesis. In one study, E2 has been associated with the decreased adipogenesis via mTOR signaling activation (Bjune et al., 2022), while in another study, it was shown that E2 increased the lipid synthesis and adipogenesis in human mesenchymal stem cells (Cohen et al., 2021). Furthermore, through the antagonistic action on the MR, Drsp has been shown to have a potent anti-adipogenic effect that is linked to a change in the transcriptional regulation of adipogenesis (Caprio et al., 2011). Therefore, to clarify the specific mechanism by which E2 regulates the adipogenesis in the breast cancer alone or in combination with Drsp, more studies are required.

In another cluster showing significant downregulation only by D_E2 samples, non-essential genes involved in WNT/β-catenin signaling, important in development, the immune modulatory IL2/STAT5 signaling pathway, and heme metabolism, which affects the oxidative stress responses and energy production. Studies have shown that estrogen can reduce the TGF-β1-induced-EMT adhesion by triggering the pathway of WNT/β-catenin signaling (J. Cao et al., 2020), and can target the early WNT/β-catenin signaling in an ER-independent manner for the regulation of uterine growth (Hou et al., 2004). Moreover, in endometrial cancer, it was shown that progesterone can block the pathyway of WNT/β-catenin signaling (Fedotcheva et al., 2021) and it counteracts the stimulation of WNT/β-catenin signaling induced by estrogen, leading to suppression of estrogen driven proliferation and differentiation (Feng et al., 2019). It has been also shown that progesterone can reverse the EMT and promote differentiation in the breast cancer cells

by blocking the positive effects of estrogen on the IL2/STAT5 signaling pathway which is interconnected with WNT/β-catenin signaling pathway (Sultan et al., 2008). In my analysis, combination of Drsp and E2 decreased the expression of genes related to these terms, and this may indicate a novel counteracting effect of Drsp acting as a progestin to E2 in these contexts.

4.6.2 Genes Affected by Drsp Alone or in Combination with Another Treatment

There were only two clusters containing genes specifically affected by Drsp. These were the mirror clusters, and Aldo alone or E2 did not have much of an effect on these genes in T47D cells. Genes upregulated by Drsp were found to be involved in fatty acid metabolism, xenobiotic metabolism, early estrogen response and androgen response. In the literature, progesterone causes increased lipid storage but Drsp has been positively associated with lipid metabolism since it caused decreased cholesterol and LDL levels (S. Li et al., 2021); our results supported these findings. Indeed, these results may indicate that Drsp affects how energy is produced and stored in cells, which might determine the efficiency of chemotherapy drugs (Kubik et al., 2022), as well as the ability of cells to respond to external stresses, estrogen signaling, and androgen signaling pathways (Mauvais-Jarvis, 2011). From this Drsp specific cluster, three genes were prioritized: PLLP, FKBP5 and MYO10. PLLP and FKBP5 were chosen in CAP-RNAseq as they both showed the highest correlation between mRNA and protein expression levels and showed significantly increased expression in Drsp, D_E2 and DEA samples alone, but not in Aldo or E2 samples alone.

PLLP (proteolipid plasmolipin) is a key player in myelin sheath composition and normal functioning of the nervous system. While it is involved in lipid raft production, Notch signaling, and intracellular transport and is possibly linked to diseases including type 2 diabetes, cancer, schizophrenia, and Alzheimer's (Shulgin et al., 2021). In the HPA data, the protein expression of PLLP was at low levels in a few breast cancer patients while it was at low to medium levels in normal breast tissues. However, in the literature, not much information is available on the expression level of PLLP, but a study using RNA-seq data from MDA-MB-231 cell lines showed that PLLP expression increased in brain metastasizing melanoma and breast cancer cells (Sato et al., 2017). It was also shown that PLLP expression was decreased in mouse cerebral cortex by E2 treatment (Humphreys et al., 2014), which was not seen in our analysis.

FKBP5, on the other hand, revealed high levels of protein expression in individuals with breast cancer and medium expression in all normal breast tissues based on the HPA data. FKBP5 is a negative regulator of AKT and was found to be upregulated in the luminal B subtype of breast cancer (Pei et al., 2009; Tong et al., 2023). By creating a complex with the heat shock proteins, FKBP5 is also important for modulating the activity of steroid receptors, including androgen, progesterone and glucocorticoid receptors (L. Li et al., 2011), and has been found to be a sensitive biomarker of corticosteroid response (Sinars et al., 2003). In addition the administration of progesterone in T47D cells increases the FKBP5 levels (Need et al., 2015).

In our analysis, these two non-essential genes showed increased expression by Drsp regardless of the presence of Aldo and E2 suggesting that this could be unique to Drsp

alone and hence could also effect Drsp containing hormone drugs recently became available for use (Guida et al., 2023).

The other prioritized gene, MYO10, was the only essential gene in this cluster, in which gene expression was increased with treatments that contained Drsp. Only a small percentage of breast cancer patients exhibited high or low expression levels of this gene in the HPA data, with the majority of patients displaying medium levels of expression. MYO10 has a function in the controlling of genome stability and inflammation in cancer (Mayca Pozo et al., 2021). Upregulated expression of MYO10 has been linked to increased aggressiveness and metastasis of breast cancer, while MYO10 silencing has been shown to decrease invasion and migration of cells (Cao et al., 2014). Furthermore, MYO10 depletion caused the inhibition of breast tumor growth via lowering genomic instability and inflammatory signaling. In general, MYO10 stimulates the growth of tumors by causing genomic instability and establishing an environment that is immunogenic for immune checkpoint inhibitors (Mayca Pozo et al., 2021). Thus, this essential gene emerges as a promising candidate for further research and clinical investigation, offering potential as a diagnostic biomarker and a therapeutic target in Drsp-containing HRT-inducible breast cancer.

On the opposite (mirror) cluster, downregulated genes with Drsp were involved in a network of pathways, known to be critical for maintaining cellular homeostasis and preventing cancer progression. Given the role of the WNT pathway in the controlling of cell proliferation and fate decisions (J. Liu et al., 2022), its downregulation may have wide-ranging implications for tumor progression and therapeutic response. Downregulation of genes in cancer pathways may also be indicative of a vulnerability

that cancer cells take advantage of or may represent attempts by cells to prevent carcinogenesis. Among these genes, SOX2, an essential gene, and CLU were prioritized, both of which exhibited high correlation of expression between their mRNA and protein levels, each. Their expression was reduced by all groups except E2 alone. Moreover, only a small percentage of breast cancer patients had protein expressions of both genes detected in HPA, which was absent from normal breast tissues.

SOX2 is a crucial biomarker for cancer stem cells and has a role in several tumor processes such as treatment resistance, recurrence, invasion, differentiation and cell proliferation (Saigusa et al., 2009; S. Yang et al., 2015). In TNBC, SOX2 encourages cell proliferation and metastasis (P. Liu et al., 2018). It was also shown that breast cancer tissues express high levels of SOX2, which correlates with a worse prognosis, but normal breast tissues do not express it (Abd El-Maqsoud & Abd El-Rehim, 2014; Dey et al., 2022). In another study using tamoxifen-resistant cells which express high level of SOX2, the inhibition of SOX2 caused the decrease in the size of cell population and restored the cell sensitivity to tamoxifen (Piva et al., 2014). The reduction of expression of SOX2 by Drsp is an interesting finding that might have importance since this does not happen when E2 is given alone. Therefore, SOX2 can be a target to prevent recurrence of breast cancer, by using Drsp alone pills, warrants further in vitro and in vivo studies.

Clusterin (CLU) has function in the immunological response, cell survival, membrane transport recycling, cell adhesion, and apoptosis (Pucci & Bettuzzi, 2009; Trougakos & Gonos, 2002), and has been linked to tumor immune cell infiltration and the prognosis of patients with breast cancer, suggesting that it can be a biomarker for immune cell infiltration and biomarker of immune cell recruitment into breast cancers (P. Yang et al.,

2023). While there are no findings about the relationship between Drsp and immune cell infiltration in the literature, its reduction by Drsp could also reflect potential for Drsp use in treating immune cell infiltration. In general, since these genes are found as biomarkers of Drsp treatment in breast cancer, they can be used as diagnostic and therapeutic targets to check the effect of Drsp reduction on cells.

4.6.3 Synergistic Effects of Drsp and E2

Combination therapies are one of the more recent solutions to cancer therapy (D'Alonzo et al., 2019). This is because combinatorial use of two distinct agents may lead to effect greater or lesser than the individual effects of the drugs used (Muir, 2015). This synergistic action may lead to better therapies as well as enhance the pro-cancer actions of drugs/hormones.

In some of the expression clusters, there was evidence that Drsp and E2 worked additively or synergistically; this has been one of the important novel findings of this thesis. In these synergistic activity clusters, both agents caused an increase (or a decrease in the mirror cluster) on a gene's expression, but their combination caused equal or larger effect sizes than their simple addition. Interestingly, presence of Aldo (DEA group) had a slight effect in these clusters in the opposite direction hence reduced the effect caused by the two agents, i.e., D_E2. The genes with synergistically downregulated expression were linked to cholesterol homeostasis, inflammation, iron homeostasis, oxidative metabolism, and long-term potentiation. The genes that were synergistically upregulated belonged to cell cycle control, proliferation and preserving genomic integrity. The cluster with the upregulated genes upon D_E2 treatment contained the highest number of essential genes

which were expressed at moderate to high levels. These findings suggested that D_E2 further downregulated metabolic activity while upregulated proliferation, while Aldo could revert these effects to a certain degree. The effect of a relatively high dose of Aldo could be due to balancing out the effect of Drsp, where the former is an agonist, and the latter is an antagonist of MR (Palacios et al., 2006). However the effects could be executed by more than one receptor, e.g., GR, which Aldo at high doses has an effect (Vassiliou et al., 2021). These two mirror clusters and their association pathways may be modulated by D_E2 but the effect could be reverted by Aldo exposure, which can also be mimicked by high salt use (Catena et al., 2017). It is intriguing to speculate that D_E2 effects could be modulated by dietary salt intake and needs further study.

When it comes to the prioritized genes of these clusters modulated by D_E2 in a synergistic manner, I have selected FOXA1, known to be targeted by ESR1 (Chaudhary et al., 2017). FOXA1 also showed a relatively high correlation between mRNA and protein levels, and was reduced by Drsp, Aldo and E2 alone samples, but much more so by the D_E2 group yet its expression was normalized in the DEA samples. FOXA1 appears to be a key mediator in nuclear receptor action as the level of FOXA1 expression are closely correlated with the levels of not just ER but also AR (Rangel et al., 2018). By remodeling chromatin and serving as a licensing factor, FOXA1 allows nuclear receptors (NRs) to access their corresponding cis-regulatory areas, which is essential for the attachment of ER and AR to the chromatin (Seachrist et al., 2021). In luminal breast cancers, FOXA1 and ESR1 expressions have a positive correlation between them (Ross & Perou, 2001; Van ’T Veer et al., 2002). It was shown that the inactivation of FOXA1 disrupts the global ER-chromatin interaction and transcriptional activity (Carroll et al., 2006; Glont et al., 2019). In the literature, FOXA1 consensus sites were shown to be

enriched in the PR-induced ER cistrome, indicating that FOXA1 expression levels control the PR-driven ER-related cell identity alterations (Mohammed et al., 2015). However, there is not a specific analysis for the FOXA1 expression levels by progesterone treatment. When E2 is provided, ESR1 levels decrease (Mori et al., 2021) hence this is reflected in FOXA1; yet when together with Drsp, the receptors could be sensitized even further.

From the mirror cluster (aka. increased expression of genes with D_E2), I prioritized FKBP4, which is a member of the immunophilin protein family, that controls the activation of steroid receptors and play a crucial part in the development of several hormone-dependent malignancies, such as prostate and breast cancer (Zong et al., 2021). In this case, FKBP4 showed increased expression by D_E2 sample and E2 alone, while Drsp and Aldo alone did not cause any increase. However, when Aldo was given into combination, the increase caused by DEA was the less than the increase caused by D_E2 and E2 alone samples. Increased FKBP4 expression has been widely observed in breast cancer cell lines and has been associated with worse survival outcomes in luminal A subtype breast cancer patients, highlighting its potential as a prognostic marker (Xiong et al., 2020; W. S. Yang et al., 2012). Through its interactions with PI3K, FKBP4 has been shown to enhance Akt activation via mTORC2 and PDK1, and the depletion of FKBP4 selectively inhibits growth and proliferation of cells in triple negative breast cancer (Mangé et al., 2019). Therefore, FKBP4 can be used as a biomarker in breast cancer that is induced by D_E2 combination however use of Aldo or high salt diet may revert the expression changes of this gene and this has to be studied further.

4.6.4 Prioritized Genes and Pathways Affected by E2

There were also clusters containing genes that were affected by E2 similarly regardless of other hormones. These genes showing downregulation were found to be enriched in PTK6 expression pathway which involves in several cellular processes, including signal transduction, cell growth, and migration (Lemmon & Schlessinger, 2010). The upregulated genes in the other mirror cluster were linked to estrogen response pathways, cell cycle regulation, DNA replication and repair, and cell division, known pathways reflecting the proliferative activity of E2 in T47D cells (Jiang et al., 2012).

PDZK1 was prioritized as the upregulated gene by E2-containing samples. As a scaffolding protein, PDZ domain containing 1 (PDZK1) has been linked to breast cancer acting as an oncogene because it controls how plasma membrane proteins are expressed on the surface of epithelial cells (Michailidou et al., 2015). PDZK1 is known as an estrogen-responsive gene and its overexpression causes an increase in cyclin D1 which plays a role in cell division (Dunbier et al., 2010; Kim et al., 2014). While estrogen can activate the PDZK1 gene, when 17β-estradiol is absent, it has been observed that enhanced expression of PDZK1 allows MCF-7 cells to proliferate free of surface attachment, suggesting that PDZK1 may have carcinogenic potential (Kim et al., 2014). Our RNAseq data analysis via CAP-RNAseq has demonstrated that this gene was not modulated by Drsp and Aldo, and its increase was solely dependent on E2.

NDRG1, which was downregulated by E2 while being upregulated by Drsp, showed reduced downward trend with the combination of Drsp and E2 when compared to E2 alone. This indicated that Drsp could modulate significant effects of E2 antagonistically. It has been proposed that NDRG1 serves as a prognostic marker for individuals with

inflammatory breast cancer (Ghafouri-Fard et al., 2023). It was also shown that silencing of NDRG1 caused the decreased proliferation of cells in TNBC (Mishra et al., 2020). In another study, expression of NDRG1 has been found to increase in response to progesterone and resulted in decreased kinase activity leading to reduced migration and invasion of breast cancer cells (Godbole et al., 2018). In my analysis, Drsp also increased this gene's expression but combinatorial treatments and E2 alone caused the decrease in its expression. These findings suggest that NDRG1 needs to be further investigated as a candidate for the breast cancer progression and for the potential interaction I discovered between Drsp and E2.

E2 treatment also downregulated the essential genes in the cluster where the genes were involved in the ubiquitin-dependent protein catabolic process, highlighting the critical role in protein turnover and cellular homeostasis (Bedford et al., 2011). The essential gene from this cluster, ERBB2, was prioritized because it is a marker for one of the important molecular subtypes of breast cancer, HER2+ (amplified) (Moasser, 2007). In the literature expression levels of ERBB2 was shown to go down in the E2 treated cells (Hurtado et al., 2008). While it showed slight downregulation by Drsp, it was highly downregulated in E2-containing samples.

One characteristic that distinguishes ERBB2-positive breast cancer from other subtypes of the disease is the amplification of the ERBB2 gene (Idossa et al., 2023). About 30% of breast cancer cases had overexpression of ERBB2 (Slamon et al., 1989). It was also shown that estrogen-ER and tamoxifen-ER complexes may directly inhibit ERBB2 transcription, which indicates a clear connection between ERBB2 expression and estrogen signaling, as this suppression is mediated by a specific regulatory element found

in the ERBB2 gene itself (Hurtado et al., 2008). This finding was also supported by our analysis which displayed a downregulation of the gene by E2-containing samples.

4.6.5 Genes Affected by DEA

The potential antagonism between Drsp and Aldo could be further studies in clusters where DEA has more effect then DE alone. In some clusters, genes were most affected by DEA samples, while others showed less effect and Aldo alone did not show much effect. The enrichment analysis of downregulated genes by DEA revealed the pathways related to apoptosis and protein localization, which indicates that the genes involving in the process of apoptosis inhibition and protein transportation to correct destinations were downregulated. In the opposite cluster of this cluster, genes were enriched in the pathways related to cell division event.

BCL2L1 from the cluster displaying downregulation by DEA samples was the essential and highly correlated gene between mRNA and protein levels. BCL2L1 expression was significantly reduced by DEA and E2 samples alone, while the reduction caused by the D_E2 sample was less than that caused by DEA and E2 samples alone, and Drsp and Aldo alone did not change its expression. It seems that Drsp affects E2 and DE2 cause less decrease, but when Aldo is given, it antagonizes the effect of Drsp so that DEA can cause as much decrease as E2 alone. BCL2L1 is an anti-apoptotic protein and was found to be essential for the resistance of ER+ breast cancer cells to treatment (Alcon et al., 2021). It was discovered that breast cancer tissues had higher levels of BCL2L1 gene expression, which was positively correlated with the presence of estrogen receptors (Olopade et al., 1997). Moreover, the presence of BCL2L1 was also shown to be

significantly correlated with the endocrine therapy (Gee et al., 1994). BCL2L1 level is frequently elevated in tamoxifen-treated ER+ breast tumors, which reduces the immediate tumor response to tamoxifen and increases long-term tamoxifen resistance (Musgrove & Sutherland, 2009). Therefore, reducing the BCL2L1 expression by DEA treatment can be a potential therapy in the breast cancer yet use of Aldo in the presence or absence of tamoxifen has to be tested whether Aldo could relieve the resistance induced by tamoxifen.

Another essential gene, CEP85 was found to be expressed more in the DEA samples, although DE2 and E2 alone caused expression levels to be lower than those of DEA. Although there is not much information about the CEP85 in breast cancer, CEP85 is an important gene for cancer cells to migrate in a certain direction and downregulation of it disrupts cell migration which is an essential cellular process (Y. Liu et al., 2020). CEP85 as a centriole duplication factor binds to STIL, thus activating PLK4 for efficient centriole assembly which is expressed abnormally in breast cancer (Y. Liu et al., 2018; Parsyan et al., 2021). PLK4 activity dysregulation has serious implications for centriole formation and is associated with the advancement of cancer (Basto et al., 2008). Low levels of PLK4 function can hinder centriole replication, while excessive PLK4 activity results in the overduplication of centrioles (Habedanck et al., 2005). Many cancer cells frequently exhibit centrosome amplification, and centrosome overduplication is found to be linked to irregularities in the mitotic spindle in breast tumors, potentially leading to genetic instability and the emergence of more aggressive characteristics as the tumor advances (D'Assoro et al., 2002). Since CEP85 expression was found to be increased by D_E2 and specifically more with DEA treatments in my data, it can be concluded that CEP85 may be synergistically affected by interactions between MR and ESR1 signaling. Therefore,

CEP85 can be a potential diagnostic biomarker in the breast cancer patients who used hormone therapy with D_E2; and needs to be further investigated with respect to its variation based on the high salt diet and its effects on its expression.

4.6.6 Effect of Aldosterone to Hormone Therapy

Aldosterone in some clusters showed distinct patterns when used in the combination. Most of the genes showed increased or decreased expression in response to D_E2 samples. However, treatment with DEA reduced this effect by almost half compared to the impact observed with D_E2 treatment. Genes that increase with D_E2 but reverse with DEA were found to be involved in cell cycle control, DNA replication and proliferation, while the opposite genes were enriched in cholesterol homeostasis and inflammation, iron homeostasis and oxidative metabolism. Moreover, genes of other clusters showed significant and more pronounce upregulation or downregulation by DEA treatments compared to other samples in the data. These genes were found to be enriched in the pathways related to apoptosis, protein localization and cell division.

Aldosterone which is a salt-retaining hormone in RAAS has a role in the regulation of blood pressure and development of renal and cardiovascular diseases (Ames et al., 2019). One important factor in the rise in cardiovascular and renal diseases among postmenopausal women has been identified as increased RAAS function caused by the decrease production of estrogen (Castelli, 1988). In the menstrual cycle of women, there is a negative correlation between circulating estrogen concentrations and blood pressure (Dubey, 2002). Adrenal renin-angiotensin-system expression and adrenal gland function are regulated by estrogen (Krishnamurthi et al., 1999). According to findings of a study,

estrogen therapy reduces the synthesis of aldosterone by upregulating AT2 receptors and downregulating AT1 receptors, which are the receptors expressed by adrenal zona glomerulosa (Jezova et al., 2003; Kakar et al., 1992; Macova et al., 2008). Moreover, certain progestins have little effect or even block the RAAS, while others cause natriuresis by acting antagonistically at the mineralocorticoid receptor (Crane & Harris, 1969; Oelkers et al., 1974). This indicates that the effects of progestins on RAAS components may differ depending on the kind of synthetic molecule and whether progestins are administered alone or in conjunction with estrogen. A study showed that HRT significantly increase the plasma aldosterone/renin ratio by increasing the aldosterone, which can be explained by the increase of angiotensinogen by estrogen and the antagonistic activity of progesterone at MR which induce natriuresis (Ahmed et al., 2017).

In the studies of this thesis, Aldosterone has been used at a concentration of 100 nM (10^{-7} M, supra-physiologic), which was previously used in thesis of Seniye Targen, PhD (Targen, 2020; Bilkent University). This dosage of aldosterone is also equivalent to high salt diet. In some cases, it was observed that aldosterone addition (DEA) caused more upregulation or downregulation of genes, while in other cases, DEA reduced the effect of D_E2 by almost half. This highlights the importance of high-salt diets used in combination with hormone therapy. A high-salt diet is one of the identified risk factors for cardiovascular disease and is thought to cause blood vessel inflammation and raise arterial pressure, leading to atherosclerosis and heart attacks (Brown et al., 2009; Willebrand & Kleinewietfeld, 2018). Aldosterone acts in hypertension by activating the distal nephron's epithelial sodium channel (ENaC) to increase the sodium reuptake. An excessive consumption of salt inhibits the systemic axis of renin-angiotensin-aldosterone (Shimosawa, 2013). This can result in higher blood pressure as the regulation of sodium

balance is disrupted. In the breast cancer, it was observed that using high salt diet increases the progression of the cancer in mice and MCF-7 cells (Chen et al., 2020). However, in a study investigating the proliferative effect of aldosterone in MCF-7 cells (no MR overexpression), cellular growth remained similar in cells treated with aldosterone across a range of concentrations (10^{-6} M to 10^{-14} M) after 48 and 72 hours (DE Azevedo Delou et al., 2014). It was reported in a pig kidney cell line that 50 and 100 nM of aldosterone caused significant increase in Nrf2 activation which is an important regulator of the cellular antioxidative defense, while 10 nM of aldosterone caused slight increase (Queisser et al., 2014). In another study, 1 nM, 10 nM and 100 nM of aldosterone levels did not cause significant effect on the proliferation in human peritoneal mesothelial cells, while 10 nM and 100 nM of aldosterone showed significant increase in the expression E-cadherin and α-SMA with the highest increase occurring with 100 nM aldosterone (M. Yu et al., 2015). Aldosterone also induced apoptosis in HUVECs in a dose-dependent manner and a significant increase was observed with 100 nM aldosterone (Zhang et al., 2016).

As confirmed in the literature, aldosterone can be affected by estrogen and act through estrogen related receptors. The effect observed by DEA combination can be related to crosstalks between receptors in the cells. Moreover, it was found that aldosterone increases the genes related to glycolysis in my MCF-7 data, and this was also shown in the literature. Glycolysis-related genes were found to be upregulated in the MR-inducible MCF-7 cells treated with 10 nM aldosterone (Doan et al., 2020). Therefore, the affected pathways and genes from DEA treatments need to be further analyzed to investigate the influence of high salt diet on the hormone therapy.

4.6.7 Genes Affected by D_E2 and DEA But Not D, E2, or Aldo

There was expression in which the genes were affected by D_E2 or DEA treatments in the same amount but not by D, E2 or A alone. These clusters also reflected the synergistic effects between PR, ESR1 and/or MR. The downregulated genes by D_E2 and DEA were found to be involved in the pathways associated with lipid metabolism, signal transduction mediated by G proteins, and environmental response such as UV-induced DNA damage, whereas the upregulated genes displayed an enrichment in the pathways related to cell division, energy metabolism, and estrogen response. Accordingly, these clusters of genes could be thought of increasing proliferative activity while altering cellular homeostasis and DNA damage negatively.

From the cluster showing upregulation of the genes, several essential genes were prioritized. One of them was RUNX1, which increased by DEA and D_E2 treatments, while E2 alone caused only a slight increase. RUNX genes are significant in cancer as they function in many systems as tumor suppressors or oncogenes. RUNX1 is the most often expressed member of the RUNX family in human breast epithelial cells (L. Wang et al., 2011). Mutations in RUNX1 are primarily observed in ER+, luminal breast cancer subtype, which suggests that RUNX1 has a tumor suppressor function (Van Bragt et al., 2014). However, in triple negative breast cancers (TNBCs), increased RUNX1 expression is associated with poor outcomes, indicating a positive, oncogenic function for RUNX1 in this subtype of breast cancer (Ferrari et al., 2014). The deletion of RUNX1 in MCF10A cells causes the increased cell proliferation and irregular morphological development (L. Wang et al., 2011). Compared to normal breast epithelial cells, breast cancer has lower levels of its expression, which gradually declines as tumor aggressiveness increases (Kadota et al., 2010). The decreased expression of RUNX1 was shown to correlate with

higher-grade primary breast tumors, which indicates the important role of RUNX1 in tumor progression and aggressiveness (Kadota et al., 2010) While RUNX1 has been identified as acting pro-oncogenic in TNBCs due to the poorest prognosis, this effect could not be found in tumors that express the estrogen receptor (Ferrari et al., 2014). It's possible that RUNX1 can alter or disrupt ER signaling, which could explain this disparity, though the exact mechanism by which RUNX1 inhibits breast cancer is yet unknown (Chimge & Frenkel, 2013). The synergistically increased RUNX1 in the presence of Drsp and E2 suggested that this could be specific to the interaction of E2 with Drsp since Aldo did not alter this. In the case of RUNX1, D_E2 treatment regardless of Aldo (or salt status) can be used to increase its expression which can cause the inhibition of breast cancer. It is also a measurable biomarker since it has high correlation between mRNA and protein levels.

Another essential gene ARF6 from the same cluster showed similar expression pattern with RUNX1. ADP-ribosylation factor (Arf) family member ARF6 is known as a key player in the regulation of membrane trafficking and cytoskeletal architecture (Donaldson, 2003). ARF6 promotes migration and cell invasion in breast cancer cells after being triggered by epidermal growth factor (EGF) (Morishige et al., 2008). Despite the fact that mRNA levels of ARF6 did not correlate with invasiveness, there was higher expression of ARF6 protein in highly invasive breast cancers compared to mildly invasive or noninvasive breast tumors and normal breast epithelial cells (Hashimoto et al., 2004). However, its mRNA and protein correlation was found to be moderate in DepMap data, and its protein expression was higher in breast cancer patients compared to normal breast tissue in HPA data. The findings about ARF6 indicate that ARF6 in breast cancer acts as a tumor promoter by breaking down cellular adhesion structures and resulting in more

aggressive cancer phenotypes (R. Xu et al., 2015). Thus, based on our analysis, ARF6 can be a biomarker in the breast cancer induced by HRT containing D_E2 but not Drsp or E2 alone.

CITED2 and TPM4 were chosen from the opposite mirror cluster where DE2 and DEA caused reduced expression in both groups, although E2 by itself had no impact. CITED2, a member of the CITED family, is a target gene of MYC. It has the ability to directly engage with the estrogen receptor and regulate various cellular processes involved in differentiation and development (Andrews et al., 2000; Chou et al., 2012). Its mRNA and protein levels were moderately correlated (r=0.65), and its protein expression was high in all breast cancer patients of HPA data. qRT-PCR has been used to investigate the relationship between CITED2 mRNA expression and the prognosis of ER-positive human breast cancer, but the findings are conflicting and CITED2's importance remains unclear (Lau et al., 2010; Van Agthoven et al., 2009). MCF-7 cells have significantly increased migratory characteristics and proliferative activity when CITED2 is expressed (Minemura et al., 2016). When compared to non-neoplastic mammary glands and normal mammary epithelium, CITED2 immunoreactivity and mRNA expression levels were considerably greater in breast cancer and invasive ductal carcinoma samples, respectively (Lau et al., 2010; Minemura et al., 2016). Since the expression of CITED2 is negatively associated with survival (Lau et al., 2010, 2013), the downregulation of CITED2 by D_E2 or DEA treatments can be significant therapy.

Tropomyosin 4 (TPM4) is an actin-binding protein from the tropomyosin family and is found in the cytoskeleton of nonmuscle cells (Lin et al., 2008). Protein expression patterns have indicated that TPM4 may be a carcinogen, and abnormal protein amounts are seen

in a number of malignancies (D. Li et al., 2006). It is linked to the worse prognosis and progression of breast cancer (C. Wang et al., 2019). Therefore, like CITED2, the downregulation of TPM4 by D_E2 or DEA treatments can be used for the suppression of cancer. DepMap mRNA-protein correlation and HPA revealed a moderate correlation between mRNA and protein levels and high protein expression in breast cancer patients.

4.7. Effects of Steroid Hormones in MR-Overexpressing MCF-7 Cells

4.7.1 Essentiality and Enrichment of the Clusters

The MCF-7 RNA-seq data analysis has yielded important insights on the distribution and role of essential genes under various treatment conditions. The highest percentage of essential genes was found in the cluster which displayed an increasing pattern in the gene expressions in all samples but highest in E2_Aldo samples. The clusters which had variations between the replicates of samples contained the lowest number of essential genes.

Essential genes in the clusters showing an increasing expression pattern in all samples were implicated in critical cellular functions such as ensuring proper protein folding and the degradation of misfolded proteins, which are pivotal for maintaining cellular homeostasis. Additionally, these genes also involve in the pathways related to cell cycle regulation and progression, as well as cell growth and responding to estrogen signals. The pathways associated with cell growth, division, cell cycle regulation, and progression were also revealed for the essential genes of a cluster that showed an increase in expression exclusively in samples containing E2. This emphasizes the importance of

cellular division and proliferation in response to estrogen and highlights the impact of the hormone on these biological processes.

While these clusters shred common terms, clusters showing the downregulation by D_E2 samples compared to control did not share any common term with any clusters. The essential genes in the cluster, which exhibited reduced expression in all samples containing Aldo but were unaffected by Drsp, were found to be involved in immune signaling and antiviral responses. These genes showed greater reduction in E2_Aldo samples, but when Drsp was added to them, the reduction was less than in E2_Aldo, indicating that Drsp might counteract some of Aldo's suppressive impact on these immune-related genes. Like this cluster, genes within other cluster that showed a similar pattern but did not reverse the effect of E2_Aldo upon Drsp addition were linked to processes such as cell growth, differentiation and apoptosis. Moreover, genes that were reduced in all samples but more so in Drsp treatment alone were enriched in pathways related to cell signaling and stress responses. This suggests that Drsp has a particular effect on controlling signaling pathways and cellular responses to stress.

On the other hand, non-essential genes within the clusters showing increased expression in D_E2 sample were associated with the estrogen response pathways. In the cluster displaying significant increase in the samples containing Aldo, non-essential genes were enriched in pathways related to lipid metabolism, adaptive mechanisms of the cell to stressors, and androgen-mediated processes. However, the genes downregulated by Aldo were found to related to immune response.

4.7.2 Prioritized Genes Modulated by Drsp

The investigation of how Drsp treatment affected gene expression patterns indicated that it had significant effects on some clusters, including one pair of mirror clusters that contained a small number of essential genes. Some genes stood out among these clusters because of their significant changes in expression and the implications for the breast cancer.

Among them, CHI3L2 was selected since it showed significant increase by Drsp. CHI3L2, which is a M2 macrophage marker, has been found to have the ability to induce the monocytes migration and act as a pro-angiogenic and monocyte recruiting factor (Litviakov et al., 2018). Studies have presented that its enhanced expression is indicative of a higher risk of distant metastases and a poor response to neoadjuvant treatment for breast cancer (T. Liu et al., 2018). This result and its higher levels of protein expression in breast cancer patients compared to normal breast tissues obtained from HPA data highlights its possible link to breast cancer development and resistance to treatment.

Another gene selected from same cluster, ENO3, did not exhibit any significant protein expression in both breast cancer and normal tissue based on the HPA analysis. However, its involvement in the fatty acid metabolism and glycolysis/gluconeogenesis pathway and its crucial importance for the metabolism of both glycogen and cholesterol (Cui et al., 2021) emphasize its importance in cellular metabolism. There are not many studies on the association of ENO3 with breast cancer, but some glycolysis-related genes, including ENO3, have been associated with poor breast cancer patient prognoses (C.-Y. Wang et al., 2023). Therefore, the upregulation of ENO3 by Drsp needs further investigation for its potential effects on cancer metabolism and patient outcomes.

From another cluster showing decrease expression by Drsp, CYP1B1 was identified as a gene of interest due to its reduced expression in response to Drsp and D_Aldo treatments. Since the other treatments did not cause any change in the expression of this gene, this suggests that only Drsp can affect this gene, while E2 suppresses the effect of Drsp on this gene. As a E2 hydroxylase involved in the metabolism of possible carcinogens, CYP1B1 is important in breast cancer, and its expression was found to be higher in tumors compared to normal tissues (McKay et al., 1995). It was shown that CYP1B1 increases the proliferation, invasion and migration of cells in MCF-7 cell (Kwon et al., 2016), and the increase expression of CYP1B1 might result in accumulation of estrogen metabolites, which could harm DNA either directly or indirectly (Yager, 2000). Also, CYP1B1 is involved in the metabolism of progesterone and testosterone which are the poor substrate for CYP1B1 (F. Li et al., 2017). Drsp's ability to inhibit CYP1B1 expression offers a possible way of preventing cancer cells from proliferating, invading, and migrating, possibly via controlling the accumulation of estrogen metabolites.

Another gene from this cluster, TXNRD1, was selected since it was highly expressed and showed increased protein expression in cancer patients compared to normal tissues. TXNRD1 is the key player in the regulation of oxidative stress, was associated with worse prognosis in breast cancer (Cadenas et al., 2010). TXNRD1 is overexpressed in higher-grade tumors and has a positive correlation with both tumor size and grade (Patwardhan et al., 2024). The effect observed by Drsp was enhanced when E2 was added. Since other E2-containing treatments also showed similar results to Drsp alone, it can be concluded that Drsp and E2 work synergistically on this gene, further reducing its expression.

Therefore, a therapeutic approach targeting this gene by Drsp aonle or combination with E2 could potentially reduce tumor aggressiveness and improve patient prognosis.

4.7.3 Prioritized Genes Modulated by Aldosterone

The effect of Aldo was observed in a cluster including genes that were downregulated by D_Aldo and E2_Aldo samples. While same reduction caused by D_Aldo or E2_Aldo was observed in DEA samples for some genes, DEA reduced the effect seen with D_Aldo or E2_Aldo for other genes.

Because of their specific patterns of expression, RELB and PREX1 become genes of interest within this cluster. RELB was chosen due to its essentiality and strong selectivity. It demonstrated considerable downregulation with D_Aldo and E2_Aldo treatments, which was partly restored when all hormones were combined, even though it remained unaffected by Drsp and D_E2 alone. RelB expression has been demonstrated at high levels in aggressive breast cancers, especially in TNBC cells compared to MCF-7 cells, and RelB increases G1/S transition and/or decreases apoptosis by upregulating Bcl-2 and Cyclin D1 resulting in increased proliferation of breast cancer cell (M. Wang et al., 2020). Therefore, the deletion of RelB inhibits the cell cycle progression and survival, migration and invasion of breast cancer cells, and increased apoptosis (M. Wang & Tang, 2018).

Another gene, PREX1 decreased by all Aldo-containing samples but more so by E2_Aldo and DEA samples compared to D_Aldo. It was shown that the migration and metastasis of breast cancer cells have been linked to PREX1, and the loss of PTEN, which triggers PI3K signaling, resulted in decreased PREX1 protein levels, but pharmacologic

suppression of PI3K increased PREX1 (Sosa et al., 2010). Moreover, expression of PREX1 might be specific to ER+ breast cancer since its protein and mRNA expression were higher in the ER+ breast cancer compared to any other cancer (Dillon & Miller, 2015). Therefore, targeting these genes by Aldo can be potential therapeutic approach for breast cancer.

Moreover, BCL6 gene was upregulated by Aldo. Research showed that expression of BCL6 which functions as a transcriptional repressor was high in the cell lines of breast cancer and promotes cell survival (Walker et al., 2015). BCL6 has antiapoptotic function and its overexpression caused the apoptosis inhibition (Logarajah et al., 2003). Targeting BCL6 with inhibitors resulted in the death of breast cancer cells (Walker et al., 2015). The effect of Aldo on BCL6 needs to be further investigated.

4.7.4 Prioritized Genes Modulated by E2

The downregulated genes by E2-containing samples were enriched in the pathways highlighting the importance of lipid signaling in maintaining cellular integrity and responding to extracellular signals. In contrast, upregulated genes by E2-containing samples were associated with estrogen response pathways, cell division and chromosome segregation during mitosis.

Among the upregulated genes, CD44 was prioritized, which increased its expression by all E2 containing samples and also Drsp alone. CD44 has role in cancer progression, metastasis and relapse of breast cancer, and is a biomarker for cancer stem cells in breast

cancer (Al-Othman et al., 2020). It is an essential regulator of the EMT pathway, which controls the onset and development of tumors (Wu et al., 2015).

Another highly correlated gene between mRNA and protein levels was PGR. In MCF-7 cells, Drsp did not affect its expression but D_Aldo samples reduced it, while others (D_E2, E2_Aldo and DEA) increased it. It is known that E2 stimulates the PGR expression in breast cell (X. Xu et al., 2004), while progesterone downregulates the PGR expression in T47D cells (Alexander et al., 1989).

4.7.5 Cross-Dataset Mapping for the Genes Modulated by Drsp

The investigation of gene expression patterns between T47D and MCF-7 cell lines, following treatments with Drsp and Aldo, provided insightful observations regarding the differential impact of these hormones across different breast cancer cell types. The genes upregulated by Drsp in T47D data were searched on MCF-7 data. These genes were mapped to the cluster of MCF-7 data which did not show any changes by Drsp but increased by Aldo-containing sample. This interesting pattern points to a distinct interaction between Drsp and Aldo's activities and the overexpression of the MR in MCF-7 cells. Specifically, this indicates that overexpression of MR in MCF-7 cells may suppress the gene-regulatory effects typically induced by Drsp, instead amplifying the response to Aldo. In the literature, a study using luciferase reporter construct showed that aldosterone treatment caused more increase of PRE-reporter signal in MCF-7 cells compared to T47D cells while a progestin (R5020) caused more signal in T47D cells (Aarts et al., 2023). These common genes were also related to androgen and estrogen responses. The increased expression of these genes in response to Aldo regardless of other

hormones in MCF-7 cells indicates the cross-talk between estrogen, mineralocorticoid and glucocorticoid receptors regulated by Aldo (Barrett Mueller et al., 2014; Leo et al., 2004), and needs to be further analyzed.

4.8 Evaluating the Interactions of Drsp, Aldo, and E2 in Breast Cancer Cell Lines by Wet-Lab Experiments

This thesis presents an extensive study that highlights the intricate interactions between Drsp, Aldo, and E2 on gene expression in breast cancer cell lines MCF-7 and T47D. The analysis reveals distinct responses depending on the hormone combinations and the cellular environment. The use of 20 nM Drsp in combination with 10 nM E2, validated through qRT-PCR and RNA sequencing, demonstrates a strategic approach to distinguish the subtle effects of hormone treatments on particular gene targets. All qRT-PCR results were analyzed with one-way ANOVA with Tukey's multiple comparisons test. This statistical approach was considered the most appropriate given the nature of the experimental design, where variations between comparison groups could only be attributed to drug treatments. Two- or three-way ANOVA could not be used due to the unavailability of data from single treatment conditions, which is necessary for the factorial design required by this type of analysis.

Key findings from this study include:

Selective Response to Hormone Treatments: Some of the previously known genes were checked by qRT-PCR in T47D and MCF-7 cell lines. PGR, which was used as a positive control of E2 effect, showed increase expression in all E2 containing samples of T47D

cells, while slight decrease was observed by Drsp, which was also seen in its RNA-seq result. The similar result was also observed in MCF-7 cell, but Drsp had no effect at all in this case. Drsp is known as its high affinity to PGR and MR and exhibits anti-mineralocorticoid activity (Oelkers, 2004). While some studies showed that progesterone decreases the expression of progesterone receptor in T47D cells (Lange et al., 2000), Drsp may also have similar effect. However, increased expression of MR in MCF-7 cells may result in Drsp interacting more with MR than with PGR and thus not affecting PGR expression. The expression of progesterone receptor, an increased target gene of the estrogen receptor, is also increased by the estrogen treatment (Lashen et al., 2023), which is consistent with our findings. Furthermore, in a study investigating the different effects in both cell lines, the MCF-7 cell line was not activated by progesterone (or R5020) treatment under estrogenic conditions, whereas the T47D cell line was shown to respond effectively to progesterone (or R5020) in the presence of estrogen (Yu et al., 2017). The differences in the effect of Drsp on PGR expression in both cell lines may also be due to the different amounts of progesterone receptor isoforms present in the two cell lines. Therefore, it is important to measure receptor levels in both cell lines to make a better clarification.

Gene Expression Patterns and Hormonal Influence: In contrast to T47D cells, which had a larger number of affected genes by Drsp alone, MCF-7 cells only had a small number of affected genes, including SGK1, ZBTB16, and HSD11B2. This indicates sensitivity to Drsp and the combination of Drsp_E2 that are particular to cell lines which contains different amount of progesterone receptor isoforms and MR levels, highlighting the significance of cellular context in hormone-based treatment. Consistent with this, it

has been shown in the literature that compared to the MCF-7 cell line, the T47D cell line is more sensitive to progesterone (Yu et al., 2017).

Complex Interactions Between Hormone Pathways: The complex interactions among MR, GR, and ESR1, in particular, how Aldo and E2 could affect target gene expression, such ZBTB16 and PER1 were also shown in this work. Period 1 (PER1), which is an MR target and an important circadian clock gene whose expression has been shown to be decreased in breast cancer, was also validated in this set (Zhao et al., 2016). Its expression was significantly decreased by D_E2 samples, while increased by D_Aldo and DEA, not changed by E2_Aldo sample. This indicates that Aldo and E2 can interact each other. Moreover, glucocorticoid receptor (GR) activation has previously been associated with ER-negative breast cancer (West et al., 2018). Significantly higher GR expression was seen in the presence of Aldo and downregulation of GR was seen by Drsp alone treatment, emphasizing an antagonistic interaction between Drsp and Aldo that may have therapeutic implications.

The intricacy of hormone regulation in cancer cells was shown by the different responses of genes including FKBP5, NDRG1 and BCL6 to treatments with Drsp, Aldo, and E2. The strong response of FKBP5 to Drsp alone in T47D cells and its upregulation by Aldo in MCF-7 cells demonstrate cell type-specific effects of the hormones. Heat shock protein 90's co-chaperone, FKBP5, is essential to the glucocorticoid receptor (GR) complex's formation, regulating GR sensitivity and activity (Binder, 2009). When the GR complex is established, FKBP5 acts as a brake, preventing GR from translocating to the nucleus to serve as a transcription factor until FKBP5 separates from the complex. In addition to the GR, FKBP5 also interacts with the MR, PR, ER, and AR receptors. FKBP5, like its effects

on GR, suppresses MR and PR activity while enhancing ER and AR activity (Gallo et al., 2007; Shrestha et al., 2015). FKBP5 containing Aldo-regulated MR binding site exhibits elevated expression by Aldo (Newfell et al., 2011). It was shown that both MR and GR occupancy are necessary for FKBP5 induction (Petrovich et al., 2014). In our data, due to overexpression of MR in MCF-7 cells, Aldo caused the increased expression of FKBP5. FKBP5 gene needs to be studied in more detail in the complex mechanism of MR and GR. Moreover, the disparity in FKBP5 expression found for the E2_Aldo sample in MCF-7 cells between RNA-seq and qRT-PCR highlights the necessity of using various validation techniques in gene expression research.

In contrast to enhanced expression of NDRG1 by Drsp and decreased expression by E2 in T47D cells, NDRG1's regulation by Aldo and its increased expression with Drsp_Aldo samples in MCF-7 cells demonstrate the gene's sensitivity to hormonal environment. In breast cancer, NDRG1 causing the inhibition of cell migration and invasion is upregulated via the progesterone treatment (Godbole et al., 2018). By controlling the expression of NDRG1 which is upregulated by SGK1 and cell migration, GR in PR-negative cells mediates the progesterone effect. The effect observed by Drsp in T47D cells but not in MR-overexpressing MCF-7 cells may be due to the higher affinity of Drsp for MR compared to PR.

The fact that BCL6 was upregulated by Aldo in MCF-7 cells and increased by Drsp without a noticeable Aldo effect in T47D cells highlights the different functions of these hormones in gene expression when different steroid hormone receptors are presented. Progesterones induce the expression of BCL6 (T. Sato et al., 2014). The BCL6 gene may also be affected by hormones for the same reason observed in the NDRG1 gene.

In conclusion, this study emphasizes the diverse impact of steroid hormones on important genes linked to the pathogenesis of breast cancer. Future research is needed to examine the broader impact of these regulatory patterns driven by various receptor types.

4.9 Modulatory Effects of E2 and Drsp on Estrogen Receptor Binding Sites in Breast Cancer

The ChIP experiment, which used the ESR1 antibody, provides insights on how the binding sites of estrogen receptors (EREs) are influenced by the presence of E2 and Drsp. The effectiveness of the ChIP process was validated by preliminary validations, such as chromatin sonication checks and qPCR validations using the GREB1 and PGR genes, indicating proper experimental execution.

Peak calling analysis revealed differences in peak intensities and distributions between samples treated with E2, Drsp_E2, and control, with samples treated with E2 alone showing more peaks overall. In the analysis of ER peaks in T47D cells, it was shown that peaks were more abundant in the samples treated with both progesterone and E2 combination compared to cells treated with E2 alone (Yu et al., 2017). Since the progesterone used in this study was not Drsp, it can be concluded that Drsp in combination with E2 may affect differently. Through the use of overlapping peak analysis, this difference was further clarified in my analysis. It showed a marked difference in the number of common peaks between samples, with the highest peaks between E2 and Drsp_E2 samples. This highlights the important influence of treatments on ERE change.

Furthermore, differences in background noise levels and antibody binding were found when transcription factors and chromatin-binding proteins were investigated using wide peak analysis in genes such as PGR and GREB1, and these results were compared with external ChIP-seq data from MCF-7 cells obtained from the literature. This comparison highlights the necessity for careful interpretation of ChIP-seq data and highlights differences in experimental findings.

This study illustrates the intricate dynamics of ERE alteration when E2 and Drsp are present, showing both common and distinct pathways impacted by these treatments. The results emphasize the importance of understanding how different steroid hormones and their receptors interact to control gene expression, as this could lead to specific therapeutic treatments in breast cancer and other diseases. However, since our Chip-seq data contains only one replicate of each sample and revealed high background due to antibody, further studies are needed.

4.10 Conclusions

One of the main novel outcomes of this thesis is the development of CAP-RNAseq that introduced a new pipeline in the plethora of online RNA-seq analysis tools, which offers invaluable analyses for the annotation and prioritizing of co-expression clusters. CAP-RNAseq in particular performs annotation of mirror clusters that exhibit inverse patterns with a network-based visualization. Subsequently, genes and/or clusters are prioritized based on factors such as "gene essentiality" and the degree of congruence between the mRNA and protein levels of individual genes or genesets. Moreover, in this thesis I have evaluated multiple published datasets to demonstrate the use of CAP-RNAseq while

analyzing a novel RNA-seq expression dataset of my own identifying the intricate cross talk between ESR1 and MR signaling in breast cancer cells, i.e. T47D and MR-overexpressing MCF7. CAP-RNAseq made the analysis of our breast cancer dataset easier and more productive and identified key co-expression clusters and candidate genes that could be evaluated for effects of hormone replacement use in breast cancer cells.

The critical function of TrkB isoforms in cellular proliferation and differentiation was highlighted by revealing novel correlations and gene expression patterns impacted by NTRK2 overexpression/depletion models through the analysis provided by CAP-RNAseq. Genes upregulated by NTRK2 overexpression are responsible for cell division and growth pathways, while NTRK2 overexpression downregulates genes involved in cell migration as well as immune response and apoptosis. The investigation of how the TrkB.T1 and TrkB.FL isoforms affect gene expression indicates that each has unique roles in regulating cellular processes. TrkB.FL may contribute to reducing cellular migration and invasion capabilities and promoting the cellular growth and energy homeostasis more effectively than TrkB.T1. Moreover, TrkB.T1 upregulated some of the genes essential for DNA replication, cell cycle progression, and cellular division, while they were downregulated by TrkB.FL. The identification of the distinct impacts of TrkB.T1 and TrkB.FL on genes offers novel perspectives on their possible functions in the development and treatment of cancer.

The analysis between senescence and essential gene expression showed that essential genes were mostly downregulated in senescent cells, impacting vital pathways for cellular homeostasis, DNA repair, and cell cycle control. The inhibition of NTRK2 in senescent cells caused the upregulation of genes related to proliferation and survival of cells and the

maintenance of cellular integrity, by reducing the senescence-induced downregulation of genes involved in cell division. This result indicates the complex nature of the senescent state and its effects on cancer development and resistance to therapy. Moreover, novel association between NTRK2 and CCND2 was revealed from these two datasets, showing the NTRK2 functions as a suppressor of CCND2 expression.

Furthermore, the investigation into the interactions of Drsp, Aldo, and E2 in breast cancer cell lines has confirmed the selective response of these cells to hormone treatments. Essential genes were found to be upregulated by the E2 treatments, having roles in cell-cycle regulation, division and growth of cell and DNA replication. Drsp increased the genes related to androgen response, indicating that Drsp can influence the signaling pathways of androgen receptors, which could mimic or counteract the actions of natural androgens, despite having progestin and antimineralocorticoid traits of Drsp, and xenobiotic metabolism, indicating the cell's improved ability to break down and remove foreign substances. The downregulated expression caused by synergistic effect of Drsp_E2 appeared in the genes related to cholesterol homeostasis, inflammation, and the synthesis of cellular energy, while the upregulated genes by synergistic effect of Drsp_E2 were found to be involved in cell division, cytoskeleton and preserving genomic integrity. Moreover, several genes associated with Drsp_E2-affected pathways have been prioritized that could be used as biomarkers for the diagnosis or treatment of breast cancer.

In conclusion, this thesis underlines that CAP-RNAseq can be used as a powerful tool to identify essential genes and pathways in the RNA-Seq data, which was demonstrated by analysis of different datasets.

4.11 Future Perspectives

CAP-RNAseq was designed and developed for the human RNA-seq dataset. However, it can be enhanced to allow the analysis of dataset from other species such as zebrafish and mouse. This would also make it possible to analyze gene expression patterns across species, which facilitate the identification of regulatory mechanisms that are both species-specific and conserved.

Further analyses can be added to CAP-RNAseq using other datasets such as drug response data or transcription factor targets. This would facilitate the discovery of possible biomarkers for drug sensitivity or resistance and help to understand how variations in transcription factor activity impact patterns of gene expression and cellular responses under different circumstances.

Moreover, copy number variation (CNV) can be added to CAP-RNAseq. Since CNVs plays a significant role in understanding genetic interactions, mutational tolerance, and the functional impact of gene amplifications or deletions, CAP-RNAseq has the potential to more precisely pinpoint genes that drive cancer, presenting novel targets for therapeutic treatment, by associating CNV data with patterns of gene expression.

DepMap updates its data quarterly. This regular updating of data allows for the inclusion of new findings and discoveries, ensuring that the most accurate and up-to-date information is available for researchers. Therefore, updated data will be added to CAP-RNAseq.

Additional datasets from HPA such as survival can be added to CAP-RNAseq, enhancing its utility in cancer research and medicine. With this advance, gene expression patterns in connection to patient outcomes might be more thoroughly analyzed, allowing scientists and medical professionals to identify prognostic biomarkers.

The prioritized genes from the RNA-seq datasets are needed to be validated by wet-lab experiments to confirm their importance in their context.

The discrepancies between RNA-seq and qRT-PCR results observed in MCF-7 and T47D should be further investigated.

The cross-talk between the receptors and hormones can be further studied in these cells lines using different experimental procedures. Since differences in the amount of receptors in cell lines can affect the responses to hormones, the level of receptors in each cell line should be measured for a better explanation.

Although concentrations of 20 nM Drsp and 10 nM E2 were used in this study, different ratios of Drsp and E2 are now used in HRT. Therefore, different doses can be used to see their effects on genes.

We will continue to use CAP-RNAseq with different datasets to understand the role of essential genes and senescence in the development of non-alcoholic fatty liver disease and liver fibrosis as well as in breast cancer with respect to hormones and TP53 status.

Appendix A. Assessing the Transient Transfection Stability and Duration

It was important to observe how many days the effect of transient transfection lasted with the vectors since the experiments with transient transfection were aimed to be performed within three days. For this purpose, cells were collected on days 2, 3, and 4 after transfection, and the presence of MR protein was checked by Western blotting with the extracted proteins. According to the results, in the presence of normal media and serum, the presence of MR could be observed with MR antibody only on day 2 (48 hours after transfection) (Figure 79A). Since MR expression could be observed three days after transfection in the previous experimental results, the second experimental plan was designed by applying serum starvation to another group in addition to normal media. In the experiment performed in the second setup, a repetition of the drug-treated setup was performed, and 24 hours after transfection, the cell media was changed, and the cells were serum starved. After 24 hours of serum starvation, cell pellets were collected on days 3 and 4. In a similar experimental setup, normal media was used instead of serum starvation for 24 hours, and cell pellets were collected on days 2, 3, and 4. The protein results obtained from these pellets show that MR protein overexpression was present, although it decreased on day 3 and even on day 4 (Figure 79B). Since STAG was used as a marker in MR expressing plasmids, the Western results were also checked with STAG antibody to confirm that the observed MR expression was due to transient transfection.

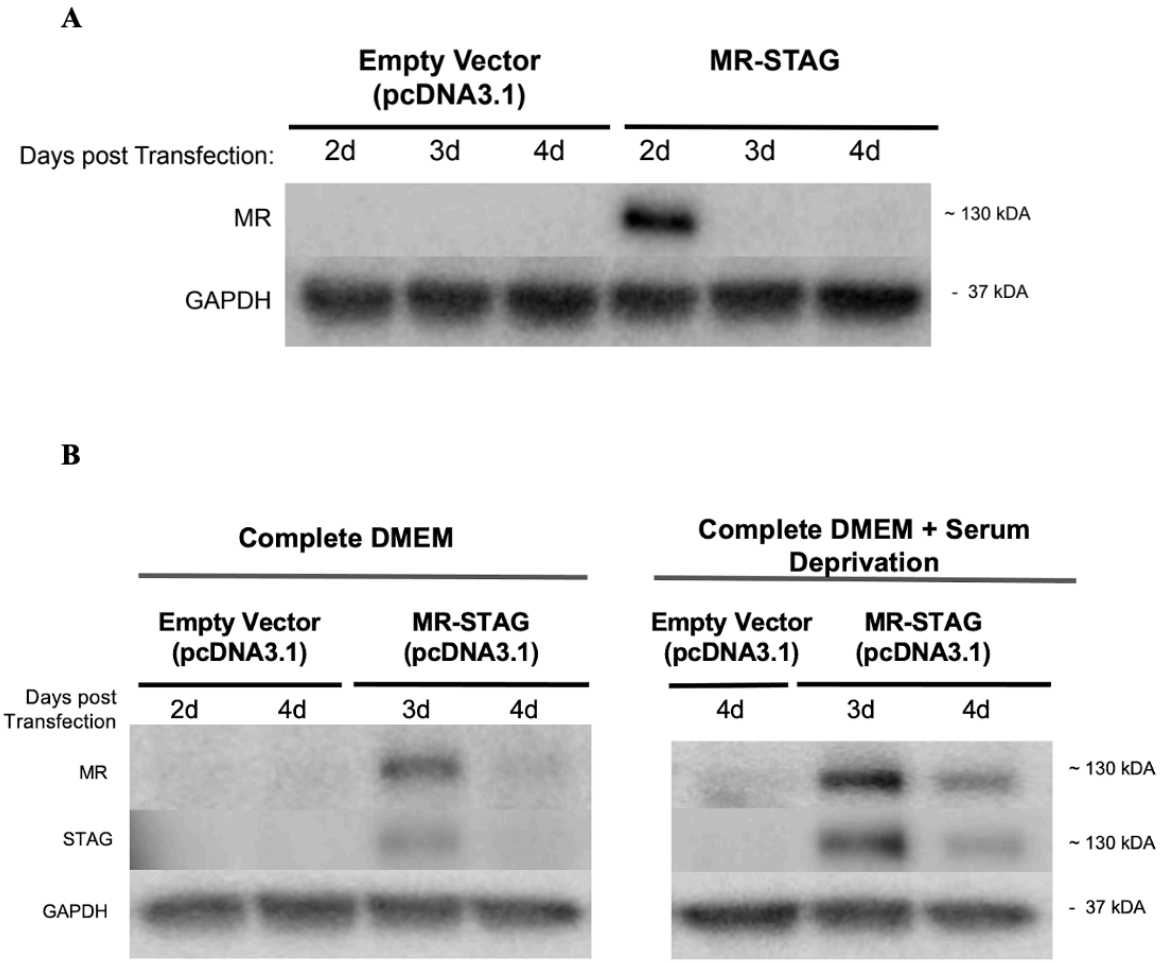


Figure 79. Western blot results with proteins extracted from cells collected on days 2, 3 and 4 post-transfections. A) Western blot results of cells collected 2, 3 and 4 days after transfection under MR and GAPDH control. B) The result on the left shows the amount of MR protein in cells in complete media collected on days 2, 3 and 4, while the panel on the right shows the results of serum-deprived cells using MR and STAG-antibodies. MR and STAG results are consistent.

Appendix B. MTT Assays

MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromidefor) assay was used for the assessing of cell viability at different cell numbers and drug doses.

Method

In order to examine the impact of varying cell counts, MCF-7 cells were seeded in 96-well plates as 3000, 5000, 7000, 8000 and 10000 cells for different groups. Otherwise, MCF-7 cells were seeded in two 96-well plates with 10,000 cells per well. Twenty-four hours after cell seeding, cells were transfected with MR-STAG, lipofectamine or empty vector using Lipofectamine 2000. Then the media of the cells was replaced with fresh media after 6 hours of transfection. The transfected cells were grown in DMEM containing 5% charcoal stripped fetal bovine serum (FBS) without phenol red for 24 hours to exclude the effects of aldosterone (Aldo), estrogen (E2) and similar steroid hormones normally present in serum. In the plates in which the effect of different cell numbers was investigated, control and estrogen (20 nM) treatments were given, while in the plates in which the effect of different drospirenone (DRPS) doses was investigated, 100 nM, 50 nM, 25 nM, 12.5 nM, 10 μM, 1 μM, 0.1 μM and 0.01 μM concentrations of 30 mM stock solution were prepared and given to MCF-7 cells in 6 replicates. Another experiment was also performed for the comparison of Drsp and Drsp-Aldo and Drsp-E2.

After the 24 hours of hormone treatments, MTT application was performed. Five mg of MTT (12 M) was dissolved in 1 ml of PBS (1X) and filtered using 0.22 μm filter into 10 ml of complete phenol red-free medium to get rid of any powder precipitation. The old

medium was removed and 110 µl of a mixture of MTT and DMEM solution (100 µl of DMEM and 10 µl of MTT) was added to each well. The plate was incubated for 4 hours in the incubator at 37°C. During the incubation period, 1 g SDS was dissolved in 10 ml of 0.01M HCl. After the incubation period, SDS/HCl solution was given as 100 µl to each well and the plate was then kept at 37°C for 16-17 hours. Then, the readings were taken at 570 nm by using the Microplate Spectrometer. The results are expressed as percentages based on the average of the wells containing MTT only and the wells belonging to the control group.

Results

Effects of cell number on cell viability

MTT experiments were performed in MCF-7 cells expressing EV and OV-MR in the presence of E2 to determine the cell number to use for future studies (Figure 80). MTT results were analyzed by three-factor ANOVA using the following criteria: drug effect (EtOH or E2), transfection (OV or EV), and cell number (3K, 5K, 7K, 8K, and 10K). As seen in Figure 80, the drug effect varied according to the number of cells seeded. The transfection did not show a significant change. Therefore, it was decided to perform the MTT studies with 5K or 8K.

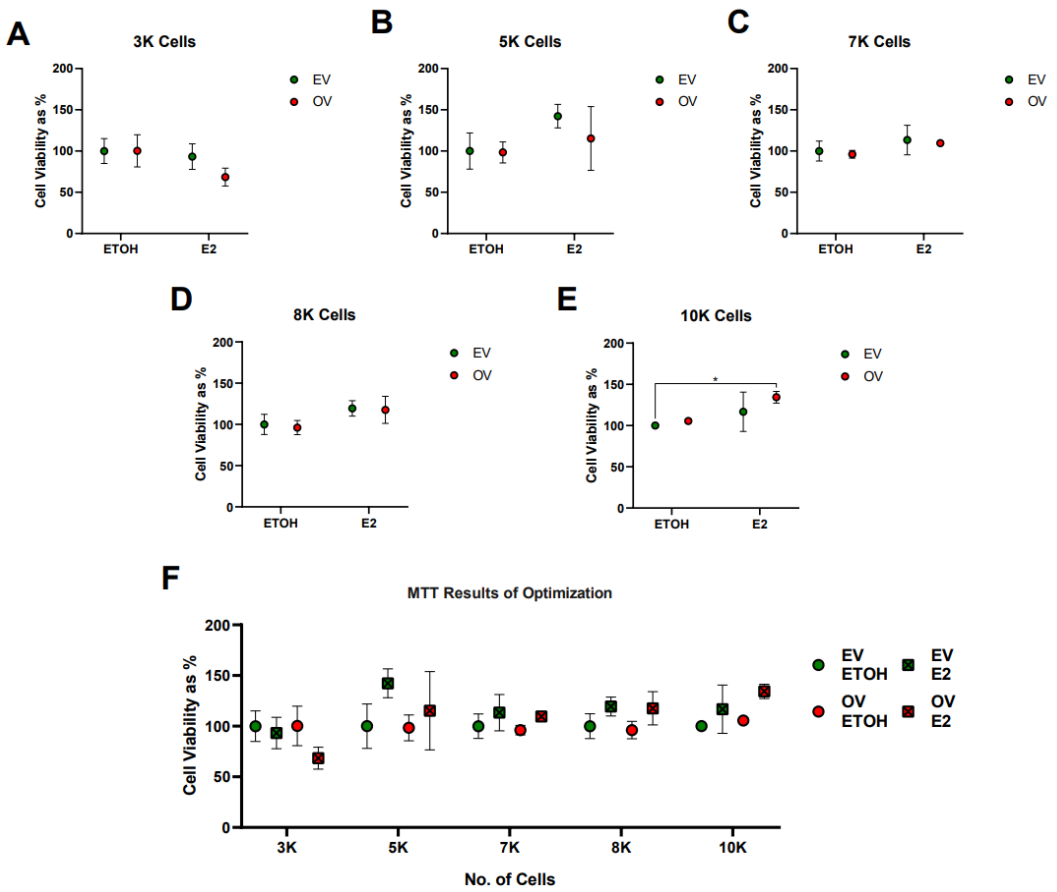


Figure 80. MTT optimization results. Analysis of the effect of transfection (EV or OV) and E2 (20 nM) in MCF-7 cells seeded with different cell numbers.

To further investigate, another MTT experiment was performed using more replicates with 5K and 8K cells threated with estrogen concentrations of 10 nM and 20 nM. Significant differences in cell viability were observed only in the groups seeded with 5K cells (Figure 81). All OV transfected cells consistently exhibited an increase in cell viability compared to EV transfected cells (except EV cells treated with 20 nM E2). In OV transfected cells, 10 nM E2 dose proved to be equally effective as 20 nM E2 at 24 hours, while EV cells showed higher cell viability in response to the higher E2 concentration. Three way ANOVA resulted that treatment alone ($F(2,48)= 6.622$, $p= 3.00e-03$), transfection alone alone ($F(1,48)= 27.209$, $p= 3.84e-06$), cell number alone $F(1,48)= 11.374$, $p= 1.00e-03$), treatment:transfection alone ($F(2,48)= 4.331$, $p= 1.90e-$

02) and transfection:cell number alone (F(1,48)= 5.438, p= 2.40e-02) were significant effects.

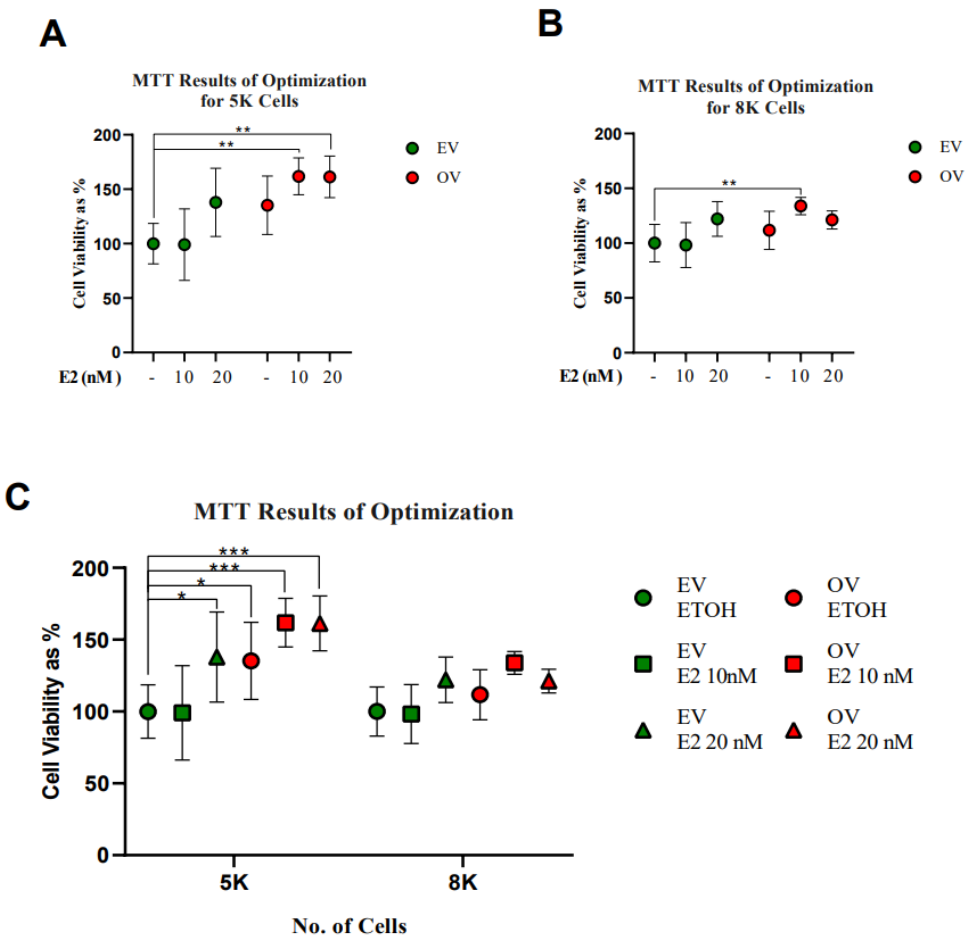


Figure 81. MTT optimization results. Analysis of the effect of transfection (EV or OV) and E2 (10 nM and 20 nM) in MCF-7 cells seeded with 5K and 8K cell numbers. Dunnett's multiple comparisons test was performed. (*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$, ****:$p<0.0001$).

Effects of different doses of Drsp on cell viability

In order to determine the dose-dependent effect of Drsp, the MTT assays were performed for MCF-7 cells transfected with MR or empty vector. Drsp treatments were given to both MR-overexpressing and naive ER+ MCF7 cell lines, and the effects of different doses on cell viability and/or growth were tested after 24 hours of treatment. In this context, 10μM, 1μM, 0.1μM and 0.01μM Drospirenone were given for 24 hours in two different 96-well plates, and cell growth was measured in cells transfected with MR or control vector. Upon analysis of the results from Plate 1 and Plate 2 using by One-Way ANOVA and Dunnett's multiple comparison tests, no statistically significant findings were obtained in both experimental setups compared to ethanol control or in multiple comparisons between doses (Figure 82). Subsequently, the two experimental setups were tested for statistical differences between them, and the two-way ANOVA results showed that there was a significant difference between the two plates ($p = 0.0030$). The drug-treated groups account for 22.30% of the total variation with a p-value approaching significance ($p=0.0641$). On the other hand, two-way ANOVA analyses for OV- and EV-transfected cells treated with Drospirenone doses (0.01, 0.1, 1, and 10 μM) showed that there was also a significant difference between the two plates transfected with EV ($p = 0.0311$). Two-way ANOVA results for MCF-7 cells transfected with OV showed that plate and dose accounted for 17.01% and 25.33% of the total variation, respectively ($p = 0.0109$ and $p = 0.0465$) (Figure 82). When the two plates were analyzed separately in OV-transfected MCF-7 cell lines, the difference between ethanol and 0.1 μM doses for Plate 2 approached significance ($p = 0.00708$). In general, a negative effect on cell viability was observed at low doses of drospirenone and a proliferative effect at increasing doses.

A) B)

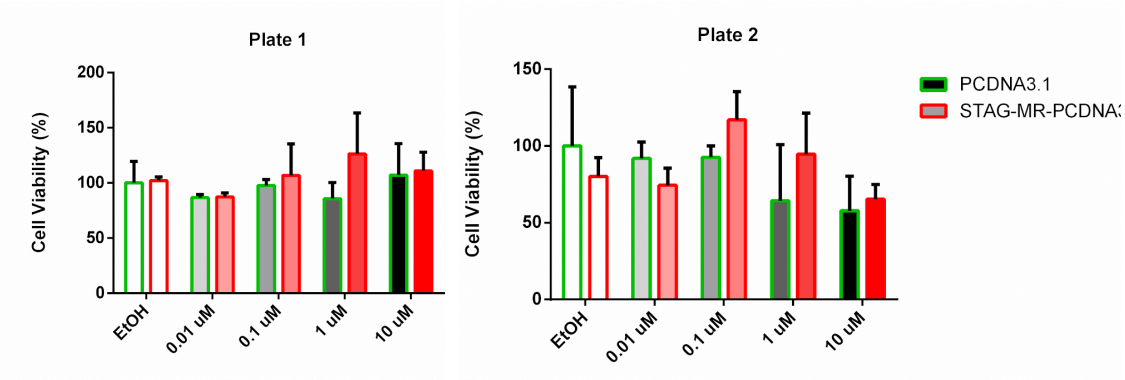


Figure 82. Comparison of cell viability according to Drospirenone concentrations on MCF-7 cell line. At 24 h incubation, cells were treated with Drospirenone (Drsp) prepared at 4 different concentrations from 0.01 µM to 10 µM. A) and B) show the results of repeating the same experimental setup in two 96-well plates with three biological replicates of each sample.

For OV- and EV-transfected cell plate treated with high doses of Drsp (12.5, 25, 50 and 100 nM), two-way ANOVA results showed that the vector accounted for 24.83% of the total variation ($p = 0.0011$). For EV-transfected MCF-7 cells, one-way ANOVA was performed and Drsp dose showed an overall significance ($p = 0.0235$, one-way ANOVA), indicating that 100 nM Drsp negatively affected cell viability in the absence of MR (EV-transfection) (Figure 83A); however, no significant change was found in OV ($p = 0.94$; Figure 83B).

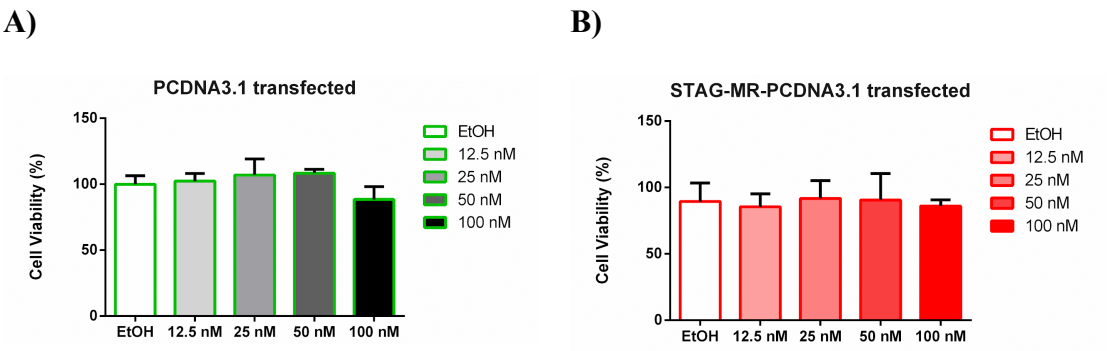


Figure 83. Comparison of cell viability according to Drospirenone concentrations on MCF-7 cell line. After 24 hours of incubation, cells were treated with Drospirenone prepared at 4 different concentrations from 12.5 nM to 100 nM. A) EV transfected cells; B) OV transfected cells.

In addition to the effects of low and high doses of Drsp alone, another MTT experiment was performed for the comparison of Drsp and the combination of Drsp_Aldo and Drsp_E2. Accordingly, OV alone increased cell survival (Figure 84). There was no significant difference between the effects of 20 nM and 100 nM Drsp doses alone. In the samples where 20 nM Drsp and 100 nM Aldo drugs were given together, a high increase in OV compared to EV was observed. However, this increase was not observed in the sample with 100 nM Drsp and 100 nM Aldo, but a decrease was observed, although not significant, compared to EV. When E2 was added to these samples, a significant increase in OV relative to EV was found in the sample with 100 nM Drsp, but not in the sample with 20 nM Drsp.

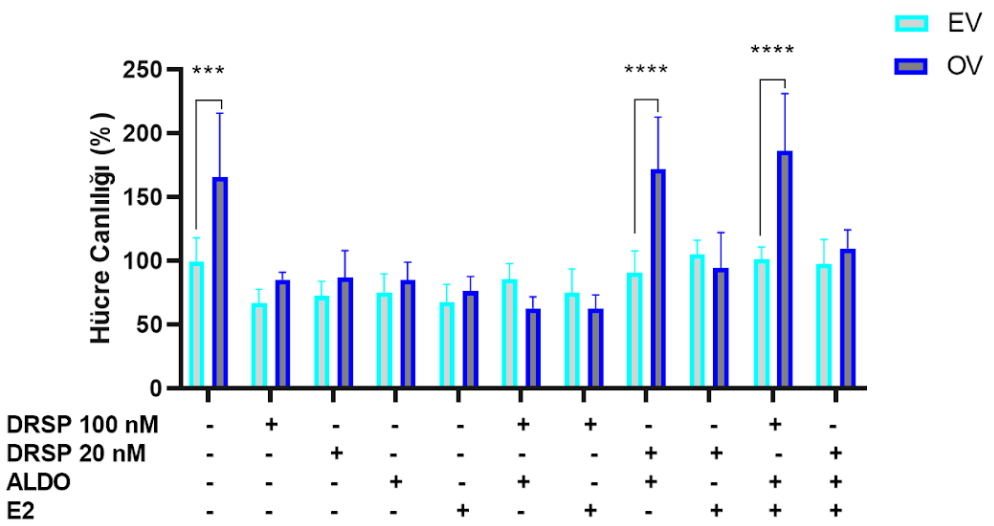


Figure 84. Effect of Drsp, Aldo, E2 and combination treatments on cell viability levels of MCF7 cells transfected with empty vector (EV) and MR expressing vector (OV) for 24 hours. Experimental results are expressed as mean ± standard deviation (n=5 for all groups). Two-way Analysis of Variance and Sidak's test for intergroup comparisons were applied for statistical analysis (*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$, ****:$p<0.0001$).

Appendix C. The Effect of High-dose Drsp and MR Overexpression in the Presence of Aldo

The studies were conducted to explore whether high doses of Drsp could counteract the effects of Aldo in both the presence and absence of MR expression. In these experiments, it was aimed to investigate the interaction of the effect of Drsp with Aldo and E2, the expression of selected MR and ESR1 targets, positive control, cell cycle, and apoptosis genes in addition to MR increase. In all qPCR experiments, TPT1 was used as a reference gene.

First, the effect of OV and EV was wanted to be tested, and preliminary studies were carried out by selecting medium and high-dose applications of Drsp. OV (MR-STAG) or EV transfected cells were treated with Drsp doses of 400 nM and 80 nM, 100 nM of Aldo, and 10 nM of E2 in different combinations as two biological replicates, but a sufficient amount of pellets or RNA could not be obtained from some replicates. The results of the qRT-PCR experiment performed by extracting cDNA from replicates from which sufficient amounts of RNA were obtained in RNA isolation. In the analyses, the EV-Aldo sample was used as a reference for normalization, and the effects of OV alone and in combination with E2 or Aldo or their combinations on gene expression were analyzed.

In each group with E2 treatment, there was at least a 4-fold increase in the mRNA level of PGR, which is the primary target of ESR1, whereas no such increase was observed in the groups without E2 (Figure 85A). In addition, MR expression was at least 10-fold higher in each MR-Stag vector (OV) treated group (Figure 85B), which indicates that MR-OV treatment worked successfully.

In this context, the addition of Aldo in the presence of E2 in high Drsp (D1) treatment may cause a synergistic increase in E2-induced PGR expression (Figure 85A). Similarly, the same is true for moderate-dose Drsp (D2) treatment, where E2 with moderate-dose Drsp can synergize more effectively without Aldo. On the other hand, the interaction of Aldo with D2 in the presence of E2 can antagonistically affect MR expression (decrease in MR amount; Figure 85B), but this did not occur at moderate dose Drsp (D2). The effect of Aldosterone on MR could not be observed without OV.

One of the most interesting findings was observed in ZBTB16 gene expression (Figure 85C). ZBTB16 is a target of both MR and GR. Since I used 100 nM Aldo, this amount is expected to have an effect on both GR and MR. The expression of ZBTB16 was slightly increased in OV-Aldo and EV-Aldo_E2 samples, indicating that ZBTB16 can interact with ESR1, both primary and secondary. However, in the presence of MR (OV), co-administration of both Drsp (D1 and D2, separately) doses with E2 decreased ZBTB16 expression, and this decrease was normalized with the addition of Aldo. This finding shows for the first time in the literature that E2 treatment may have an antagonistic effect inversely proportional to MR expression and that MR can be neutralized by 100 nM Aldo. Uncovering the ZBTB16 gene interaction network has become an important element in unraveling the interaction between Aldo-E2.

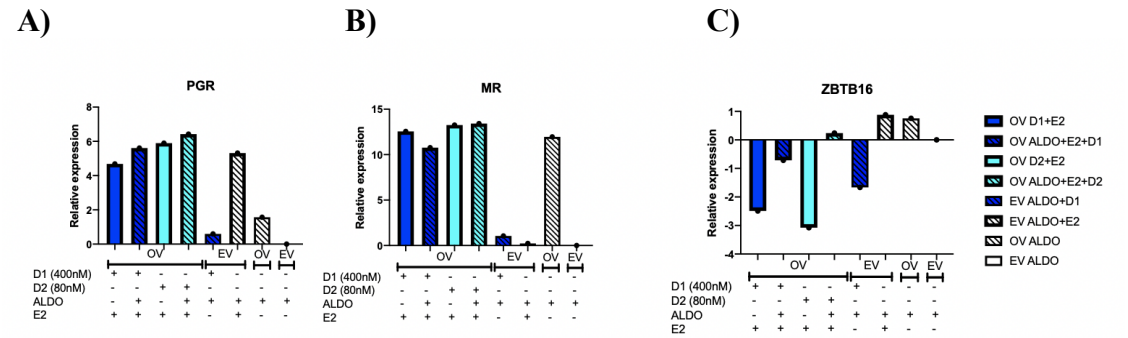


Figure 85. qRT-PCR results normalized to EV Aldo (D1:400 nM, D2:80 nM, Aldo:100 nM and E2: 10 nM). mRNA expression of A) PGR B) MR C) ZBTB16 genes.

For the analysis of proliferation-related genes, CCDN1, CCNE2, and RRM2 gene expressions were tested on this set (Figure 86). Transcriptional activation of the CCND1 gene has an important role in G(1)-phase progression and it is a direct target gene of E2 signaling (Stanelle, 2002). CCNE2 gene is a secondary estrogen target required for cell cycle control in the late G1 and early S phase (Fagundes & Teixeira, 2021). Accordingly, medium-dose Drsp (D2) in the presence or absence of Aldo caused a greater increase in proliferation markers. Moreover, even in the presence of EV treatment, these three proliferation genes were shown to increase with E2.

The other genes investigated in this set were CHK1, BAX, and BCL2. The CHEK1 gene induces DNA damage, and a decrease in CHK1 expression can induce apoptosis (Prince et al., 2016). When Aldo was added to the presence of E2 in MR-OV cells, it was discovered that CHK1 expression rose, indicating that Aldo may have an important role in preventing DNA damage. The effect of E2 on CHK1 was more apparent in medium-

dose Drsp (D2) compared to high-dose Drsp (D1). Aldo alone showed a greater increase in MR transfected cells, and E2 treatment caused an increase in CHK1, independent of MR, in EV Aldo-treated cells. For BCL2 which is an antiapoptotic gene, no statistically significant difference was found, although a decrease was observed in samples with Aldo. As observed in CHK1, medium-dose Drsp (D2) was more effective in the BAX gene.

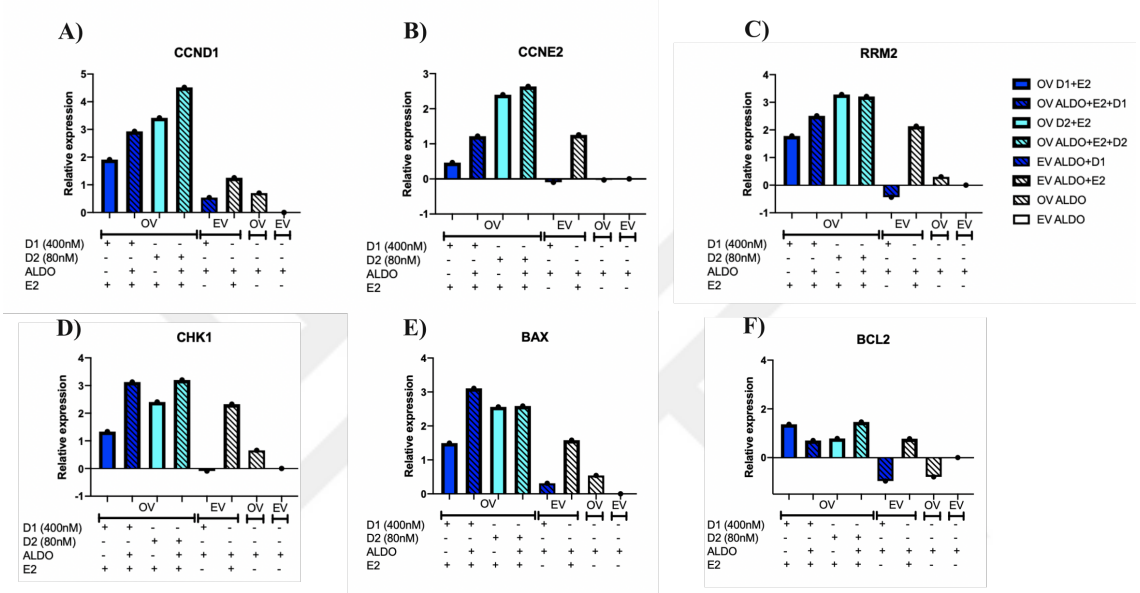


Figure 86. qRT-PCR results in MCF-7 cells using cell cycle genes normalized to EV Aldo (D1: 400 nM, D2: 80 nM, Aldo: 100 nM and E2: 10 nM). Expression of A) CCND1, B) CCNE2, C) RRM2, D) CHK1, E) BAX and F) BCL2 genes.

Appendix D. Investigation of the Effect of Drsp Doses in T47D Cells

RNAs extracted from T47D cells treated with two different doses of Drsp were analyzed for MR-targeted NEDD4L, HSD11B2 genes, and ESR1-targeted TFF1 as well as CLDN1 genes, but no significant difference was observed, although a downward trend was observed in the CLDN1 gene at a 40 nM Drsp dose (Figure 87). The changes in mRNA expression of CCND1 and GATA3 genes were also examined with 20 nM and 40 nM Drsp doses, but no significant difference was observed. This indicated that the 20 nM dose could be used in further analysis.

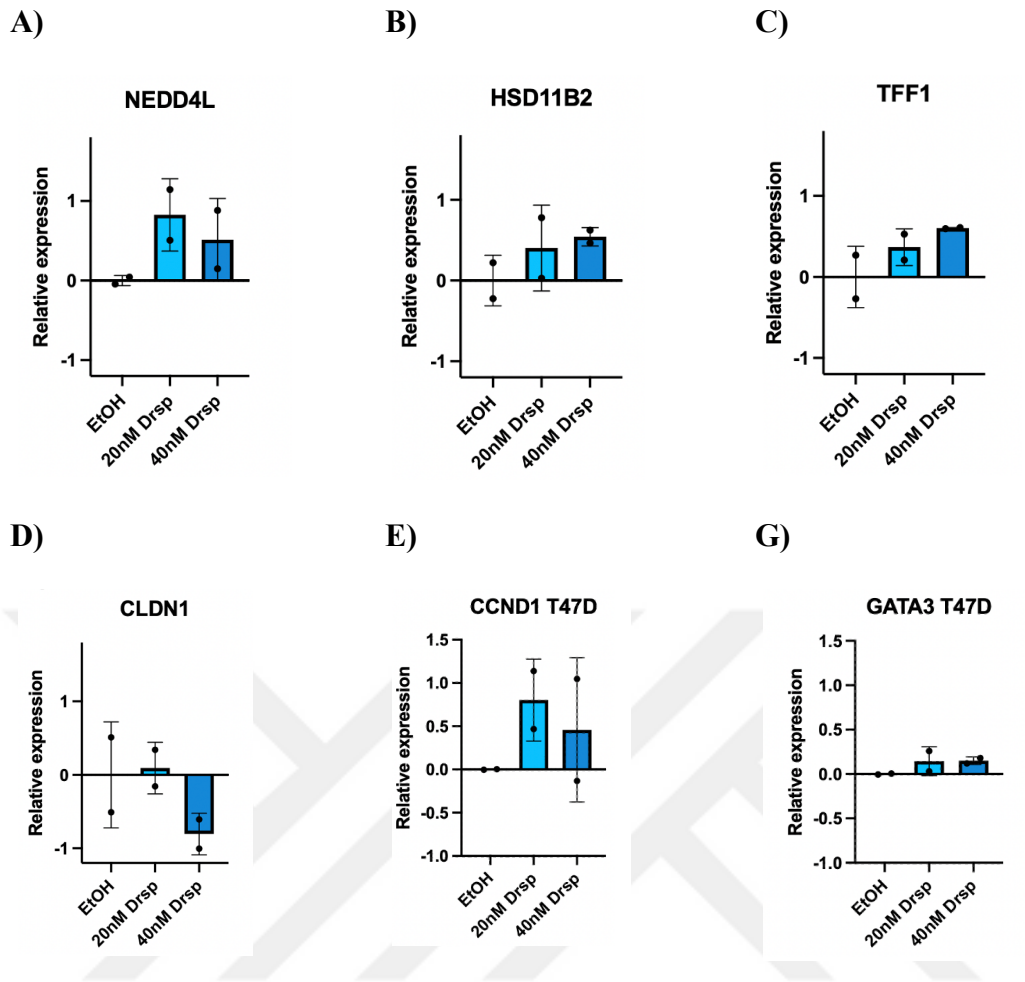


Figure 87. qRT-PCR results with RNAs from T47D cells (*: p<0.05, **: p<0.01, ***: p<0.001, ****:p<0.0001). One-way ANOVA results for genes are as follows: F(2,3)=2.192 p=0.259 for NEDD4L; F(2,3)=1.211 p=0.412 for HSD11B2; F(2,3)=2.847 p=0.203 for TFF1; F(2,3)=2.023 p=0.278 for CLDN1; F(2,3)=1.056, p=0.45 for CCND1 and F(2,3)=1.539 p=0.347 for GATA3.

Appendix E. Chromatin Immunoprecipitation Sequencing (ChIP-Seq)

Chromatin immunoprecipitation (ChIP) experiments were performed for control, E2 and Drsp_E2 combination samples as one replicate of each to investigate the binding dynamics of genomic elements under these hormonal conditions.

Methods for ChIP

For each sample, 20 million T47D cells were seeded in 3 different 10 cm dishes. Since T47D cells can endogenously produce MR, they were grown for 24 hours without transfection in RPMI containing 5% Charcoal stripped FBS and no phenol red, then treated with ethanol only (for control), 10 nM E2 or 20 nM Drsp with 10 nM E2, and chromatin was fixed after 24 hours to preserve the protein-DNA interactions at the time of treatment.

Crosslink

For the crosslink, 10x formaldehyde solution was prepared, compromising 1.46 ml Paro Fix (containing 50 mM Hepes pH 8.0 (NaOH), 0.5 mM EGTA, 1 mM EDTA and 100 mM NaCl) and 0.54 ml formaldehyde. From this solution, 1 ml was added to cell culture medium through the 10 cm dish plate's wall, achieving a final concentration of 1x. Then, the plates were shaken at low speed on a rotator at room temperature for 15 minutes. Glycine (550 ul of 2.5 M) was then given to the chromatin to stop the crosslinking reaction and avoid over-fixation at a final concentration of 0.125 M, and the plates were kept at 4°C for 10 minutes, which allowed the quenching reaction to proceed effectively.

Preparation of Chromatin

Cells were rinsed with 8 ml ice-cold PBS to remove any residual medium. Then, they were detached from the culture dish using a cell scraper in 1 ml PBS and collected in 15 ml falcon tubes on ice. These cell suspensions in 15 ml falcon tubes were centrifuged at 600 g 4°C for 5 minutes, followed by the removal of supernatant. Each pellet was resuspended in 12 ml Paro Rinse 1 (prepared with 10 mM Tris pH 8.0, 10 mM EDTA pH 8.0, 0.5 mM EGTA, 0.25% Triton X-100 and containing Protease Inhibitor Cocktail (PIC)) and incubated on ice for 5 minutes. The cells were then centrifuged again at 600 g 4°C for 5 minutes and resuspended in 12 ml Paro Rinse 2 (prepared with 10 mM Tris pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 200 mM NaCl and containing PIC) and incubated on ice for 5 minutes. After a final centrifugation at 600 g for 5 minutes at 4°C and removal of supernatant, the pellets were resuspended in 1 ml Lysis buffer (50 mM HEPES/KOH, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% DOC, 0.1% SDS and 411.5ml dH2O) containing PIC (40 µl) and incubated on ice for at least 10 minutes. At this stage, the samples were kept at -80 for a while and transported to Izmir for the rest of the procedure.

Before sonication, 50ul of chromatin was first taken from each sample as an "uncut" chromatin control and stored for control use. The remaining samples were sonicated using a Covaris sonicator to shear the DNA into fragments. After the sonication, 50 ul of sonicated chromatin was separated and stored as "input" chromatin. Reversal of crosslink was performed for both uncut and input chromatin samples.

Reversal of Crosslink for Input Fraction

After the sonication and preparation of input samples, a set of procedures was carried out for the reversal of crosslink and DNA purification. The volume of input samples was increased to 200 ul volume by adding TE. Then, 4 ul of RNase A from 10 mg/ml stock was added the samples to make final concentration of 0.2 mg/ml. Samples were incubated for 30 min at 37°C. Following the RNase A treatment, 20 ul of 10% SDS, 4 ul of 5M NaCl and 2 ul of 20 mg/ml Proteinase K were added to samples, with the final concentrations of 1% SDS, 100 mM NaCl, 200 ug/ml Proteinase K. This mixture was incubated at 55°C for at least 2.5 hours, followed by an overnight incubation at 65°C. Next day, samples (except uncut) were centrifuged at maximum speed at room temperature for 10 minutes, and the supernatant was taken to another 1.5 ml eppendorf.

DNA from samples was eluted with Zymo DNA Clean&Concentrator kit, following the manufacturer's instructions. After the DNA purification, the uncut and input chromatin samples were run on an agarose gel in order to confirm the efficiency of the sonication procedure and check that the chromatin size was between 200-500bp fragments. The next stages were then only carried out when appropriate fragments were observed.

Pre-blocking of Magnetic Beads

For pre-blocking of the magnetic beads, 60 µl Dynabeads was added to a 1.5 ml eppendorf tube, followed by 1 ml TE pH 8.0 and mixed by inverting. The tubes were placed on a magnetic stand, and the beads were allowed to be collected towards the stand while the remaining solution was allowed to become clear. The clear solution was removed from the tube without disturbing the beads collected at the side ıf the tube. This washing step

was repeated. After the second wash, beads were resuspended in 100 ul TE buffer, 10ul of 10mg/ml BSA, 10ul of 10mg/ml tRNA (denaturated for 5 min at 95°C before used). The tubes were incubated at 4°C for 2 hours with overhead shaking. Following the incubation, beads were washed 3 times with 1 ml of TE as in the beginning to remove any unbound blocking agents, and finally pre-blocked beads were dissolved in 60ul of TE and kept at 4°C.

Immunoprecipitation

Immunoprecipitation was started after pre-blocking of the magnetic beads. After adding 10ul of tRNA (10mg/ml) blocked magnetic beads per 70 ug of not pre-cleared chromatin, samples were shaken at 4°C for 1 hour. Then, the tubes were placed on a magnetic stand to collect the supernatants for ChIP. After adding 5 µl of antibody (ESR1) to each tube, the samples were incubated overnight at 4°C with upside down shaking.

The next day, 50 µl of pre-blocked magnetic beads were added to each sample and incubated at 4°C for 3 hours with upside down shaking. The tubes were then placed on a magnetic stand, and the supernatants were discarded. The remaining beads were redissolved in 1 ml of Lysis buffer containing PIC and shaken for 5 minutes, then collected on the magnetic stand and supernatants were discarded. After repeating this process once more, the beads were washed separately with 1ml DOC buffer (20 ul of 1M Tris pH 8.0, 0.5 ml of 1M LiCl, 10 ul NP-40, 0.1 ml of 10% DOC, 4 ul of 0.5M EDTA and 1366 ul dH2O) containing PIC and 1ml TE pH 8.0 containing PIC, collected on a magnetic stand. After the removal of supernatant, pellets were spinned for 3 min at 960 g at 4°C and all remaining supernatants (residual TE) were removed with a gel loading tip. The beads were redissolved in 100 µl freshly made elution buffer (50 ul of 10% SDS, 50

ul of 1M NaHCO3 and 400 ul dH2O) and incubated in a shaker incubator (1000 rpm) at 25°C for 15 min, vortexing every few minutes. They were spinned at 11000 rpm for 2 minutes at room temperature and 100 µl of supernatant containing the eluted DNA was carefully transferred to a new tube. This step was repeated with 100 ul elution buffer to increase the yield, pooling the elutes to a final volume of 200 µl.

Reversal of Crosslink for IP Fraction

To each 200 ul of sample volume, 4 ul RNase A (10 mg/ml stock) was added to achieve a final concentration of 0.2 mg/ml RNase A, and this mixture was incubated at 37°C for 30 minutes. Then, 4 µl of 0.5 M EDTA, 8 µl of Tris pH 6.5 1M, 0.5 µl of Proteinase K (20 mg/ml stock) were added per 200 ul volume to achieve a final concentrations of 10 mM EDTA, 40 mM Tris pH 6.5, 50 µg/ml Proteinase K, and incubated at 55°C for at least 2.5 hours, then overnight at 65°C. The next day, the samples were centrifuged at room temperature for 10 minutes at maximum speed, then the supernatants were taken into another 1.5 ml eppendorf tubes, and ChIP DNA was isolated with Zymo DNA Clean & Concentrator.

Prior to sequencing the DNA extracted from the samples, qPCRs using SYBR green were run as a preliminary test in order to check the accuracy of the ChIP process. As a positive control, the TPRG1 primer was designed based on the ChIP results in the literature., Moreover, GREB1 and PGR genes were also tested in order to detect an increase in the presence of E2. The primer regions designed for these genes were selected by uploading high quality T47D ChIP-seq data to the UCSC genome browser using the Cistrome database (Mei et al., 2017; Zheng et al., 2019), focusing on regions with consistent

coverage across samples. Primers were designed by selecting regions with similar coverage at the same location in all uploaded data.

Sequencing and Analysis of ChIP Samples

After the qPCR validation of ChIP samples, they were sent to Macrogen for sequencing. Upon receiving the samples, Macrogen performed DNA extraction and quality control (QC). Macrogen asked for our approval to proceed with library construction with qualified samples in the Figure 88A, we confirmed the use of all samples. The samples for which library preparations were made using the TruSeq DNA Sample prep Kit were sequenced as pair-end. In Figure 88B, it was observed that the sequenced samples were compatible with each other, and bioinformatics analysis was started.

A)

#	Sample Name	Conc. (ng/ul)	Final Volume (ul)	Total Amount (ug)	Result*	
1	Contr_Inp_DNA	214.253	13	2.785	Hold	Need to confirm
2	Contr_Chip_DNA	0.982	9	0.009	Fail	Quantity, Need to confirm
3	E2_Chip_DNA	1.176	16	0.019	Hold	Need to confirm
4	DE2_Chip_DNA	1.231	13	0.016	Hold	Need to confirm

B)

Sample ID	Total read bases (bp)	Total reads	GC(%)	AT(%)	Q20(%)	Q30
Contr_Chip_DNA	10,115,441,378	66,989,678	43.82	56.18	94.65	87.
Contr_Inp_DNA	10,669,780,498	70,660,798	41.05	58.95	95.53	89.
DE2_Chip_DNA	11,072,169,224	73,325,624	43.83	56.17	95.09	88.
E2_Chip_DNA	11,394,229,876	75,458,476	44.32	55.68	95.57	89.

- Sample ID : Sample name.
- Total read bases : Total number of bases sequenced.
- Total reads : Total number of reads. For Illumina paired-end sequencing, this value refers to the sur read 1 and read 2.
- GC(%) : GC content.
- AT(%) : AT content.
- Q20(%) : Ratio of bases that have phred quality score of over 20.
- Q30(%) : Ratio of bases that have phred quality score of over 30.

Figure 88. Concentration information before library preparation (A) and reads of raw data (B).

For the analysis of ChIP-seq data, Galaxy platform was used. After uploading the samples into the platform, alignment of the reads to the reference genome hg38 was performed by using BWA (Burrows-Wheeler Aligner) (H. Li & Durbin, 2009b, 2010). Upon aligning the data, MACS2 peak calling tool was run to find both wide and narrow peaks (Feng et al., 2012; Y. Zhang et al., 2008). To generate coverage data, Wig/BedGraph-to-bigWig tool of Galaxy was used to obtained bigwig files, which are frequently used for viewing and analyzing coverage data, from bedgraph files provided by MACS2 peak calling.

Appendix F. Analysis of GSE74620 dataset

After uploading the raw count data of GSE74620 to CAP-RNAseq, vst+ANOVA was performed with a threshold of p value <0.00005. Remaining genes were clustered for 12 clusters with default parameters and MSigDB analysis was performed for Hallmark gene set collection using only essential genes (Figure 89).

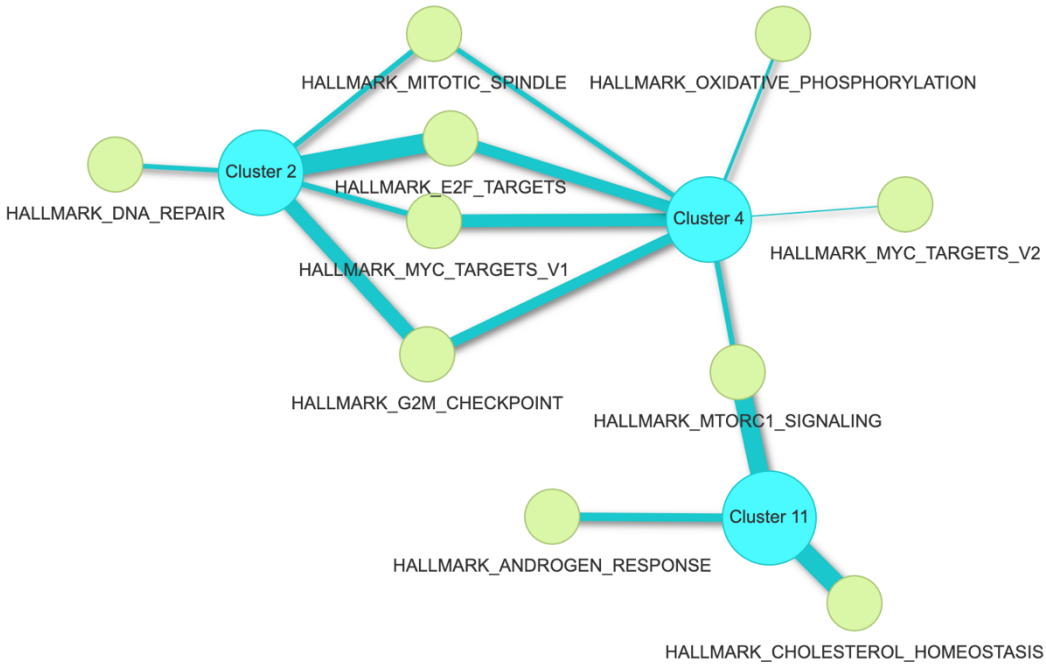


Figure 89. Hallmark enrichment of essential genes for cluster 2, cluster 4 and cluster 11 when comparing CDK4 inhibitor versus Serum Starvation groups.

REFERENCES

Aarts, M. T., Wagner, M., Van Der Wal, T., Van Boxtel, A. L., & Van Amerongen, R. (2023). A molecular toolbox to study progesterone receptor signaling. *Journal of Mammary Gland Biology and Neoplasia*, *28*(1), 24. https://doi.org/10.1007/s10911-023-09550-0

Abd El-Maqsoud, N. M. R., & Abd El-Rehim, D. M. (2014). Clinicopathologic Implications of EpCAM and Sox2 Expression in Breast Cancer. *Clinical Breast Cancer*, *14*(1), e1–e9. https://doi.org/10.1016/j.clbc.2013.09.006

Acosta, J. C., O'Loghlen, A., Banito, A., Guijarro, M. V., Augert, A., Raguz, S., Fumagalli, M., Da Costa, M., Brown, C., Popov, N., Takatsu, Y., Melamed, J., d'Adda Di Fagagna, F., Bernard, D., Hernando, E., & Gil, J. (2008). Chemokine Signaling via the CXCR2 Receptor Reinforces Senescence. *Cell*, *133*(6), 1006–1018. https://doi.org/10.1016/j.cell.2008.03.038

Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., Kerlavage, A. R., McCombie, W. R., & Venter, J. C. (1991). Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project. *Science*, *252*(5013), 1651–1656. https://doi.org/10.1126/science.2047873

Adie, E. A., Adams, R. R., Evans, K. L., Porteous, D. J., & Pickard, B. S. (2005). Speeding disease gene discovery by sequence based candidate prioritization. *BMC Bioinformatics*, *6*(1), 55. https://doi.org/10.1186/1471-2105-6-55

Adie, E. A., Adams, R. R., Evans, K. L., Porteous, D. J., & Pickard, B. S. (2006). SUSPECTS: Enabling fast and effective prioritization of positional candidates. *Bioinformatics*, *22*(6), 773–774. https://doi.org/10.1093/bioinformatics/btk031

Aerts, S., Lambrechts, D., Maity, S., Van Loo, P., Coessens, B., De Smet, F., Tranchevent, L.-C., De Moor, B., Marynen, P., Hassan, B., Carmeliet, P., & Moreau, Y. (2006). Gene prioritization through genomic data fusion. *Nature Biotechnology*, *24*(5), 537–544. https://doi.org/10.1038/nbt1203

Agarwal, M. K., Mirshahi, F., Mirshahi, M., & Rostene, W. (1993). Immunochemical Detection of the Mineralocorticoid Receptor in Rat Brain. *Neuroendocrinology*, *58*(5), 575–580. https://doi.org/10.1159/000126593

Ahmed, A. H., Gordon, R. D., Ward, G., Wolley, M., McWhinney, B. C., Ungerer, J. P., & Stowasser, M. (2017). Effect of Combined Hormonal Replacement Therapy on the Aldosterone/Renin Ratio in Postmenopausal Women. *The Journal of Clinical Endocrinology & Metabolism*, *102*(7), 2329–2334. https://doi.org/10.1210/jc.2016-3851

Akhmedov, M., Martinelli, A., Geiger, R., & Kwee, I. (2020). Omics Playground: A comprehensive self-service platform for visualization, analytics and exploration of Big Omics Data. *NAR Genomics and Bioinformatics*, *2*(1), lqz019. https://doi.org/10.1093/nargab/lqz019

Alcon, C., Gómez Tejeda Zañudo, J. G. T., Albert, R., Wagle, N., Scaltriti, M., Letai, A., Samitier, J., & Montero, J. (2021). ER+ Breast Cancer Strongly Depends on MCL-1 and BCL-xL Anti-Apoptotic Proteins. *Cells*, *10*(7), 1659. https://doi.org/10.3390/cells10071659

Alexander, I. E., Clarke, C. L., Shine, J., & Sutherland, R. L. (1989). Progestin Inhibition of Progesterone Receptor Gene Expression in Human these steroids is

determined in part by the cellular Breast Cancer Cells. *Molecular Endocrinology*, *3*(9), 1377–1386. https://doi.org/10.1210/mend-3-9-1377

Alexandrova, R., Desislav Dinev, Gavrilova-Valcheva, I., & Gavrilov, I. (2019). *Cell cultures as model systems in breast cancer research*. *7*(2), 073–079. https://doi.org/10.5281/ZENODO.2579322

Alghamri, M. S., McClellan, B. L., Hartlage, C. S., Haase, S., Faisal, S. M., Thalla, R., Dabaja, A., Banerjee, K., Carney, S. V., Mujeeb, A. A., Olin, M. R., Moon, J. J., Schwendeman, A., Lowenstein, P. R., & Castro, M. G. (2021). Targeting Neuroinflammation in Brain Cancer: Uncovering Mechanisms, Pharmacological Targets, and Neuropharmaceutical Developments. *Frontiers in Pharmacology*, *12*, 680021. https://doi.org/10.3389/fphar.2021.680021

Alkarkhi, A. F. M., & Alqaraghuli, W. A. A. (2019). Cluster Analysis. In *Easy Statistics for Food Science with R* (pp. 177–186). Elsevier. https://doi.org/10.1016/B978-0-12-814262-2.00011-X

Almende B.V. and Contributors, & Thieurmel, B. (2022). *visNetwork: Network Visualization using "vis.js" Library}* (R package version 2.1.2) [Computer software]. https://CRAN.R-project.org/package=visNetwork

Al-Othman, N., Alhendi, A., Ihbaisha, M., Barahmeh, M., Alqaraleh, M., & Al-Momany, B. Z. (2020). Role of CD44 in breast cancer. *Breast Disease*, *39*(1), 1–13. https://doi.org/10.3233/BD-190409

Altar, C. A., Vawter, M. P., & Ginsberg, S. D. (2009). Target identification for CNS diseases by transcriptional profiling. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *34*(1), 18–54. https://doi.org/10.1038/npp.2008.172

Altman, N., & Krzywinski, M. (2017). Clustering. *Nature Methods*, *14*(6), 545–546. https://doi.org/10.1038/nmeth.4299

Ames, M. K., Atkins, C. E., & Pitt, B. (2019). The renin-angiotensin-aldosterone system and its suppression. *Journal of Veterinary Internal Medicine*, *33*(2), 363–382. https://doi.org/10.1111/jvim.15454

Anbalagan, M., & Rowan, B. G. (2015). Estrogen receptor alpha phosphorylation and its functional impact in human breast cancer. *Molecular and Cellular Endocrinology*, *418*, 264–272. https://doi.org/10.1016/j.mce.2015.01.016

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*(10), R106. https://doi.org/10.1186/gb-2010-11-10-r106

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxford, England)*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Anderson, L., & Seilhamer, J. (1997). A comparison of selected mRNA and protein abundances in human liver. *ELECTROPHORESIS*, *18*(3–4), 533–537. https://doi.org/10.1002/elps.1150180333

Andrews, J. E., O'Neill, M. J., Binder, M., Shioda, T., & Sinclair, A. H. (2000). Isolation and expression of a novel member of the CITED family. *Mechanisms of Development*, *95*(1–2), 305–308. https://doi.org/10.1016/S0925-4773(00)00362-2

Andrews, S. (n.d.). *FastQC A Quality Control tool for High Throughput Sequence Data.* http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Anerillas, C., Herman, A. B., Munk, R., Garrido, A., Lam, K.-W. G., Payea, M. J., Rossi, M., Tsitsipatis, D., Martindale, J. L., Piao, Y., Mazan-Mamczarz, K., Fan, J., Cui, C.-Y., De, S., Abdelmohsen, K., De Cabo, R., & Gorospe, M. (2022a). A

BDNF-TrkB autocrine loop enhances senescent cell viability. *Nature Communications*, *13*(1), 6228. https://doi.org/10.1038/s41467-022-33709-8

Anerillas, C., Herman, A. B., Munk, R., Garrido, A., Lam, K.-W. G., Payea, M. J., Rossi, M., Tsitsipatis, D., Martindale, J. L., Piao, Y., Mazan-Mamczarz, K., Fan, J., Cui, C.-Y., De, S., Abdelmohsen, K., De Cabo, R., & Gorospe, M. (2022b). Author Correction: A BDNF-TrkB autocrine loop enhances senescent cell viability. *Nature Communications*, *13*(1), 7540. https://doi.org/10.1038/s41467-022-35154-z

Angius, F., Spolitu, S., Uda, S., Deligia, S., Frau, A., Banni, S., Collu, M., Accossu, S., Madeddu, C., Serpe, R., & Batetta, B. (2015). High-density lipoprotein contribute to G0-G1/S transition in Swiss NIH/3T3 fibroblasts. *Scientific Reports*, *5*(1), 17812. https://doi.org/10.1038/srep17812

Apostolou, P., & Fostira, F. (2013). Hereditary Breast Cancer: The Era of New Susceptibility Genes. *BioMed Research International*, *2013*, 1–11. https://doi.org/10.1155/2013/747318

Arab, H., Alharbi, A. J., Oraif, A., Sagr, E., Al Madani, H., Abduljabbar, H., Bajouh, O. S., Faden, Y., & Sabr, Y. (2019). The Role Of Progestogens In Threatened And Idiopathic Recurrent Miscarriage. *International Journal of Women's Health*, *Volume 11*, 589–596. https://doi.org/10.2147/IJWH.S224159

Aragon, T. J. (2020). *epitools: Epidemiology Tools* (R package version 0.5-10.1) [Computer software]. Bioconductor. https://CRAN.R-project.org/package=epitools

Archer, D. F. (2007). Drospirenone and estradiol: A new option for the postmenopausal woman. *Climacteric*, *10*(sup1), 3–10. https://doi.org/10.1080/13697130601114859

Arnold, S. F., Vorojeikina, D. P., & Notides, A. C. (1995). Phosphorylation of Tyrosine 537 on the Human Estrogen Receptor Is Required for Binding to an Estrogen Response Element. *Journal of Biological Chemistry*, *270*(50), 30205–30212. https://doi.org/10.1074/jbc.270.50.30205

Aromolaran, O., Beder, T., Oswald, M., Oyelade, J., Adebiyi, E., & Koenig, R. (2020). Essential gene prediction in Drosophila melanogaster using machine learning approaches based on sequence and functional features. *Computational and Structural Biotechnology Journal*, *18*, 612–621. https://doi.org/10.1016/j.csbj.2020.02.022

Arriza, J. L., Simerly, R. B., Swanson, L. W., & Evans, R. M. (1988). The neuronal mineralocorticoid eeceptor as a mediator of glucocorticoid response. *Neuron*, *1*(9), 887–900. https://doi.org/10.1016/0896-6273(88)90136-5

Arriza, J. L., Weinberger, C., Cerelli, G., Glaser, T. M., Handelin, B. L., Housman, D. E., & Evans, R. M. (1987). Cloning of Human Mineralocorticoid Receptor Complementary DNA: Structural and Functional Kinship with the Glucocorticoid Receptor. *Science*, *237*(4812), 268–275. https://doi.org/10.1126/science.3037703

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Asi, N., Mohammed, K., Haydour, Q., Gionfriddo, M. R., Vargas, O. L. M., Prokop, L. J., Faubion, S. S., & Murad, M. H. (2016). Progesterone vs. synthetic progestins

and the risk of breast cancer: A systematic review and meta-analysis. *Systematic Reviews*, *5*(1), 121. https://doi.org/10.1186/s13643-016-0294-5

Attali, D. (2020). *shinydisconnect: Show a Nice Message When a "Shiny" App Disconnects or Errors* (R package version 0.1.0) [Computer software]. https://CRAN.R-project.org/package=shinydisconnect

Attali, D. (2021). *shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds* (R package version 2.1.0) [Computer software]. https://CRAN.R-project.org/package=shinyjs

Avecilla, G., Spealman, P., Matthews, J., Caudal, E., Schacherer, J., & Gresham, D. (2023). Copy number variation alters local and global mutational tolerance. *Genome Research*, *33*(8), 1340–1353. https://doi.org/10.1101/gr.277625.122

Axelrod, M. L., Wang, Y., Xu, Y., Sun, X., Bejan, C. A., Gonzalez-Ericsson, P. I., Nunnery, S., Bergman, R. E., Donaldson, J., Guerrero-Zotano, A. L., Massa, C., Seliger, B., Sanders, M., Mayer, I. A., & Balko, J. M. (2022). Peripheral Blood Monocyte Abundance Predicts Outcomes in Patients with Breast Cancer. *Cancer Research Communications*, *2*(5), 286–292. https://doi.org/10.1158/2767-9764.CRC-22-0023

Babraham Institute. (n.d.). *FastQC: a quality control tool for high throughput sequence data.* [Computer software]. Retrieved February 25, 2024, from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bailey, E. (2022). *shinyBS: Twitter Bootstrap Components for Shiny* (R package version 0.61.1) [Computer software]. https://CRAN.R-project.org/package=shinyBS

Bainbridge, M. N., Warren, R. L., Hirst, M., Romanuik, T., Zeng, T., Go, A., Delaney, A., Griffith, M., Hickenbotham, M., Magrini, V., Mardis, E. R., Sadar, M. D., Siddiqui, A. S., Marra, M. A., & Jones, S. J. (2006). Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. *BMC Genomics*, *7*(1), 246. https://doi.org/10.1186/1471-2164-7-246

Baldessari, D., Shin, Y., Krebs, O., König, R., Koide, T., Vinayagam, A., Fenger, U., Mochii, M., Terasaka, C., Kitayama, A., Peiffer, D., Ueno, N., Eils, R., Cho, K. W., & Niehrs, C. (2005). Global gene expression profiling and cluster analysis in Xenopus laevis. *Mechanisms of Development*, *122*(3), 441–475. https://doi.org/10.1016/j.mod.2004.11.007

Baron, J. A., Johnson, C. S.-B., Schor, M. A., Olley, D., Nickel, L., Felix, V., Munro, J. B., Bello, S. M., Bearer, C., Lichenstein, R., Bisordi, K., Koka, R., Greene, C., & Schriml, L. M. (2024). The DO-KB Knowledgebase: A 20-year journey developing the disease open science ecosystem. *Nucleic Acids Research*, *52*(D1), D1305–D1314. https://doi.org/10.1093/nar/gkad1051

Barrett Mueller, K., Lu, Q., Mohammad, N. N., Luu, V., McCurley, A., Williams, G. H., Adler, G. K., Karas, R. H., & Jaffe, I. Z. (2014). Estrogen Receptor Inhibits Mineralocorticoid Receptor Transcriptional Regulatory Function. *Endocrinology*, *155*(11), 4461–4472. https://doi.org/10.1210/en.2014-1270

Barros, L. A., Tufik, S., & Andersen, M. L. (2015). The role of progesterone in memory: An overview of three decades. *Neuroscience & Biobehavioral Reviews*, *49*, 193–204. https://doi.org/10.1016/j.neubiorev.2014.11.015

Bashour, N. M., & Wray, S. (2012). Progesterone Directly and Rapidly Inhibits GnRH Neuronal Activity via Progesterone Receptor Membrane Component 1. *Endocrinology*, *153*(9), 4457–4469. https://doi.org/10.1210/en.2012-1122

Basto, R., Brunk, K., Vinadogrova, T., Peel, N., Franz, A., Khodjakov, A., & Raff, J. W. (2008). Centrosome Amplification Can Initiate Tumorigenesis in Flies. *Cell*, *133*(6), 1032–1042. https://doi.org/10.1016/j.cell.2008.05.039

Baxter, G. T., Radeke, M. J., Kuo, R. C., Makrides, V., Hinkle, B., Hoang, R., Medina-Selby, A., Coit, D., Valenzuela, P., & Feinstein, S. C. (1997). Signal transduction mediated by the truncated trkB receptor isoforms, trkB.T1 and trkB.T2. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *17*(8), 2683–2690. https://doi.org/10.1523/JNEUROSCI.17-08-02683.1997

Baxter, M. E., Miller, H. A., Chen, J., Williams, B. J., & Frieboes, H. B. (2023). Metabolomic differentiation of tumor core versus edge in glioma. *Neurosurgical Focus*, *54*(6), E4. https://doi.org/10.3171/2023.3.FOCUS2379

Bedford, L., Lowe, J., Dick, L. R., Mayer, R. J., & Brownell, J. E. (2011). Ubiquitin-like protein conjugation and the ubiquitin–proteasome system as drug targets. *Nature Reviews Drug Discovery*, *10*(1), 29–46. https://doi.org/10.1038/nrd3321

Behan, F. M., Iorio, F., Picco, G., Gonçalves, E., Beaver, C. M., Migliardi, G., Santos, R., Rao, Y., Sassi, F., Pinnelli, M., Ansari, R., Harper, S., Jackson, D. A., McRae, R., Pooley, R., Wilkinson, P., Van Der Meer, D., Dow, D., Buser-Doepner, C., … Garnett, M. J. (2019). Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. *Nature*, *568*(7753), 511–516. https://doi.org/10.1038/s41586-019-1103-9

Bei, Y., Huang, Q., Shen, J., Shi, J., Shen, C., Xu, P., Chang, H., Xia, X., Xu, L., Ji, B., & Chen, J. (2017). IGFBP6 Regulates Cell Apoptosis and Migration in Glioma. *Cellular and Molecular Neurobiology*, *37*(5), 889–898. https://doi.org/10.1007/s10571-016-0426-4

Belda-Iniesta, C., De Castro Carpeño, J., Sereno, M., González-Barón, M., & Perona, R. (2008). Epidermal growth factor receptor and glioblastoma multiforme: Molecular basis for a new approach. *Clinical and Translational Oncology*, *10*(2), 73–77. https://doi.org/10.1007/s12094-008-0159-z

Benevolenskaya, E. V., Islam, A. B. M. M. K., Ahsan, H., Kibriya, M. G., Jasmine, F., Wolff, B., Al-Alem, U., Wiley, E., Kajdacsy-Balla, A., Macias, V., & Rauscher, G. H. (2016). DNA methylation and hormone receptor status in breast cancer. *Clinical Epigenetics*, *8*(1), 17. https://doi.org/10.1186/s13148-016-0184-7

Bentink, S., Haibe-Kains, B., Risch, T., Fan, J.-B., Hirsch, M. S., Holton, K., Rubio, R., April, C., Chen, J., Wickham-Garcia, E., Liu, J., Culhane, A., Drapkin, R., Quackenbush, J., & Matulonis, U. A. (2012). Angiogenic mRNA and microRNA Gene Expression Signature Predicts a Novel Subtype of Serous Ovarian Cancer. *PLoS ONE*, *7*(2), e30269. https://doi.org/10.1371/journal.pone.0030269

Beral, V. & Million Women Study Collaborators. (2003). Breast cancer and hormone-replacement therapy in the Million Women Study. *The Lancet*, *362*(9382), 419–427. https://doi.org/10.1016/S0140-6736(03)14065-2

Beral, V., Reeves, G., Bull, D., Green, J., & for the Million Women Study Collaborators. (2011). Breast Cancer Risk in Relation to the Interval Between Menopause and Starting Hormone Therapy. *JNCI Journal of the National Cancer Institute*, *103*(4), 296–305. https://doi.org/10.1093/jnci/djq527

Berger, S., Wolfer, D. P., Selbach, O., Alter, H., Erdmann, G., Reichardt, H. M., Chepkova, A. N., Welzl, H., Haas, H. L., Lipp, H.-P., & Schütz, G. (2006). Loss of the limbic mineralocorticoid receptor impairs behavioral plasticity. *Proceedings of the National Academy of Sciences*, *103*(1), 195–200. https://doi.org/10.1073/pnas.0503878102

Bermejo-Haro, M. Y., Camacho-Pacheco, R. T., Brito-Pérez, Y., & Mancilla-Herrera, I. (2023). The hormonal physiology of immune components in breast milk and

their impact on the infant immune response. *Molecular and Cellular Endocrinology*, *572*, 111956. https://doi.org/10.1016/j.mce.2023.111956

Bertucci, F., Finetti, P., Rougemont, J., Charafe-Jauffret, E., Cervera, N., Tarpin, C., Nguyen, C., Xerri, L., Houlgatte, R., Jacquemier, J., Viens, P., & Birnbaum, D. (2005). Gene Expression Profiling Identifies Molecular Subtypes of Inflammatory Breast Cancer. *Cancer Research*, *65*(6), 2170–2178. https://doi.org/10.1158/0008-5472.CAN-04-4115

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Springer US. https://doi.org/10.1007/978-1-4757-0450-1

Birch, J., & Gil, J. (2020). Senescence and the SASP: Many therapeutic avenues. *Genes & Development*, *34*(23–24), 1565–1576. https://doi.org/10.1101/gad.343129.120

Bird, S. T., Pepe, S. R., Etminan, M., Liu, X., Brophy, J. M., & Delaney, J. A. (2011). The association between drospirenone and hyperkalemia: A comparative-safety study. *BMC Clinical Pharmacology*, *11*(1), 23. https://doi.org/10.1186/1472-6904-11-23

Bjune, J.-I., Strømland, P. P., Jersin, R. Å., Mellgren, G., & Dankel, S. N. (2022). Metabolic and Epigenetic Regulation by Estrogen in Adipocytes. *Frontiers in Endocrinology*, *13*, 828780. https://doi.org/10.3389/fendo.2022.828780

Blank, E. W., Wong, P.-Y., Lakshmanaswamy, R., Guzman, R., & Nandi, S. (2008). Both ovarian hormones estrogen and progesterone are necessary for hormonal mammary carcinogenesis in ovariectomized ACI rats. *Proceedings of the National Academy of Sciences*, *105*(9), 3527–3532. https://doi.org/10.1073/pnas.0710535105

Blomen, V. A., Májek, P., Jae, L. T., Bigenzahn, J. W., Nieuwenhuis, J., Staring, J., Sacco, R., Van Diemen, F. R., Olk, N., Stukalov, A., Marceau, C., Janssen, H., Carette, J. E., Bennett, K. L., Colinge, J., Superti-Furga, G., & Brummelkamp, T. R. (2015). Gene essentiality and synthetic lethality in haploid human cells. *Science*, *350*(6264), 1092–1096. https://doi.org/10.1126/science.aac7557

Bonomi, S., Gallo, S., Catillo, M., Pignataro, D., Biamonti, G., & Ghigna, C. (2013). Oncogenic Alternative Splicing Switches: Role in Cancer Progression and Prospects for Therapy. *International Journal of Cell Biology*, *2013*, 1–17. https://doi.org/10.1155/2013/962038

Boopalan, T., Arumugam, A., Parada, J., Saltzstein, E., & Lakshmanaswamy, R. (2015). Receptor activator for nuclear factor-κB ligand signaling promotes progesterone-mediated estrogen-induced mammary carcinogenesis. *Cancer Science*, *106*(1), 25–33. https://doi.org/10.1111/cas.12571

Booth, R. E., Johnson, J. P., & Stockand, J. D. (2002). ALDOSTERONE. *Advances in Physiology Education*, *26*(1), 8–20. https://doi.org/10.1152/advan.00051.2001

Boudot, A., Kerdivel, G., Habauzit, D., Eeckhoute, J., Le Dily, F., Flouriot, G., Samson, M., & Pakdel, F. (2011). Differential Estrogen-Regulation of CXCL12 Chemokine Receptors, CXCR4 and CXCR7, Contributes to the Growth Effect of Estrogens in Breast Cancer Cells. *PLoS ONE*, *6*(6), e20898. https://doi.org/10.1371/journal.pone.0020898

Bray, J. D., Jelinsky, S., Ghatge, R., Bray, J. A., Tunkey, C., Saraf, K., Jacobsen, B. M., Richer, J. K., Brown, E. L., Winneker, R. C., Horwitz, K. B., & Lyttle, C. R. (2005). Quantitative analysis of gene regulation by seven clinically relevant progestins suggests a highly similar mechanism of action through progesterone receptors in T47D breast cancer cells. *The Journal of Steroid Biochemistry and Molecular Biology*, *97*(4), 328–341. https://doi.org/10.1016/j.jsbmb.2005.06.032

Brennan, C. W., Verhaak, R. G. W., McKenna, A., Campos, B., Noushmehr, H., Salama, S. R., Zheng, S., Chakravarty, D., Sanborn, J. Z., Berman, S. H., Beroukhim, R., Bernard, B., Wu, C.-J., Genovese, G., Shmulevich, I., Barnholtz-Sloan, J., Zou, L., Vegesna, R., Shukla, S. A., … McLendon, R. (2013). The Somatic Genomic Landscape of Glioblastoma. *Cell*, *155*(2), 462–477. https://doi.org/10.1016/j.cell.2013.09.034

Brisken, C. (2013). Progesterone signalling in breast cancer: A neglected hormone coming into the limelight. *Nature Reviews Cancer*, *13*(6), 385–396. https://doi.org/10.1038/nrc3518

Brodeur, G. M., Minturn, J. E., Ho, R., Simpson, A. M., Iyer, R., Varela, C. R., Light, J. E., Kolla, V., & Evans, A. E. (2009). Trk receptor expression and inhibition in neuroblastomas. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *15*(10), 3244–3250. https://doi.org/10.1158/1078-0432.CCR-08-1815

Brooks, J., Cairns, P., & Zeleniuch-Jacquotte, A. (2009). Promoter methylation and the detection of breast cancer. *Cancer Causes & Control*, *20*(9), 1539–1550. https://doi.org/10.1007/s10552-009-9415-y

Brown, I. J., Tzoulaki, I., Candeias, V., & Elliott, P. (2009). Salt intakes around the world: Implications for public health. *International Journal of Epidemiology*, *38*(3), 791–813. https://doi.org/10.1093/ije/dyp139

Browne, F., Wang, H., & Zheng, H. (2014). An integrative network-driven pipeline for the prioritization of Alzheimer's disease genes. *2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 385–390. https://doi.org/10.1109/BIBM.2014.6999189

Bullard, J. H., Purdom, E., Hansen, K. D., & Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics*, *11*, 94. https://doi.org/10.1186/1471-2105-11-94

Cadenas, C., Franckenstein, D., Schmidt, M., Gehrmann, M., Hermes, M., Geppert, B., Schormann, W., Maccoux, L. J., Schug, M., Schumann, A., Wilhelm, C., Freis, E., Ickstadt, K., Rahnenführer, J., Baumbach, J. I., Sickmann, A., & Hengstler, J. G. (2010). Role of thioredoxin reductase 1 and thioredoxin interacting protein in prognosis of breast cancer. *Breast Cancer Research*, *12*(3), R44. https://doi.org/10.1186/bcr2599

Cahill, M. A. (2007). Progesterone receptor membrane component 1: An integrative review. *The Journal of Steroid Biochemistry and Molecular Biology*, *105*(1–5), 16–36. https://doi.org/10.1016/j.jsbmb.2007.02.002

Cai, M., McNamara, K., Yamazaki, Y., Harada, N., Miyashita, M., Tada, H., Ishida, T., & Sasano, H. (2022). The role of mineralocorticoids and glucocorticoids under the impact of 11β-hydroxysteroid dehydrogenase in human breast lesions. *Medical Molecular Morphology*, *55*(2), 110–122. https://doi.org/10.1007/s00795-022-00312-1

Campisi, J. (2005). Senescent Cells, Tumor Suppression, and Organismal Aging: Good Citizens, Bad Neighbors. *Cell*, *120*(4), 513–522. https://doi.org/10.1016/j.cell.2005.02.003

Campisi, J. (2013). Aging, Cellular Senescence, and Cancer. *Annual Review of Physiology*, *75*(1), 685–705. https://doi.org/10.1146/annurev-physiol-030212-183653

Campisi, J., & d'Adda Di Fagagna, F. (2007). Cellular senescence: When bad things happen to good cells. *Nature Reviews Molecular Cell Biology*, *8*(9), 729–740. https://doi.org/10.1038/nrm2233

Campos, S., Davey, P., Hird, A., Pressnail, B., Bilbao, J., Aviv, R. I., Symons, S., Pirouzmand, F., Sinclair, E., Culleton, S., Desa, E., Goh, P., & Chow, E. (2009). Brain metastasis from an unknown primary, or primary brain tumour? A diagnostic dilemma. *Current Oncology (Toronto, Ont.)*, *16*(1), 62–66.

Cao, J., Liu, D., Zhao, S., Yuan, L., Huang, Y., Ma, J., Yang, Z., Shi, B., Wang, L., & Wei, J. (2020). Estrogen attenuates TGF-β1-induced EMT in intrauterine adhesion by activating Wnt/β-catenin signaling pathway. *Brazilian Journal of Medical and Biological Research*, *53*(8), e9794. https://doi.org/10.1590/1414-431x20209794

Cao, R., Chen, J., Zhang, X., Zhai, Y., Qing, X., Xing, W., Zhang, L., Malik, Y. S., Yu, H., & Zhu, X. (2014). Elevated expression of myosin X in tumours contributes to breast cancer aggressiveness and metastasis. *British Journal of Cancer*, *111*(3), 539–550. https://doi.org/10.1038/bjc.2014.298

Caprio, M., Antelmi, A., Chetrite, G., Muscat, A., Mammi, C., Marzolla, V., Fabbri, A., Zennaro, M.-C., & Fève, B. (2011). Antiadipogenic Effects of the Mineralocorticoid Receptor Antagonist Drospirenone: Potential Implications for the Treatment of Metabolic Syndrome. *Endocrinology*, *152*(1), 113–125. https://doi.org/10.1210/en.2010-0674

Carim-Todd, L., Bath, K. G., Fulgenzi, G., Yanpallewar, S., Jing, D., Barrick, C. A., Becker, J., Buckley, H., Dorsey, S. G., Lee, F. S., & Tessarollo, L. (2009). Endogenous truncated TrkB.T1 receptor regulates neuronal complexity and TrkB kinase receptor function in vivo. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(3), 678–685. https://doi.org/10.1523/JNEUROSCI.5060-08.2009

Carroll, J. S., Meyer, C. A., Song, J., Li, W., Geistlinger, T. R., Eeckhoute, J., Brodsky, A. S., Keeton, E. K., Fertuck, K. C., Hall, G. F., Wang, Q., Bekiranov, S., Sementchenko, V., Fox, E. A., Silver, P. A., Gingeras, T. R., Liu, X. S., & Brown, M. (2006). Genome-wide analysis of estrogen receptor binding sites. *Nature Genetics*, *38*(11), 1289–1297. https://doi.org/10.1038/ng1901

Caruso, J. C., & Cliff, N. (1997). Empirical Size, Coverage, and Power of Confidence Intervals for Spearman's Rho. *Educational and Psychological Measurement*, *57*(4), 637–654. https://doi.org/10.1177/0013164497057004009

Carvajal, A., Espinoza, N., Kato, S., Pinto, M., Sadarangani, A., Monso, C., Aranda, E., Villalon, M., Richer, J. K., Horwitz, K. B., Brosens, J. J., & Owen, G. I. (2005). Progesterone pre-treatment potentiates EGF pathway signaling in the breast cancer cell line ZR-75. *Breast Cancer Research and Treatment*, *94*(2), 171–183. https://doi.org/10.1007/s10549-005-7726-6

Castelli, W. P. (1988). Cardiovascular disease in women. *American Journal of Obstetrics and Gynecology*, *158*(6), 1553–1560. https://doi.org/10.1016/0002-9378(88)90189-5

Castrén, M., Patchev, V. K., Almeida, O. F., Holsboer, F., Trapp, T., & Castrén, E. (1995). Regulation of rat mineralocorticoid receptor expression in neurons by progesterone. *Endocrinology*, *136*(9), 3800–3806. https://doi.org/10.1210/endo.136.9.7649087

Catena, C., Colussi, G. L., Brosolo, G., Bertin, N., Novello, M., Palomba, A., & Sechi, L. A. (2017). Salt, Aldosterone, and Parathyroid Hormone: What Is the

Relevance for Organ Damage? *International Journal of Endocrinology*, *2017*, 1–8. https://doi.org/10.1155/2017/4397028

Ceccarelli, M., Barthel, F. P., Malta, T. M., Sabedot, T. S., Salama, S. R., Murray, B. A., Morozova, O., Newton, Y., Radenbaugh, A., Pagnotta, S. M., Anjum, S., Wang, J., Manyam, G., Zoppoli, P., Ling, S., Rao, A. A., Grifford, M., Cherniack, A. D., Zhang, H., … Zmuda, E. (2016). Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell*, *164*(3), 550–563. https://doi.org/10.1016/j.cell.2015.12.028

Centers for Disease Control and Prevention. (n.d.). *What Are the Risk Factors for Breast Cancer?* Retrieved February 11, 2024, from https://www.cdc.gov/cancer/breast/basic_info/risk_factors.htm

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C., & Schultz, N. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discovery*, *2*(5), 401–404. https://doi.org/10.1158/2159-8290.CD-12-0095

Chalhoub, N., Zhu, G., Zhu, X., & Baker, S. J. (2009). Cell type specificity of PI3K signaling in Pdk1- and Pten-deficient brains. *Genes & Development*, *23*(14), 1619–1624. https://doi.org/10.1101/gad.1799609

Chang, H. Y., Thomson, J. A., & Chen, X. (2013). Microarray Analysis of Stem Cells and Differentiation. In *Handbook of Stem Cells* (pp. 399–407). Elsevier. https://doi.org/10.1016/B978-0-12-385942-6.00034-2

Chang, W. (2021). *shinythemes: Themes for Shiny* (R package version 1.2.0}) [Computer software]. https://CRAN.R-project.org/package=shinythemes

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2023). *shiny: Web Application Framework for R* (R package version 1.7.4.1) [Computer software]. https://CRAN.R-project.org/package=shiny

Chao, M. V. (2003). Neurotrophins and their receptors: A convergence point for many signalling pathways. *Nature Reviews. Neuroscience*, *4*(4), 299–309. https://doi.org/10.1038/nrn1078

Chao, M. V., Rajagopal, R., & Lee, F. S. (2006). Neurotrophin signalling in health and disease. *Clinical Science (London, England: 1979)*, *110*(2), 167–173. https://doi.org/10.1042/CS20050163

Chapman, N. M., Boothby, M. R., & Chi, H. (2020). Metabolic coordination of T cell quiescence and activation. *Nature Reviews. Immunology*, *20*(1), 55–70. https://doi.org/10.1038/s41577-019-0203-y

Chaturvedi, V., Qin, J.-Z., Denning, M. F., Choubey, D., Diaz, M. O., & Nickoloff, B. J. (1999). Apoptosis in Proliferating, Senescent, and Immortalized Keratinocytes. *Journal of Biological Chemistry*, *274*(33), 23358–23367. https://doi.org/10.1074/jbc.274.33.23358

Chaudhary, S., Krishna, B. M., & Mishra, S. K. (2017). A novel FOXA1/ESR1 interacting pathway: A study of Oncomine™ breast cancer microarrays. *Oncology Letters*, *14*(2), 1247–1264. https://doi.org/10.3892/ol.2017.6329

Chauvin, S., Cohen-Tannoudji, J., & Guigon, C. J. (2022). Estradiol Signaling at the Heart of Folliculogenesis: Its Potential Deregulation in Human Ovarian Pathologies. *International Journal of Molecular Sciences*, *23*(1), 512. https://doi.org/10.3390/ijms23010512

Cheang, M. C. U., Martin, M., Nielsen, T. O., Prat, A., Voduc, D., Rodriguez-Lescure, A., Ruiz, A., Chia, S., Shepherd, L., Ruiz-Borrego, M., Calvo, L., Alba, E.,

Carrasco, E., Caballero, R., Tu, D., Pritchard, K. I., Levine, M. N., Bramwell, V. H., Parker, J., … Carey, L. A. (2015). Defining Breast Cancer Intrinsic Subtypes by Quantitative Receptor Expression. *The Oncologist*, *20*(5), 474–482. https://doi.org/10.1634/theoncologist.2014-0372

Chen, G., Gharib, T. G., Huang, C.-C., Taylor, J. M. G., Misek, D. E., Kardia, S. L. R., Giordano, T. J., Iannettoni, M. D., Orringer, M. B., Hanash, S. M., & Beer, D. G. (2002). Discordant Protein and mRNA Expression in Lung Adenocarcinomas. *Molecular & Cellular Proteomics*, *1*(4), 304–313. https://doi.org/10.1074/mcp.M200008-MCP200

Chen, G., Gharib, T. G., Huang, C.-C., Thomas, D. G., Shedden, K. A., Taylor, J. M. G., Kardia, S. L. R., Misek, D. E., Giordano, T. J., Iannettoni, M. D., Orringer, M. B., Hanash, S. M., & Beer, D. G. (2002). Proteomic analysis of lung adenocarcinoma: Identification of a highly expressed set of proteins in tumors. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *8*(7), 2298–2305.

Chen, J., Bardes, E. E., Aronow, B. J., & Jegga, A. G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Research*, *37*(Web Server), W305–W311. https://doi.org/10.1093/nar/gkp427

Chen, J., Liu, X., Huang, H., Zhang, F., Lu, Y., & Hu, H. (2020). High salt diet may promote progression of breast tumor through eliciting immune response. *International Immunopharmacology*, *87*, 106816. https://doi.org/10.1016/j.intimp.2020.106816

Chen, J., Xu, H., Aronow, B. J., & Jegga, A. G. (2007). Improved human disease candidate gene prioritization using mouse phenotype. *BMC Bioinformatics*, *8*(1), 392. https://doi.org/10.1186/1471-2105-8-392

Chen, P., Li, B., & Ou-Yang, L. (2022). Role of estrogen receptors in health and disease. *Frontiers in Endocrinology*, *13*, 839005. https://doi.org/10.3389/fendo.2022.839005

Chen, Z., Chen, S., & Qiang, X. (2022). Identification of Biomarker in Brain-specific Gene Regulatory Network Using Structural Controllability Analysis. *Frontiers in Bioinformatics*, *2*, 812314. https://doi.org/10.3389/fbinf.2022.812314

Chimge, N.-O., & Frenkel, B. (2013). The RUNX family in breast cancer: Relationships with estrogen signaling. *Oncogene*, *32*(17), 2121–2130. https://doi.org/10.1038/onc.2012.328

Chlebowski, R. T., Kuller, L. H., Prentice, R. L., Stefanick, M. L., Manson, J. E., Gass, M., Aragaki, A. K., Ockene, J. K., Lane, D. S., Sarto, G. E., Rajkovic, A., Schenken, R., Hendrix, S. L., Ravdin, P. M., Rohan, T. E., Yasmeen, S., & Anderson, G. (2009). Breast Cancer after Use of Estrogen plus Progestin in Postmenopausal Women. *New England Journal of Medicine*, *360*(6), 573–587. https://doi.org/10.1056/NEJMoa0807684

Chlebowski, R. T., Manson, J. E., Anderson, G. L., Cauley, J. A., Aragaki, A. K., Stefanick, M. L., Lane, D. S., Johnson, K. C., Wactawski-Wende, J., Chen, C., Qi, L., Yasmeen, S., Newcomb, P. A., & Prentice, R. L. (2013). Estrogen Plus Progestin and Breast Cancer Incidence and Mortality in the Women's Health Initiative Observational Study. *JNCI: Journal of the National Cancer Institute*, *105*(8), 526–535. https://doi.org/10.1093/jnci/djt043

Choi, Y. (2022). Estrogen Receptor β Expression and Its Clinical Implication in Breast Cancers: Favorable or Unfavorable? *Journal of Breast Cancer*, *25*(2), 75. https://doi.org/10.4048/jbc.2022.25.e9

Choi, Y., Kim, H., & Pollack, S. (2022). ERβ Isoforms Have Differential Clinical Significance in Breast Cancer Subtypes and Subgroups. *Current Issues in Molecular Biology*, *44*(4), 1564–1586. https://doi.org/10.3390/cimb44040107

Chou, Y.-T., Hsieh, C.-H., Chiou, S.-H., Hsu, C.-F., Kao, Y.-R., Lee, C.-C., Chung, C.-H., Wang, Y.-H., Hsu, H.-S., Pang, S.-T., Shieh, Y.-S., & Wu, C.-W. (2012). CITED2 functions as a molecular switch of cytokine-induced proliferation and quiescence. *Cell Death & Differentiation*, *19*(12), 2015–2028. https://doi.org/10.1038/cdd.2012.91

Clay, D. E., & Fox, D. T. (2021). DNA Damage Responses during the Cell Cycle: Insights from Model Organisms and Beyond. *Genes*, *12*(12), 1882. https://doi.org/10.3390/genes12121882

Cloonan, N., Forrest, A. R. R., Kolle, G., Gardiner, B. B. A., Faulkner, G. J., Brown, M. K., Taylor, D. F., Steptoe, A. L., Wani, S., Bethel, G., Robertson, A. J., Perkins, A. C., Bruce, S. J., Lee, C. C., Ranade, S. S., Peckham, H. E., Manning, J. M., McKernan, K. J., & Grimmond, S. M. (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature Methods*, *5*(7), 613–619. https://doi.org/10.1038/nmeth.1223

Clusan, L., Ferrière, F., Flouriot, G., & Pakdel, F. (2023). A Basic Review on Estrogen Receptor Signaling Pathways in Breast Cancer. *International Journal of Molecular Sciences*, *24*(7), 6834. https://doi.org/10.3390/ijms24076834

Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, *38*(6), 1767–1771. https://doi.org/10.1093/nar/gkp1137

Cohen, I. C., Cohenour, E. R., Harnett, K. G., & Schuh, S. M. (2021). BPA, BPAF and TMBPF Alter Adipogenesis and Fat Accumulation in Human Mesenchymal Stem Cells, with Implications for Obesity. *International Journal of Molecular Sciences*, *22*(10), 5363. https://doi.org/10.3390/ijms22105363

Collaborative Group on Hormonal Factors in Breast Cancer. (1996). Breast cancer and hormonal contraceptives: Collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. *The Lancet*, *347*(9017), 1713–1727. https://doi.org/10.1016/S0140-6736(96)90806-5

Collaborative Group on Hormonal Factors in Breast Cancer. (1997). Breast cancer and hormone replacement therapy: Collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Collaborative Group on Hormonal Factors in Breast Cancer. *Lancet (London, England)*, *350*(9084), 1047–1059.

Condon, J. C., Hardy, D. B., Kovaric, K., & Mendelson, C. R. (2006). Up-Regulation of the Progesterone Receptor (PR)-C Isoform in Laboring Myometrium by Activation of Nuclear Factor-κB May Contribute to the Onset of Labor through Inhibition of PR Function. *Molecular Endocrinology*, *20*(4), 764–775. https://doi.org/10.1210/me.2005-0242

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, *17*, 13. https://doi.org/10.1186/s13059-016-0881-8

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex Genome

Engineering Using CRISPR/Cas Systems. *Science*, *339*(6121), 819–823. https://doi.org/10.1126/science.1231143

Cotarelo, C. L., Schad, A., Kirkpatrick, C. J., Sleeman, J. P., Springer, E., Schmidt, M., & Thaler, S. (2016). Detection of cellular senescence within human invasive breast carcinomas distinguishes different breast tumor subtypes. *Oncotarget*, *7*(46), 74846–74859. https://doi.org/10.18632/oncotarget.12432

Crane, M. G., & Harris, J. J. (1969). Plasma Renin Activity and Aldosterone Excretion Rate in Normal Subjects. II. Effect of Oral Contraceptive Agents. *The Journal of Clinical Endocrinology & Metabolism*, *29*(4), 558–562. https://doi.org/10.1210/jcem-29-4-558

Cribbs, A. (2022). *scAmpDesign: What the Package Does (Title Case)* (R package version 0.1.0) [Computer software]. https://github.com/Acribbs/scAmpDesign

Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., Jupe, S., Kalatskaya, I., Mahajan, S., May, B., Ndegwa, N., Schmidt, E., Shamovsky, V., Yung, C., Birney, E., … Stein, L. (2011). Reactome: A database of reactions, pathways and biological processes. *Nucleic Acids Research*, *39*(Database), D691–D697. https://doi.org/10.1093/nar/gkq1018

Cui, H., Guo, D., Zhang, X., Zhu, Y., Wang, Z., Jin, Y., Guo, W., & Zhang, S. (2021). ENO3 Inhibits Growth and Metastasis of Hepatocellular Carcinoma via Wnt/β-Catenin Signaling Pathway. *Frontiers in Cell and Developmental Biology*, *9*, 797102. https://doi.org/10.3389/fcell.2021.797102

Cuilla, K. (2022). *reactablefmtr: Streamlined Table Styling and Formatting for Reactable* (R package version 2.0.0) [Computer software]. https://CRAN.R-project.org/package=reactablefmtr

Curran, S. P., & Ruvkun, G. (2007). Lifespan Regulation by Evolutionarily Conserved Genes Essential for Viability. *PLoS Genetics*, *3*(4), e56. https://doi.org/10.1371/journal.pgen.0030056

Dai, Z., Wang, L., Wang, X., Zhao, B., Zhao, W., Bhardwaj, S. S., Ye, J., Yin, Z., Zhang, J., & Zhao, S. (2018). Oxymatrine induces cell cycle arrest and apoptosis and suppresses the invasion of human glioblastoma cells through the EGFR/PI3K/Akt/mTOR signaling pathway and STAT3. *Oncology Reports*, *40*(2), 867–876. https://doi.org/10.3892/or.2018.6512

D'Alonzo, M., Bounous, V. E., Villa, M., & Biglia, N. (2019). Current Evidence of the Oncological Benefit-Risk Profile of Hormone Replacement Therapy. *Medicina*, *55*(9), 573. https://doi.org/10.3390/medicina55090573

D'Assoro, A. B., Barrett, S. L., Folk, C., Negron, V. C., Boeneman, K., Busby, R., Whitehead, C., Stivala, F., Lingle, W. L., & Salisbury, J. L. (2002). Amplified Centrosomes in Breast Cancer: A Potential Indicator of Tumor Aggressiveness. *Breast Cancer Research and Treatment*, *75*(1), 25–34. https://doi.org/10.1023/A:1016550619925

DE Azevedo Delou, J. M., Lopes, A. G., & Capella, M. A. M. (2014). Effects of Angiotensin, vasopressin and aldosterone on proliferation of mcf-7 cells and their sensitivity to Doxorubicin. *Anticancer Research*, *34*(4), 1843–1848.

De Bruijn, I., Kundra, R., Mastrogiacomo, B., Tran, T. Ngoc., Sikina, L., Mazor, T., Li, X., Ochoa, A., Zhao, G., Lai, B., Abeshouse, A., Baiceanu, D., Ciftci, E., Dogrusoz, U., Dufilie, A., Erkoc, Z., Garcia Lara, E., Fu, Z., Gross, B. E., … Schultz, N. (2023). Analysis and Visualization of Longitudinal Genomic and Clinical Data from the AACR Project GENIE Biopharma Collaborative in cBioPortal. *Cancer Research*. https://doi.org/10.1158/0008-5472.CAN-23-0816

De Godoy, L. M. F., Olsen, J. V., Cox, J., Nielsen, M. L., Hubner, N. C., Fröhlich, F., Walther, T. C., & Mann, M. (2008). Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature*, *455*(7217), 1251–1254. https://doi.org/10.1038/nature07341

De Sousa Abreu, R., Penalva, L. O., Marcotte, E. M., & Vogel, C. (2009). Global signatures of protein and mRNA expression levels. *Molecular BioSystems*, 10.1039.b908315d. https://doi.org/10.1039/b908315d

de Souto, M. C., Costa, I. G., De Araujo, D. S., Ludermir, T. B., & Schliep, A. (2008). Clustering cancer gene expression data: A comparative study. *BMC Bioinformatics*, *9*(1), 497. https://doi.org/10.1186/1471-2105-9-497

Debacq-Chainiaux, F., Leduc, C., Verbeke, A., & Toussaint, O. (2012). UV, stress and aging. *Dermato-Endocrinology*, *4*(3), 236–240. https://doi.org/10.4161/derm.23652

DeMayo, F. J., & Lydon, J. P. (2020). 90 YEARS OF PROGESTERONE: New insights into progesterone receptor signaling in the endometrium required for embryo implantation. *Journal of Molecular Endocrinology*, *65*(1), T1–T14. https://doi.org/10.1530/JME-19-0212

Dempster, J. M., Boyle, I., Vazquez, F., Root, D. E., Boehm, J. S., Hahn, W. C., Tsherniak, A., & McFarland, J. M. (2021). Chronos: A cell population dynamics model of CRISPR experiments that improves inference of gene fitness effects. *Genome Biology*, *22*(1), 343. https://doi.org/10.1186/s13059-021-02540-7

Dempster, J. M., Rossen, J., Kazachkova, M., Pan, J., Kugener, G., Root, D. E., & Tsherniak, A. (2019). *Extracting Biological Insights from the Project Achilles Genome-Scale CRISPR Screens in Cancer Cell Lines* [Preprint]. Cancer Biology. https://doi.org/10.1101/720243

Desmet, C. J., & Peeper, D. S. (2006). The neurotrophic receptor TrkB: A drug target in anti-cancer therapy? *Cellular and Molecular Life Sciences: CMLS*, *63*(7–8), 755–759. https://doi.org/10.1007/s00018-005-5490-8

Dey, A., Kundu, M., Das, S., Jena, B. C., & Mandal, M. (2022). Understanding the function and regulation of Sox2 for its therapeutic potential in breast cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, *1877*(2), 188692. https://doi.org/10.1016/j.bbcan.2022.188692

Diep, C. H., Ahrendt, H., & Lange, C. A. (2016). Progesterone induces progesterone receptor gene (PGR) expression via rapid activation of protein kinase pathways required for cooperative estrogen receptor alpha (ER) and progesterone receptor (PR) genomic action at ER/PR target genes. *Steroids*, *114*, 48–58. https://doi.org/10.1016/j.steroids.2016.09.004

Diep, C. H., Knutson, T. P., & Lange, C. A. (2016). Active FOXO1 Is a Key Determinant of Isoform-Specific Progesterone Receptor Transactivation and Senescence Programming. *Molecular Cancer Research*, *14*(2), 141–162. https://doi.org/10.1158/1541-7786.MCR-15-0431

Dilley, T. K., Bowden, G. T., & Chen, Q. M. (2003). Novel mechanisms of sublethal oxidant toxicity: Induction of premature senescence in human fibroblasts confers tumor promoter activity. *Experimental Cell Research*, *290*(1), 38–48. https://doi.org/10.1016/S0014-4827(03)00308-2

Dillies, M.-A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, G., Castel, D., Estelle, J., Guernec, G., Jagla, B., Jouneau, L., Laloë, D., Le Gall, C., Schaëffer, B., Le Crom, S., Guedj, M., Jaffrézic, F., & French StatOmique Consortium. (2013). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data

analysis. *Briefings in Bioinformatics*, *14*(6), 671–683. https://doi.org/10.1093/bib/bbs046

Dillon, L. M., & Miller, T. W. (2015). The PREX1/Rac signaling axis: Potential as a biomarker and therapeutic target in breast cancer. *Molecular & Cellular Oncology*, *2*(3), e996016. https://doi.org/10.1080/23723556.2014.996016

Dimri, G. P., Lee, X., Basile, G., Acosta, M., Scott, G., Roskelley, C., Medrano, E. E., Linskens, M., Rubelj, I., & Pereira-Smith, O. (1995). A biomarker that identifies senescent human cells in culture and in aging skin in vivo. *Proceedings of the National Academy of Sciences*, *92*(20), 9363–9367. https://doi.org/10.1073/pnas.92.20.9363

Dixon, S. J., Lemberg, K. M., Lamprecht, M. R., Skouta, R., Zaitsev, E. M., Gleason, C. E., Patel, D. N., Bauer, A. J., Cantley, A. M., Yang, W. S., Morrison, B., & Stockwell, B. R. (2012). Ferroptosis: An iron-dependent form of nonapoptotic cell death. *Cell*, *149*(5), 1060–1072. https://doi.org/10.1016/j.cell.2012.03.042

Doan, T. B., Cheung, V., Clyne, C. D., Hilton, H. N., Eriksson, N., Young, M. J., Funder, J. W., Muscat, G. E. O., Fuller, P. J., Clarke, C. L., & Graham, J. D. (2020). A tumour suppressive relationship between mineralocorticoid and retinoic acid receptors activates a transcriptional program consistent with a reverse Warburg effect in breast cancer. *Breast Cancer Research*, *22*(1), 122. https://doi.org/10.1186/s13058-020-01355-x

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dolgalev, I. (2022). *msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format* (R package version 7.5.1) [Computer software]. Bioconductor. https://CRAN.R-project.org/package=msigdbr

Donaldson, J. G. (2003). Multiple Roles for Arf6: Sorting, Structuring, and Signaling at the Plasma Membrane. *Journal of Biological Chemistry*, *278*(43), 41573–41576. https://doi.org/10.1074/jbc.R300026200

Duan, J., Ruan, B., Song, P., Fang, Z., Yue, Z., Liu, J., Dou, G., Han, H., & Wang, L. (2022). Shear stress–induced cellular senescence blunts liver regeneration through Notch–sirtuin 1–P21/P16 axis. *Hepatology*, *75*(3), 584–599. https://doi.org/10.1002/hep.32209

Dubey, R. (2002). Sex hormones and hypertension. *Cardiovascular Research*, *53*(3), 688–708. https://doi.org/10.1016/S0008-6363(01)00527-2

Dunbier, A. K., Anderson, H., Ghazoui, Z., Folkerd, E. J., A'Hern, R., Crowder, R. J., Hoog, J., Smith, I. E., Osin, P., Nerurkar, A., Parker, J. S., Perou, C. M., Ellis, M. J., & Dowsett, M. (2010). Relationship Between Plasma Estradiol Levels and Estrogen-Responsive Gene Expression in Estrogen Receptor–Positive Breast Cancer in Postmenopausal Women. *Journal of Clinical Oncology*, *28*(7), 1161–1167. https://doi.org/10.1200/JCO.2009.23.9616

Duplessis, T. T., Williams, C. C., Hill, S. M., & Rowan, B. G. (2011). Phosphorylation of Estrogen Receptor α at Serine 118 Directs Recruitment of Promoter Complexes and Gene-Specific Transcription. *Endocrinology*, *152*(6), 2517–2526. https://doi.org/10.1210/en.2010-1281

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., & Huber, W. (2005). BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics*, *21*(16), 3439–3440. https://doi.org/10.1093/bioinformatics/bti525

Durinck, S., Spellman, P. T., Birney, E., & Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, *4*(8), 1184–1191. https://doi.org/10.1038/nprot.2009.97

Dvinge, H., Kim, E., Abdel-Wahab, O., & Bradley, R. K. (2016). RNA splicing factors as oncoproteins and tumour suppressors. *Nature Reviews. Cancer*, *16*(7), 413–430. https://doi.org/10.1038/nrc.2016.51

Dwane, L., Behan, F. M., Gonçalves, E., Lightfoot, H., Yang, W., van der Meer, D., Shepherd, R., Pignatelli, M., Iorio, F., & Garnett, M. J. (2021). Project Score database: A resource for investigating cancer cell dependencies and prioritizing therapeutic targets. *Nucleic Acids Research*, *49*(D1), D1365–D1372. https://doi.org/10.1093/nar/gkaa882

Edwards, C. R. W., & Stewart, P. M. (1991). The cortisol-cortisone shuttle and the apparent specificity of glucocorticoid and mineralocorticoid receptors. *The Journal of Steroid Biochemistry and Molecular Biology*, *39*(5), 859–865. https://doi.org/10.1016/0960-0760(91)90036-5

Edwards, M., & Can, A. S. (2024). Progestins. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK563211/

Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*, *95*(25), 14863–14868. https://doi.org/10.1073/pnas.95.25.14863

Elstein, M., & Furniss, H. A. (1995). Advances in oral hormonal contraception. *Zentralblatt Fur Gynakologie*, *117*(11), 559–565.

Engelman, J. A., Luo, J., & Cantley, L. C. (2006). The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. *Nature Reviews. Genetics*, *7*(8), 606–619. https://doi.org/10.1038/nrg1879

Enmark, E., Pelto-Huikko, M., Grandien, K., Lagercrantz, S., Lagercrantz, J., Fried, G., Nordenskjöld, M., & Gustafsson, J.-Å. (1997). Human Estrogen Receptor β-Gene Structure, Chromosomal Localization, and Expression Pattern1. *The Journal of Clinical Endocrinology & Metabolism*, *82*(12), 4258–4265. https://doi.org/10.1210/jcem.82.12.4470

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Knowledge Discovery and Data Mining*, *96*, 226--231.

Evans, C., Hardin, J., & Stoebel, D. M. (2018). Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Briefings in Bioinformatics*, *19*(5), 776–792. https://doi.org/10.1093/bib/bbx008

Fackler, M. J., Umbricht, C. B., Williams, D., Argani, P., Cruz, L.-A., Merino, V. F., Teo, W. W., Zhang, Z., Huang, P., Visvananthan, K., Marks, J., Ethier, S., Gray, J. W., Wolff, A. C., Cope, L. M., & Sukumar, S. (2011). Genome-wide Methylation Analysis Identifies Genes Specific to Breast Cancer Hormone Receptor Status and Risk of Recurrence. *Cancer Research*, *71*(19), 6195–6207. https://doi.org/10.1158/0008-5472.CAN-11-1630

Fagundes, R., & Teixeira, L. K. (2021). Cyclin E/CDK2: DNA Replication, Replication Stress and Genomic Instability. *Frontiers in Cell and Developmental Biology*, *9*, 774845. https://doi.org/10.3389/fcell.2021.774845

Faivre, E. J., Daniel, A. R., Hillard, C. J., & Lange, C. A. (2008). Progesterone receptor rapid signaling mediates serine 345 phosphorylation and tethering to specificity protein 1 transcription factors. *Molecular Endocrinology (Baltimore, Md.)*, *22*(4), 823–837. https://doi.org/10.1210/me.2007-0437

FDA label, Bayer Healthcare Pharmaceuticals Inc. (2006). *YAZ (drospirenone/ethinyl estradiol).* https://www.accessdata.fda.gov/drugsatfda_docs/label/2012/021676s012lbl.pdf

FDA label, Exceltis USA Inc. (2019). *SLYND (drospirenone).* https://www.accessdata.fda.gov/drugsatfda_docs/label/2019/211367s000lbl.pdf

Fedotcheva, T. A., Fedotcheva, N. I., & Shimanovsky, N. L. (2021). Progestins as Anticancer Drugs and Chemosensitizers, New Targets and Applications. *Pharmaceutics*, *13*(10), 1616. https://doi.org/10.3390/pharmaceutics13101616

Feng, J., Liu, T., Qin, B., Zhang, Y., & Liu, X. S. (2012). Identifying ChIP-seq enrichment using MACS. *Nature Protocols*, *7*(9), 1728–1740. https://doi.org/10.1038/nprot.2012.101

Feng, M., Zhang, T., & Ma, H. (2019). Progesterone ameliorates the endometrial polyp by modulating the signaling pathway of Wnt and β-catenin via regulating the expression of H19 and miR-152. *Journal of Cellular Biochemistry*, *120*(6), 10164–10174. https://doi.org/10.1002/jcb.28301

Fenner, B. M. (2012). Truncated TrkB: Beyond a dominant negative receptor. *Cytokine & Growth Factor Reviews*, *23*(1–2), 15–24. https://doi.org/10.1016/j.cytogfr.2012.01.002

Ferrari, N., Mohammed, Z. M. A., Nixon, C., Mason, S. M., Mallon, E., McMillan, D. C., Morris, J. S., Cameron, E. R., Edwards, J., & Blyth, K. (2014). Expression of RUNX1 Correlates with Poor Patient Prognosis in Triple Negative Breast Cancer. *PLoS ONE*, *9*(6), e100759. https://doi.org/10.1371/journal.pone.0100759

Ferreira-Gonzalez, S., Lu, W.-Y., Raven, A., Dwyer, B., Man, T. Y., O'Duibhir, E., Lewis, P. J. S., Campana, L., Kendall, T. J., Bird, T. G., Tarrats, N., Acosta, J.-C., Boulter, L., & Forbes, S. J. (2018). Paracrine cellular senescence exacerbates biliary injury and impairs regeneration. *Nature Communications*, *9*(1), 1020. https://doi.org/10.1038/s41467-018-03299-5

Filardo, E. J., Quinn, J. A., Bland, K. I., & Frackelton, A. R. (2000). Estrogen-Induced Activation of Erk-1 and Erk-2 Requires the G Protein-Coupled Receptor Homolog, GPR30, and Occurs via Trans-Activation of the Epidermal Growth Factor Receptor through Release of HB-EGF. *Molecular Endocrinology*, *14*(10), 1649–1660. https://doi.org/10.1210/mend.14.10.0532

Filardo, E. J., & Thomas, P. (2012). Minireview: G Protein-Coupled Estrogen Receptor-1, GPER-1: Its Mechanism of Action and Role in Female Reproductive Cancer, Renal and Vascular Physiology. *Endocrinology*, *153*(7), 2953–2962. https://doi.org/10.1210/en.2012-1061

Fitsiou, E., Pulido, T., Campisi, J., Alimirah, F., & Demaria, M. (2021). Cellular Senescence and the Senescence-Associated Secretory Phenotype as Drivers of Skin Photoaging. *Journal of Investigative Dermatology*, *141*(4), 1119–1126. https://doi.org/10.1016/j.jid.2020.09.031

Flasch, M., Bueschl, C., Del Favero, G., Adam, G., Schuhmacher, R., Marko, D., & Warth, B. (2022). Elucidation of xenoestrogen metabolism by non-targeted, stable isotope-assisted mass spectrometry in breast cancer cells. *Environment International*, *158*, 106940. https://doi.org/10.1016/j.envint.2021.106940

Fletcher, S. M., Resurreccion, A. V. A., & Misra, S. K. (1993). MEASURING AND MODELING CONSUMER ACCEPTANCE. In *Postharvest Handling* (pp. 149–166). Elsevier. https://doi.org/10.1016/B978-0-08-092576-9.50013-6

Flister, M. J., & Bergom, C. (2018). Genetic Modifiers of the Breast Tumor Microenvironment. *Trends in Cancer*, *4*(6), 429–444. https://doi.org/10.1016/j.trecan.2018.04.003

Florian, C. L., Preece, N. E., Bhakoo, K. K., Williams, S. R., & Noble, M. (1995). Characteristic metabolic profiles revealed by ^{1}H NMR spectroscopy for three types of human brain and nervous system tumours. *NMR in Biomedicine*, *8*(6), 253–264. https://doi.org/10.1002/nbm.1940080605

Food and Drug Administration. (2006). *Medical review Yaz (NDA 21–676).* https://www.accessdata.fda.gov/drugsatfda_docs/nda/2006/021676s000_MEDR_pt1.pdf

Formby, B., & Wiley, T. S. (1999). Bcl-2, survivin and variant CD44 v7-v10 are downregulated and p53 is upregulated in breast cancer cells by progesterone: Inhibition of cell growth and induction of apoptosis. *Molecular and Cellular Biochemistry*, *202*(1–2), 53–61. https://doi.org/10.1023/a:1007081021483

Forsburg, S. L. (2004). Eukaryotic MCM proteins: Beyond replication initiation. *Microbiology and Molecular Biology Reviews: MMBR*, *68*(1), 109–131. https://doi.org/10.1128/MMBR.68.1.109-131.2004

Fotherby, K., & Caldwell, A. D. S. (1994). New progestogens in oral contraception. *Contraception*, *49*(1), 1–32. https://doi.org/10.1016/0010-7824(94)90106-6

Fournier, A., Berrino, F., Riboli, E., Avenel, V., & Clavel-Chapelon, F. (2005). Breast cancer risk in relation to different types of hormone replacement therapy in the E3N-EPIC cohort. *International Journal of Cancer*, *114*(3), 448–454. https://doi.org/10.1002/ijc.20710

Fournier, M. L., Paulson, A., Pavelka, N., Mosley, A. L., Gaudenz, K., Bradford, W. D., Glynn, E., Li, H., Sardiu, M. E., Fleharty, B., Seidel, C., Florens, L., & Washburn, M. P. (2010). Delayed Correlation of mRNA and Protein Expression in Rapamycin-treated Cells and a Role for Ggc1 in Cellular Sensitivity to Rapamycin. *Molecular & Cellular Proteomics*, *9*(2), 271–284. https://doi.org/10.1074/mcp.M900415-MCP200

Freischel, A. R., Teer, J. K., Luddy, K., Cunningham, J., Artzy-Randrup, Y., Epstein, T., Tsai, K. Y., Berglund, A., Cleveland, J. L., Gillies, R. J., Brown, J. S., & Gatenby, R. A. (2022). Evolutionary Analysis of TCGA Data Using Over- and Under- Mutated Genes Identify Key Molecular Pathways and Cellular Functions in Lung Cancer Subtypes. *Cancers*, *15*(1), 18. https://doi.org/10.3390/cancers15010018

Freudenberg, J., & Propping, P. (2002). A similarity-based method for genome-wide prediction ofdisease-relevant human genes. *Bioinformatics*, *18*(suppl_2), S110–S115. https://doi.org/10.1093/bioinformatics/18.suppl_2.S110

Freyhult, E., Landfors, M., Önskog, J., Hvidsten, T. R., & Rydén, P. (2010). Challenges in microarray class discovery: A comprehensive examination of normalization, gene selection and clustering. *BMC Bioinformatics*, *11*(1), 503. https://doi.org/10.1186/1471-2105-11-503

Frigo, D. E., Bondesson, M., & Williams, C. (2021). Nuclear receptors: From molecular mechanisms to therapeutics. *Essays in Biochemistry*, *65*(6), 847–856. https://doi.org/10.1042/EBC20210020

Fu, Dong, Tan, Yin, Zhang, Zhao, Ye, & Wu. (2019). Identification of Shoot Differentiation-Related Genes in Populus euphratica Oliv. *Genes*, *10*(12), 1034. https://doi.org/10.3390/genes10121034

Fuentes, N., & Silveyra, P. (2019). Estrogen receptor signaling mechanisms. In *Advances in Protein Chemistry and Structural Biology* (Vol. 116, pp. 135–170). Elsevier. https://doi.org/10.1016/bs.apcsb.2019.01.001

Fuhrmann, U., Krattenmacher, R., Slater, E. P., & Fritzemeier, K. H. (1996). The novel progestin drospirenone and its natural counterpart progesterone: Biochemical profile and antiandrogenic potential. *Contraception*, *54*(4), 243–251. https://doi.org/10.1016/s0010-7824(96)00195-3

Fujikado, N., Saijo, S., & Iwakura, Y. (2006). Identification of arthritis-related gene clusters by microarray analysis of two independent mouse models for rheumatoid arthritis. *Arthritis Research & Therapy*, *8*(4), R100. https://doi.org/10.1186/ar1985

Funder, J. W., Feldman, D., & Edelman, I. S. (1972). Specific aldosterone binding in rat kidney and parotid. *Journal of Steroid Biochemistry*, *3*(2), 209–218. https://doi.org/10.1016/0022-4731(72)90052-0

Funder, J. W., Pearce, P. T., Smith, R., & Smith, A. I. (1988). Mineralocorticoid Action: Target Tissue Specificity Is Enzyme, Not Receptor, Mediated. *Science*, *242*(4878), 583–585. https://doi.org/10.1126/science.2845584

Futcher, B., Latter, G. I., Monardo, P., McLaughlin, C. S., & Garrels, J. I. (1999). A sampling of the yeast proteome. *Molecular and Cellular Biology*, *19*(11), 7357–7368. https://doi.org/10.1128/MCB.19.11.7357

Gaddy, V. T., Barrett, J. T., Delk, J. N., Kallab, A. M., Porter, A. G., & Schoenlein, P. V. (2004). Mifepristone Induces Growth Arrest, Caspase Activation, and Apoptosis of Estrogen Receptor-Expressing, Antiestrogen-Resistant Breast Cancer Cells. *Clinical Cancer Research*, *10*(15), 5215–5225. https://doi.org/10.1158/1078-0432.CCR-03-0637

Galigniana, M. D., Echeverría, P. C., Erlejman, A. G., & Piwien-Pilipuk, G. (2010). Role of molecular chaperones and TPR-domain proteins in the cytoplasmic transport of steroid receptors and their passage through the nuclear pore. *Nucleus*, *1*(4), 299–308. https://doi.org/10.4161/nucl.1.4.11743

Galili, T., O'Callaghan, A., Sidi, J., & Sievert, C. (2018). heatmaply: An R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics*, *34*(9), 1600–1602. https://doi.org/10.1093/bioinformatics/btx657

Gallez, A., Dias Da Silva, I., Wuidar, V., Foidart, J.-M., & Péqueux, C. (2021). Estetrol and Mammary Gland: Friends or Foes? *Journal of Mammary Gland Biology and Neoplasia*, *26*(3), 297–308. https://doi.org/10.1007/s10911-021-09497-0

Gallia, G. L., Rand, V., Siu, I.-M., Eberhart, C. G., James, C. D., Marie, S. K. N., Oba-Shinjo, S. M., Carlotti, C. G., Caballero, O. L., Simpson, A. J. G., Brock, M. V., Massion, P. P., Carson, B. S., & Riggins, G. J. (2006). *PIK3CA* Gene Mutations in Pediatric and Adult Glioblastoma Multiforme. *Molecular Cancer Research*, *4*(10), 709–714. https://doi.org/10.1158/1541-7786.MCR-06-0172

Gao, C. X., Dwyer, D., Zhu, Y., Smith, C. L., Du, L., Filia, K. M., Bayer, J., Menssink, J. M., Wang, T., Bergmeir, C., Wood, S., & Cotton, S. M. (2023). An overview of clustering methods with guidelines for application in mental health research. *Psychiatry Research*, *327*, 115265. https://doi.org/10.1016/j.psychres.2023.115265

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., & Schultz, N. (2013). Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Science Signaling*, *6*(269). https://doi.org/10.1126/scisignal.2004088

Garami, A., Zwartkruis, F. J. T., Nobukuni, T., Joaquin, M., Roccio, M., Stocker, H., Kozma, S. C., Hafen, E., Bos, J. L., & Thomas, G. (2003). Insulin activation of Rheb, a mediator of mTOR/S6K/4E-BP signaling, is inhibited by TSC1 and 2. *Molecular Cell*, *11*(6), 1457–1466. https://doi.org/10.1016/s1097-2765(03)00220-x

Garcia, J. H., Jain, S., & Aghi, M. K. (2021). Metabolic Drivers of Invasion in Glioblastoma. *Frontiers in Cell and Developmental Biology*, *9*, 683276. https://doi.org/10.3389/fcell.2021.683276

García-Gutiérrez, L., Delgado, M. D., & León, J. (2019). MYC Oncogene Contributions to Release of Cell Cycle Brakes. *Genes*, *10*(3), 244. https://doi.org/10.3390/genes10030244

García-Sáenz, M., Ibarra-Salce, R., Pozos-Varela, F. J., Mena-Ureta, T. S., Flores-Villagómez, S., Santana-Mata, M., De Los Santos-Aguilar, R. G., Uribe-Cortés, D., & Ferreira-Hermosillo, A. (2023). Understanding Progestins: From Basics to Clinical Applicability. *Journal of Clinical Medicine*, *12*(10), 3388. https://doi.org/10.3390/jcm12103388

Gatalica, Z., Xiu, J., Swensen, J., & Vranic, S. (2019). Molecular characterization of cancers with NTRK gene fusions. *Modern Pathology: An Official Journal of the United States and Canadian Academy of Pathology, Inc*, *32*(1), 147–153. https://doi.org/10.1038/s41379-018-0118-3

Gavrilovic, I. T., & Posner, J. B. (2005). Brain metastases: Epidemiology and pathophysiology. *Journal of Neuro-Oncology*, *75*(1), 5–14. https://doi.org/10.1007/s11060-004-8093-6

Ge, S. X., Son, E. W., & Yao, R. (2018). iDEP: An integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics*, *19*(1), 534. https://doi.org/10.1186/s12859-018-2486-6

Gee, J. M., Robertson, J. F., Ellis, I. O., Willsher, P., McClelland, R. A., Hoyle, H. B., Kyme, S. R., Finlay, P., Blamey, R. W., & Nicholson, R. I. (1994). Immunocytochemical localization of BCL-2 protein in human breast cancers and its relationship to a series of prognostic markers and response to endocrine therapy. *International Journal of Cancer*, *59*(5), 619–628. https://doi.org/10.1002/ijc.2910590508

Geller, D. S., Farhi, A., Pinkerton, N., Fradley, M., Moritz, M., Spitzer, A., Meinke, G., Tsai, F. T. F., Sigler, P. B., & Lifton, R. P. (2000). Activating Mineralocorticoid Receptor Mutation in Hypertension Exacerbated by Pregnancy. *Science*, *289*(5476), 119–123. https://doi.org/10.1126/science.289.5476.119

Gellersen, B., Fernandes, M. S., & Brosens, J. J. (2009). Non-genomic progesterone actions in female reproduction. *Human Reproduction Update*, *15*(1), 119–138. https://doi.org/10.1093/humupd/dmn044

Gerdes, D., Wehling, M., Leube, B., & Falkenstein, E. (1998). Cloning and tissue expression of two putative steroid membrane receptors. *Biological Chemistry*, *379*(7), 907–911. https://doi.org/10.1515/bchm.1998.379.7.907

Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., & Weissman, J. S. (2003). Global analysis of protein expression in yeast. *Nature*, *425*(6959), 737–741. https://doi.org/10.1038/nature02046

Ghafouri-Fard, S., Ahmadi Teshnizi, S., Hussen, B. M., Taheri, M., & Sharifi, G. (2023). A review on the role of NDRG1 in different cancers. *Molecular Biology Reports*, *50*(7), 6251–6264. https://doi.org/10.1007/s11033-023-08540-z

Girault, I., Lerebours, F., Amarir, S., Tozlu, S., Tubiana-Hulin, M., Lidereau, R., & Bièche, I. (2003). Expression analysis of estrogen receptor alpha coregulators in

breast carcinoma: Evidence that NCOR1 expression is predictive of the response to tamoxifen. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *9*(4), 1259–1266.

Glass, K., Huttenhower, C., Quackenbush, J., & Yuan, G.-C. (2013). Passing Messages between Biological Networks to Refine Predicted Interactions. *PLoS ONE*, *8*(5), e64832. https://doi.org/10.1371/journal.pone.0064832

Glont, S.-E., Chernukhin, I., & Carroll, J. S. (2019). Comprehensive Genomic Analysis Reveals that the Pioneering Function of FOXA1 Is Independent of Hormonal Signaling. *Cell Reports*, *26*(10), 2558-2565.e3. https://doi.org/10.1016/j.celrep.2019.02.036

Godbole, M., Chandrani, P., Gardi, N., Dhamne, H., Patel, K., Yadav, N., Gupta, S., Badwe, R., & Dutt, A. (2017). miR-129-2 mediates down-regulation of progesterone receptor in response to progesterone in breast cancer cells. *Cancer Biology & Therapy*, *18*(10), 801–805. https://doi.org/10.1080/15384047.2017.1373216

Godbole, M., Togar, T., Patel, K., Dharavath, B., Yadav, N., Janjuha, S., Gardi, N., Tiwary, K., Terwadkar, P., Desai, S., Prasad, R., Dhamne, H., Karve, K., Salunkhe, S., Kawle, D., Chandrani, P., Dutt, S., Gupta, S., Badwe, R. A., & Dutt, A. (2018). Up-regulation of the kinase gene SGK1 by progesterone activates the AP-1–NDRG1 axis in both PR-positive and -negative breast cancer cells. *Journal of Biological Chemistry*, *293*(50), 19263–19276. https://doi.org/10.1074/jbc.RA118.002894

Goeman, J. J., & Bühlmann, P. (2007). Analyzing gene expression data in terms of gene sets: Methodological issues. *Bioinformatics*, *23*(8), 980–987. https://doi.org/10.1093/bioinformatics/btm051

Goletiani, N. V., Keith, D. R., & Gorsky, S. J. (2007). Progesterone: Review of safety for clinical studies. *Experimental and Clinical Psychopharmacology*, *15*(5), 427–444. https://doi.org/10.1037/1064-1297.15.5.427

Gomez-Sanchez, E., & Gomez-Sanchez, C. E. (2014). The Multifaceted Mineralocorticoid Receptor. In R. Terjung (Ed.), *Comprehensive Physiology* (1st ed., pp. 965–994). Wiley. https://doi.org/10.1002/cphy.c130044

González-Suárez, E. (2011). RANKL inhibition: A promising novel strategy for breast cancer treatment. *Clinical and Translational Oncology*, *13*(4), 222–228. https://doi.org/10.1007/s12094-011-0646-5

Gonzalez-Suarez, E., Branstetter, D., Armstrong, A., Dinh, H., Blumberg, H., & Dougall, W. C. (2007). RANK Overexpression in Transgenic Mice with Mouse Mammary Tumor Virus Promoter-Controlled RANK Increases Proliferation and Impairs Alveolar Differentiation in the Mammary Epithelia and Disrupts Lumen Formation in Cultured Epithelial Acini. *Molecular and Cellular Biology*, *27*(4), 1442–1454. https://doi.org/10.1128/MCB.01298-06

Gorgoulis, V., Adams, P. D., Alimonti, A., Bennett, D. C., Bischof, O., Bishop, C., Campisi, J., Collado, M., Evangelou, K., Ferbeyre, G., Gil, J., Hara, E., Krizhanovsky, V., Jurk, D., Maier, A. B., Narita, M., Niedernhofer, L., Passos, J. F., Robbins, P. D., … Demaria, M. (2019). Cellular Senescence: Defining a Path Forward. *Cell*, *179*(4), 813–827. https://doi.org/10.1016/j.cell.2019.10.005

Gosden, J. R., Middleton, P. G., & Rout, D. (1986). Localization of the human oestrogen receptor gene to chromosome 6q24→q27 by in situ hybridization. *Cytogenetic and Genome Research*, *43*(3–4), 218–220. https://doi.org/10.1159/000132325

Graham, J. D., & Clarke, C. L. (1997). Physiological action of progesterone in target tissues. *Endocrine Reviews*, *18*(4), 502–519. https://doi.org/10.1210/edrv.18.4.0308

Graham, J. D., Yager, M. L., Hill, H. D., Byth, K., O'Neill, G. M., & Clarke, C. L. (2005). Altered Progesterone Receptor Isoform Expression Remodels Progestin Responsiveness of Breast Cancer Cells. *Molecular Endocrinology*, *19*(11), 2713–2735. https://doi.org/10.1210/me.2005-0126

Greenbaum, D., Colangelo, C., Williams, K., & Gerstein, M. (2003). Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biology*, *4*(9), 117. https://doi.org/10.1186/gb-2003-4-9-117

Griffin, J. L., & Kauppinen, R. A. (2007). A metabolomics perspective of human brain tumours. *The FEBS Journal*, *274*(5), 1132–1139. https://doi.org/10.1111/j.1742-4658.2007.05676.x

Grillo, A., Salvi, L., Coruzzi, P., Salvi, P., & Parati, G. (2019). Sodium Intake and Hypertension. *Nutrients*, *11*(9), 1970. https://doi.org/10.3390/nu11091970

Grimm, S. L., Hartig, S. M., & Edwards, D. P. (2016). Progesterone Receptor Signaling Mechanisms. *Journal of Molecular Biology*, *428*(19), 3831–3849. https://doi.org/10.1016/j.jmb.2016.06.020

Grindberg, R. V., Yee-Greenbaum, J. L., McConnell, M. J., Novotny, M., O'Shaughnessy, A. L., Lambert, G. M., Araúzo-Bravo, M. J., Lee, J., Fishman, M., Robbins, G. E., Lin, X., Venepally, P., Badger, J. H., Galbraith, D. W., Gage, F. H., & Lasken, R. S. (2013). RNA-sequencing from single nuclei. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(49), 19802–19807. https://doi.org/10.1073/pnas.1319700110

Grossmann, C., Ruhs, S., Langenbruch, L., Mildenberger, S., Strätz, N., Schumann, K., & Gekle, M. (2012). Nuclear Shuttling Precedes Dimerization in Mineralocorticoid Receptor Signaling. *Chemistry & Biology*, *19*(6), 742–751. https://doi.org/10.1016/j.chembiol.2012.04.014

Gu, Z., Steinmetz, L. M., Gu, X., Scharfe, C., Davis, R. W., & Li, W.-H. (2003). Role of duplicate genes in genetic robustness against null mutations. *Nature*, *421*(6918), 63–66. https://doi.org/10.1038/nature01198

Guida, M., Quercitelli, L., De Franciscis, P., Ferrara, C., Marietta, M., Iaccheri, M., Facchinetti, F., Capasso, F., & Grandi, G. (2023). A new progestin-only pill (POP): The impact of drospirenone-only pill 4 mg 24 + 4 on coagulation markers and bleeding patterns. *The European Journal of Contraception & Reproductive Health Care*, *28*(6), 308–312. https://doi.org/10.1080/13625187.2023.2276668

Guo, W., Tzioutziou, N. A., Stephen, G., Milne, I., Calixto, C. P., Waugh, R., Brown, J. W. S., & Zhang, R. (2021). 3D RNA-seq: A powerful and flexible tool for rapid and accurate differential expression and alternative splicing analysis of RNA-seq data for biologists. *RNA Biology*, *18*(11), 1574–1587. https://doi.org/10.1080/15476286.2020.1858253

Gupta, V., You, Y., Gupta, V., Klistorner, A., & Graham, S. (2013). TrkB Receptor Signalling: Implications in Neurodegenerative, Psychiatric and Proliferative Disorders. *International Journal of Molecular Sciences*, *14*(5), 10122–10142. https://doi.org/10.3390/ijms140510122

Gygi, S. P., Rochon, Y., Franza, B. R., & Aebersold, R. (1999). Correlation between protein and mRNA abundance in yeast. *Molecular and Cellular Biology*, *19*(3), 1720–1730. https://doi.org/10.1128/MCB.19.3.1720

Haapasalo, A., Saarelainen, T., Moshnyakov, M., Arumäe, U., Kiema, T. R., Saarma, M., Wong, G., & Castrén, E. (1999). Expression of the naturally occurring truncated trkB neurotrophin receptor induces outgrowth of filopodia and processes in neuroblastoma cells. *Oncogene*, *18*(6), 1285–1296. https://doi.org/10.1038/sj.onc.1202401

Habedanck, R., Stierhof, Y.-D., Wilkinson, C. J., & Nigg, E. A. (2005). The Polo kinase Plk4 functions in centriole duplication. *Nature Cell Biology*, *7*(11), 1140–1146. https://doi.org/10.1038/ncb1320

Habermann, B. H., & Haering, M. (2021). RNfuzzyApp: An R shiny RNA-seq data analysis app for visualisation, differential expression analysis, time-series clustering and enrichment analysis. *F1000Research 2021 10:654*, *10*, 654. https://doi.org/10.12688/f1000research.54533.2

Hamilton, K. J., Hewitt, S. C., Arao, Y., & Korach, K. S. (2017). Estrogen Hormone Biology. In *Current Topics in Developmental Biology* (Vol. 125, pp. 109–146). Elsevier. https://doi.org/10.1016/bs.ctdb.2016.12.005

Hampel, B., Malisan, F., Niederegger, H., Testi, R., & Jansen-Dürr, P. (2004). Differential regulation of apoptotic cell death in senescent human cells. *Experimental Gerontology*, *39*(11–12), 1713–1721. https://doi.org/10.1016/j.exger.2004.05.010

Han, S., Liu, Y., Cai, S. J., Qian, M., Ding, J., Larion, M., Gilbert, M. R., & Yang, C. (2020). IDH mutation in glioma: Molecular mechanisms and potential therapeutic targets. *British Journal of Cancer*, *122*(11), 1580–1589. https://doi.org/10.1038/s41416-020-0814-x

Hardcastle, T. J., & Kelly, K. A. (2010). baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics*, *11*(1), 422. https://doi.org/10.1186/1471-2105-11-422

Hart, T., Brown, K. R., Sircoulomb, F., Rottapel, R., & Moffat, J. (2014). Measuring error rates in genomic perturbation screens: Gold standards for human functional genomics. *Molecular Systems Biology*, *10*(7), 733. https://doi.org/10.15252/msb.20145216

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K. R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., Mero, P., Dirks, P., Sidhu, S., Roth, F. P., Rissland, O. S., Durocher, D., Angers, S., & Moffat, J. (2015). High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell*, *163*(6), 1515–1526. https://doi.org/10.1016/j.cell.2015.11.015

Hartigan, J. A., & Hartigan, P. M. (1985). The Dip Test of Unimodality. *The Annals of Statistics*, *13*(1). https://doi.org/10.1214/aos/1176346577

Hartmann, M., Brigadski, T., Erdmann, K. S., Holtmann, B., Sendtner, M., Narz, F., & Lessmann, V. (2004). Truncated TrkB receptor-induced outgrowth of dendritic filopodia involves the p75 neurotrophin receptor. *Journal of Cell Science*, *117*(Pt 24), 5803–5814. https://doi.org/10.1242/jcs.01511

Hashimoto, S., Onodera, Y., Hashimoto, A., Tanaka, M., Hamaguchi, M., Yamada, A., & Sabe, H. (2004). Requirement for Arf6 in breast cancer invasive activities. *Proceedings of the National Academy of Sciences*, *101*(17), 6647–6652. https://doi.org/10.1073/pnas.0401753101

Hashimshony, T., Wagner, F., Sher, N., & Yanai, I. (2012). CEL-Seq: Single-cell RNA-Seq by multiplexed linear amplification. *Cell Reports*, *2*(3), 666–673. https://doi.org/10.1016/j.celrep.2012.08.003

Hayflick, L. (1965). The limited in vitro lifetime of human diploid cell strains. *Experimental Cell Research*, *37*(3), 614–636. https://doi.org/10.1016/0014-4827(65)90211-9

Hellwig, B., Hengstler, J. G., Schmidt, M., Gehrmann, M. C., Schormann, W., & Rahnenführer, J. (2010). Comparison of scores for bimodality of gene expression distributions and genome-wide evaluation of the prognostic relevance of high-scoring genes. *BMC Bioinformatics*, *11*(1), 276. https://doi.org/10.1186/1471-2105-11-276

Herranz, N., & Gil, J. (2018). Mechanisms and functions of cellular senescence. *Journal of Clinical Investigation*, *128*(4), 1238–1246. https://doi.org/10.1172/JCI95148

Hewitt, S. C., Winuthayanon, W., & Korach, K. S. (2016). What's new in estrogen receptor action in the female reproductive tract. *Journal of Molecular Endocrinology*, *56*(2), R55–R71. https://doi.org/10.1530/JME-15-0254

Hilton, H. N., Graham, J. D., Kantimm, S., Santucci, N., Cloosterman, D., Huschtscha, L. I., Mote, P. A., & Clarke, C. L. (2012). Progesterone and estrogen receptors segregate into different cell subpopulations in the normal human breast. *Molecular and Cellular Endocrinology*, *361*(1–2), 191–201. https://doi.org/10.1016/j.mce.2012.04.010

Hinohara, K., Wu, H.-J., Vigneau, S., McDonald, T. O., Igarashi, K. J., Yamamoto, K. N., Madsen, T., Fassl, A., Egri, S. B., Papanastasiou, M., Ding, L., Peluffo, G., Cohen, O., Kales, S. C., Lal-Nag, M., Rai, G., Maloney, D. J., Jadhav, A., Simeonov, A., … Polyak, K. (2018). KDM5 Histone Demethylase Activity Links Cellular Transcriptomic Heterogeneity to Therapeutic Resistance. *Cancer Cell*, *34*(6), 939-953.e9. https://doi.org/10.1016/j.ccell.2018.10.014

Hirschberg, A. L., Tani, E., Brismar, K., & Lundström, E. (2019). Effects of drospirenone and norethisterone acetate combined with estradiol on mammographic density and proliferation of breast epithelial cells—A prospective randomized trial. *Maturitas*, *126*, 18–24. https://doi.org/10.1016/j.maturitas.2019.04.205

Holst, J. P., Soldin, O. P., Guo, T., & Soldin, S. J. (2004). Steroid hormones: Relevance and measurement in the clinical laboratory. *Clinics in Laboratory Medicine*, *24*(1), 105–118. https://doi.org/10.1016/j.cll.2004.01.004

Hopp, T. A., Weiss, H. L., Hilsenbeck, S. G., Cui, Y., Allred, D. C., Horwitz, K. B., & Fuqua, S. A. W. (2004). Breast Cancer Patients with Progesterone Receptor PR-A-Rich Tumors Have Poorer Disease-Free Survival Rates. *Clinical Cancer Research*, *10*(8), 2751–2760. https://doi.org/10.1158/1078-0432.CCR-03-0141

Horwitz, K. B., & Freidenberg, G. R. (1985). Growth inhibition and increase of insulin receptors in antiestrogen-resistant T47DCO human breast cancer cells by progestins: Implications for endocrine therapies. *Cancer Research*, *45*(1), 167–173.

Hou, X., Tan, Y., Li, M., Dey, S. K., & Das, S. K. (2004). Canonical Wnt Signaling Is Critical to Estrogen-Mediated Uterine Growth. *Molecular Endocrinology*, *18*(12), 3035–3049. https://doi.org/10.1210/me.2004-0259

Huang, B., Omoto, Y., Iwase, H., Yamashita, H., Toyama, T., Coombes, R. C., Filipovic, A., Warner, M., & Gustafsson, J.-Å. (2014). Differential expression of estrogen receptor α, β1, and β2 in lobular and ductal breast cancer. *Proceedings of the National Academy of Sciences*, *111*(5), 1933–1938. https://doi.org/10.1073/pnas.1323719111

Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K.

D., Irizarry, R. A., Lawrence, M., Love, M. I., MacDonald, J., Obenchain, V., Oleś, A. K., … Morgan, M. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*, *12*(2), 115–121. https://doi.org/10.1038/nmeth.3252

Huda, N., Liu, G., Hong, H., Yan, S., Khambu, B., & Yin, X.-M. (2019). Hepatic senescence, the good and the bad. *World Journal of Gastroenterology*, *25*(34), 5069–5081. https://doi.org/10.3748/wjg.v25.i34.5069

Hudson, W. H., Youn, C., & Ortlund, E. A. (2014). Crystal Structure of the Mineralocorticoid Receptor DNA Binding Domain in Complex with DNA. *PLoS ONE*, *9*(9), e107000. https://doi.org/10.1371/journal.pone.0107000

Huh, M.-H. (2002). Setting the Number of Clusters in K-Means Clustering. In Y. Baba, A. J. Hayter, K. Kanefuji, & S. Kuriki (Eds.), *Recent Advances in Statistical Research and Data Analysis* (pp. 115–124). Springer Japan. https://doi.org/10.1007/978-4-431-68544-9_5

Humphreys, G. I., Ziegler, Y. S., & Nardulli, A. M. (2014). 17β-Estradiol Modulates Gene Expression in the Female Mouse Cerebral Cortex. *PLoS ONE*, *9*(11), e111975. https://doi.org/10.1371/journal.pone.0111975

Hung, C.-M., Garcia-Haro, L., Sparks, C. A., & Guertin, D. A. (2012). mTOR-Dependent Cell Survival Mechanisms. *Cold Spring Harbor Perspectives in Biology*, *4*(12), a008771–a008771. https://doi.org/10.1101/cshperspect.a008771

Hurtado, A., Holmes, K. A., Geistlinger, T. R., Hutcheson, I. R., Nicholson, R. I., Brown, M., Jiang, J., Howat, W. J., Ali, S., & Carroll, J. S. (2008). Regulation of ERBB2 by oestrogen receptor–PAX2 determines response to tamoxifen. *Nature*, *456*(7222), 663–666. https://doi.org/10.1038/nature07483

Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., & Hood, L. (2001). Integrated Genomic and Proteomic Analyses of a Systematically Perturbed Metabolic Network. *Science*, *292*(5518), 929–934. https://doi.org/10.1126/science.292.5518.929

Idossa, D., Borrero, M., & Blaes, A. (2023). ERBB2-Low (Also Known as HER2-Low) Breast Cancer. *JAMA Oncology*, *9*(4), 576. https://doi.org/10.1001/jamaoncol.2022.6889

Ignatov, T., Treeck, O., Kalinski, T., Ortmann, O., & Ignatov, A. (2020). GPER-1 expression is associated with a decreased response rate to primary tamoxifen therapy of breast cancer patients. *Archives of Gynecology and Obstetrics*, *301*(2), 565–571. https://doi.org/10.1007/s00404-019-05384-6

Ignatov, T., Weißenborn, C., Poehlmann, A., Lemke, A., Semczuk, A., Roessner, A., Costa, S. D., Kalinski, T., & Ignatov, A. (2013). GPER-1 Expression Decreases During Breast Cancer Tumorigenesis. *Cancer Investigation*, *31*(5), 309–315. https://doi.org/10.3109/07357907.2013.789901

Imelfort, M., Duran, C., Batley, J., & Edwards, D. (2009). Discovering genetic polymorphisms in next-generation sequencing data. *Plant Biotechnology Journal*, *7*(4), 312–317. https://doi.org/10.1111/j.1467-7652.2009.00406.x

Inoki, K., Li, Y., Zhu, T., Wu, J., & Guan, K.-L. (2002). TSC2 is phosphorylated and inhibited by Akt and suppresses mTOR signalling. *Nature Cell Biology*, *4*(9), 648–657. https://doi.org/10.1038/ncb839

Iqbal, J., Andrew, R., Cruden, N. L., Kenyon, C. J., Hughes, K. A., Newby, D. E., Hadoke, P. W. F., & Walker, B. R. (2014). Displacement of Cortisol From Human Heart by Acute Administration of a Mineralocorticoid Receptor Antagonist. *The Journal of Clinical Endocrinology & Metabolism*, *99*(3), 915–922. https://doi.org/10.1210/jc.2013-2049

Ishihama, Y., Schmidt, T., Rappsilber, J., Mann, M., Hartl, F. U., Kerner, M. J., & Frishman, D. (2008). Protein abundance profiling of the Escherichia coli cytosol. *BMC Genomics*, *9*(1), 102. https://doi.org/10.1186/1471-2164-9-102

Itaya, M. (1995). An estimation of minimal genome size required for life. *FEBS Letters*, *362*(3), 257–260. https://doi.org/10.1016/0014-5793(95)00233-Y

Ivanisenko, V. A., Saik, O. V., Ivanisenko, N. V., Tiys, E. S., Ivanisenko, T. V., Demenkov, P. S., & Kolchanov, N. A. (2015). ANDSystem: An Associative Network Discovery System for automated literature mining in the field of biology. *BMC Systems Biology*, *9*(Suppl 2), S2. https://doi.org/10.1186/1752-0509-9-S2-S2

Jacobsen, B. M., & Horwitz, K. B. (2012). Progesterone receptors, their isoforms and progesterone regulated transcription. *Molecular and Cellular Endocrinology*, *357*(1–2), 18–29. https://doi.org/10.1016/j.mce.2011.09.016

Jacobsen, B. M., Schittone, S. A., Richer, J. K., & Horwitz, K. B. (2005). Progesterone-Independent Effects of Human Progesterone Receptors (PRs) in Estrogen Receptor-Positive Breast Cancer: PR Isoform-Specific Gene Regulation and Tumor Biology. *Molecular Endocrinology*, *19*(3), 574–587. https://doi.org/10.1210/me.2004-0287

Jalilvand, A., Akbari, B., Mirakabad, F. Z., & Ghaderi, F. (2019). *Disease gene prioritization using network topological analysis from a sequence based human functional linkage network*. https://doi.org/10.48550/ARXIV.1904.06973

Jančík, S., Drábek, J., Radzioch, D., & Hajdúch, M. (2010). Clinical Relevance of KRAS in Human Cancers. *Journal of Biomedicine and Biotechnology*, *2010*, 1–13. https://doi.org/10.1155/2010/150960

Jaskowiak, P. A., Costa, I. G., & Campello, R. J. G. B. (2018). Clustering of RNA-Seq samples: Comparison study on cancer data. *Methods*, *132*, 42–49. https://doi.org/10.1016/j.ymeth.2017.07.023

Jayapal, K. P., Philp, R. J., Kok, Y.-J., Yap, M. G. S., Sherman, D. H., Griffin, T. J., & Hu, W.-S. (2008). Uncovering Genes with Divergent mRNA-Protein Dynamics in Streptomyces coelicolor. *PLoS ONE*, *3*(5), e2097. https://doi.org/10.1371/journal.pone.0002097

Jazwinski, S. M. (2000). Aging and longevity genes. *Acta Biochimica Polonica*, *47*(2), 269–279. https://doi.org/10.18388/abp.2000_4007

Jeon, J., Nim, S., Teyra, J., Datti, A., Wrana, J. L., Sidhu, S. S., Moffat, J., & Kim, P. M. (2014). A systematic approach to identify novel cancer drug targets using machine learning, inhibitor design and high-throughput screening. *Genome Medicine*, *6*(7), 57. https://doi.org/10.1186/s13073-014-0057-7

Jeong, H., Mason, S. P., Barabási, A.-L., & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, *411*(6833), 41–42. https://doi.org/10.1038/35075138

Jezova, M., Armando, I., Bregonzio, C., Yu, Z.-X., Qian, S., Ferrans, V. J., Imboden, H., & Saavedra, J. M. (2003). Angiotensin II AT1 and AT2 Receptors Contribute to Maintain Basal Adrenomedullary Norepinephrine Synthesis and Tyrosine Hydroxylase Transcription. *Endocrinology*, *144*(5), 2092–2101. https://doi.org/10.1210/en.2002-0019

Jia, M., Dahlman-Wright, K., & Gustafsson, J.-Å. (2015). Estrogen receptor alpha and beta in health and disease. *Best Practice & Research Clinical Endocrinology & Metabolism*, *29*(4), 557–568. https://doi.org/10.1016/j.beem.2015.04.008

Jiang, J., Yuan, J., Hu, Z., Xu, M., Zhang, Y., Long, M., Fan, Y., Montone, K. T., Tanyi, J. L., Tavana, O., Chan, H. M., Zhang, L., & Hu, X. (2022). Systematic

Pan-Cancer Characterization of Nuclear Receptors Identifies Potential Cancer Biomarkers and Therapeutic Targets. *Cancer Research*, *82*(1), 46–59. https://doi.org/10.1158/0008-5472.CAN-20-3458

Jiang, Z., Guo, J., Shen, J., Jin, M., Xie, S., & Wang, L. (2012). The role of estrogen receptor alpha in mediating chemoresistance in breast cancer cells. *Journal of Experimental & Clinical Cancer Research*, *31*(1), 42. https://doi.org/10.1186/1756-9966-31-42

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, *337*(6096), 816–821. https://doi.org/10.1126/science.1225829

Jones, D. H., Nakashima, T., Sanchez, O. H., Kozieradzki, I., Komarova, S. V., Sarosi, I., Morony, S., Rubin, E., Sarao, R., Hojilla, C. V., Komnenovic, V., Kong, Y.-Y., Schreiber, M., Dixon, S. J., Sims, S. M., Khokha, R., Wada, T., & Penninger, J. M. (2006). Regulation of cancer cell migration and bone metastasis by RANKL. *Nature*, *440*(7084), 692–696. https://doi.org/10.1038/nature04524

Jones, D. T. W., Hutter, B., Jäger, N., Korshunov, A., Kool, M., Warnatz, H.-J., Zichner, T., Lambert, S. R., Ryzhova, M., Quang, D. A. K., Fontebasso, A. M., Stütz, A. M., Hutter, S., Zuckermann, M., Sturm, D., Gronych, J., Lasitschka, B., Schmidt, S., Seker-Cin, H., … International Cancer Genome Consortium PedBrain Tumor Project. (2013). Recurrent somatic alterations of FGFR1 and NTRK2 in pilocytic astrocytoma. *Nature Genetics*, *45*(8), 927–932. https://doi.org/10.1038/ng.2682

Juncker, A. S., Jensen, L. J., Pierleoni, A., Bernsel, A., Tress, M. L., Bork, P., Von Heijne, G., Valencia, A., Ouzounis, C. A., Casadio, R., & Brunak, S. (2009). Sequence-based feature prediction and annotation of proteins. *Genome Biology*, *10*(2), 206. https://doi.org/10.1186/gb-2009-10-2-206

Kadota, M., Yang, H. H., Gomez, B., Sato, M., Clifford, R. J., Meerzaman, D., Dunn, B. K., Wakefield, L. M., & Lee, M. P. (2010). Delineating Genetic Alterations for Tumor Progression in the MCF10A Series of Breast Cancer Cell Lines. *PLoS ONE*, *5*(2), e9201. https://doi.org/10.1371/journal.pone.0009201

Kagawa, C. M. (1958). Blocking Urinary Electrolyte Effects of Desoxycorticosterone with Progesterone in Rats. *Experimental Biology and Medicine*, *99*(3), 705–707. https://doi.org/10.3181/00379727-99-24471

Kakar, S. S., Sellers, J. C., Devor, D. C., Musgrove, L. C., & Neill, J. D. (1992). Angiotensin II type-1 receptor subtype cDNAs: Differential tissue expression and hormonal regulation. *Biochemical and Biophysical Research Communications*, *183*(3), 1090–1096. https://doi.org/10.1016/S0006-291X(05)80302-X

Källberg, D., Vidman, L., & Rydén, P. (2021). Comparison of Methods for Feature Selection in Clustering of High-Dimensional RNA-Sequencing Data to Identify Cancer Subtypes. *Frontiers in Genetics*, *12*, 632620. https://doi.org/10.3389/fgene.2021.632620

Kanduri, C., & Järvelä, I. (2016). *GenRank: An R/Bioconductor package for prioritization of candidate genes* [Preprint]. Bioinformatics. https://doi.org/10.1101/048264

Kaore, S. N., Langade, D. K., Yadav, V. K., Sharma, P., Thawani, V. R., & Sharma, R. (2012). Novel actions of progesterone: What we know today and what will be the scenario in the future? *The Journal of Pharmacy and Pharmacology*, *64*(8), 1040–1062. https://doi.org/10.1111/j.2042-7158.2012.01464.x

Kassambara, A., & Mundt, F. (2020). *Factoextra: Extract and Visualize the Results of Multivariate Data Analyses* (R package version 1.0.7) [Computer software]. Bioconductor. https://CRAN.R-project.org/package=factoextra

Kastner, P., Krust, A., Turcotte, B., Stropp, U., Tora, L., Gronemeyer, H., & Chambon, P. (1990). Two distinct estrogen-regulated promoters generate transcripts encoding the two functionally different human progesterone receptor forms A and B. *The EMBO Journal*, *9*(5), 1603–1614. https://doi.org/10.1002/j.1460-2075.1990.tb08280.x

Kato, S. (2001). Estrogen receptor-mediated cross-talk with growth factor signaling pathways. *Breast Cancer*, *8*(1), 3–9. https://doi.org/10.1007/BF02967472

Katsu, Y., & Baker, M. E. (2021). Aldosterone. In *Handbook of Hormones* (pp. 951–953). Elsevier. https://doi.org/10.1016/B978-0-12-820649-2.00262-X

Katsu, Y., Oka, K., & Baker, M. E. (2018). Evolution of human, chicken, alligator, frog, and zebrafish mineralocorticoid receptors: Allosteric influence on steroid specificity. *Science Signaling*, *11*(537), eaao1520. https://doi.org/10.1126/scisignal.aao1520

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley. https://doi.org/10.1002/9780470316801

Kawamoto, S., Matsumoto, Y., Mizuno, K., Okubo, K., & Matsubara, K. (1996). Expression profiles of active genes in human and mouse livers. *Gene*, *174*(1), 151–158. https://doi.org/10.1016/0378-1119(96)00512-4

Kennewell, P. D. (2007). Major Drug Introductions. In *Comprehensive Medicinal Chemistry II* (pp. 97–249). Elsevier. https://doi.org/10.1016/B0-08-045044-X/00003-1

Khan, J. A., Bellance, C., Guiochon-Mantel, A., Lombès, M., & Loosfelt, H. (2012). Differential Regulation of Breast Cancer-Associated Genes by Progesterone Receptor Isoforms PRA and PRB in a New Bi-Inducible Breast Cancer Cell Line. *PLoS ONE*, *7*(9), e45993. https://doi.org/10.1371/journal.pone.0045993

Khan, S. S., & Ahmad, A. (2004). Cluster center initialization algorithm for K-means clustering. *Pattern Recognition Letters*, *25*(11), 1293–1302. https://doi.org/10.1016/j.patrec.2004.04.007

Khandelwal, N., Simpson, J., Taylor, G., Rafique, S., Whitehouse, A., Hiscox, J., & Stark, L. A. (2011). Nucleolar NF-κB/RelA mediates apoptosis by causing cytoplasmic relocalization of nucleophosmin. *Cell Death & Differentiation*, *18*(12), 1889–1903. https://doi.org/10.1038/cdd.2011.79

Kim, C. H., & Cho, Y. S. (2010). Selection and optimization of MCF-7 cell line for screening selective inhibitors of 11beta-hydroxysteroid dehydrogenase 2. *Cell Biochemistry and Function*, *28*(6), 440–447. https://doi.org/10.1002/cbf.1674

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. https://doi.org/10.1038/nmeth.3317

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kim, D.-U., Hayles, J., Kim, D., Wood, V., Park, H.-O., Won, M., Yoo, H.-S., Duhig, T., Nam, M., Palmer, G., Han, S., Jeffery, L., Baek, S.-T., Lee, H., Shim, Y. S., Lee, M., Kim, L., Heo, K.-S., Noh, E. J., … Hoe, K.-L. (2010). Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. *Nature Biotechnology*, *28*(6), 617–623. https://doi.org/10.1038/nbt.1628

Kim, H., Abd Elmageed, Z. Y., Davis, C., El-Bahrawy, A. H., Naura, A. S., Ekaidi, I., Abdel-Mageed, A. B., & Boulares, A. H. (2014). Correlation between PDZK1, Cdc37, Akt and Breast Cancer Malignancy: The Role of PDZK1 in Cell Growth through Akt Stabilization by Increasing and Interacting with Cdc37. *Molecular Medicine*, *20*(1), 270–279. https://doi.org/10.2119/molmed.2013.00166

Kim, M. Y., Woo, E. M., Chong, Y. T. E., Homenko, D. R., & Kraus, W. L. (2006). Acetylation of Estrogen Receptor α by p300 at Lysines 266 and 268 Enhances the Deoxyribonucleic Acid Binding and Transactivation Activities of the Receptor. *Molecular Endocrinology*, *20*(7), 1479–1493. https://doi.org/10.1210/me.2005-0531

Kim, N., Kang, H., Jo, A., Yoo, S.-A., & Lee, H.-O. (2023). Perspectives on single-nucleus RNA sequencing in different cell types and tissues. *Journal of Pathology and Translational Medicine*, *57*(1), 52–59. https://doi.org/10.4132/jptm.2022.12.19

Kim, S., Kim, A., Shin, J.-Y., & Seo, J.-S. (2020). The tumor immune microenvironmental analysis of 2,033 transcriptomes across 7 cancer types. *Scientific Reports*, *10*(1), 9536. https://doi.org/10.1038/s41598-020-66449-0

Kim, S.-S., Harford, J. B., Pirollo, K. F., & Chang, E. H. (2015). Effective treatment of glioblastoma requires crossing the blood–brain barrier and targeting tumors including cancer stem cells: The promise of nanomedicine. *Biochemical and Biophysical Research Communications*, *468*(3), 485–489. https://doi.org/10.1016/j.bbrc.2015.06.137

Kitange, G. J., Carlson, B. L., Schroeder, M. A., Grogan, P. T., Lamont, J. D., Decker, P. A., Wu, W., James, C. D., & Sarkaria, J. N. (2009). Induction of MGMT expression is associated with temozolomide resistance in glioblastoma xenografts. *Neuro-Oncology*, *11*(3), 281–291. https://doi.org/10.1215/15228517-2008-090

Kleihues, P., Cavenee, W., & editors (Eds.). (2000). *Pathology and genetics of tumours of the nervous system* (2nd ed.). IARC Press.

Klein, R., Conway, D., Parada, L. F., & Barbacid, M. (1990). The trkB tyrosine protein kinase gene codes for a second neurogenic receptor that lacks the catalytic kinase domain. *Cell*, *61*(4), 647–656. https://doi.org/10.1016/0092-8674(90)90476-u

Klinge, C. M. (2001). Estrogen receptor interaction with estrogen response elements. *Nucleic Acids Research*, *29*(14), 2905–2919. https://doi.org/10.1093/nar/29.14.2905

Klipping, C., Marr, J., & Korner, P. (2006). Ovulation inhibition effects of two low-dose oral contraceptive dosing regimens following intentional dosing errors. *Obstetrics & Gynecology*, *107*(4), 49S.

Kobayashi, K., Ehrlich, S. D., Albertini, A., Amati, G., Andersen, K. K., Arnaud, M., Asai, K., Ashikaga, S., Aymerich, S., Bessieres, P., Boland, F., Brignell, S. C., Bron, S., Bunai, K., Chapuis, J., Christiansen, L. C., Danchin, A., Débarbouillé, M., Dervyn, E., … Ogasawara, N. (2003). Essential *Bacillus subtilis* genes. *Proceedings of the National Academy of Sciences*, *100*(8), 4678–4683. https://doi.org/10.1073/pnas.0730515100

Koch, C. M., Chiu, S. F., Akbarpour, M., Bharat, A., Ridge, K. M., Bartom, E. T., & Winter, D. R. (2018). A Beginner's Guide to Analysis of RNA Sequencing Data. *American Journal of Respiratory Cell and Molecular Biology*, *59*(2), 145–157. https://doi.org/10.1165/rcmb.2017-0430TR

Koebele, S. V., Poisson, M. L., Palmer, J. M., Berns-Leone, C., Northup-Smith, S. N., Peña, V. L., Strouse, I. M., Bulen, H. L., Patel, S., Croft, C., & Bimonte-Nelson, H. A. (2022). Evaluating the Cognitive Impacts of Drospirenone, a Spironolactone-Derived Progestin, Independently and in Combination With Ethinyl Estradiol in Ovariectomized Adult Rats. *Frontiers in Neuroscience*, *16*, 885321. https://doi.org/10.3389/fnins.2022.885321

Kohler, B. A., Sherman, R. L., Howlader, N., Jemal, A., Ryerson, A. B., Henry, K. A., Boscoe, F. P., Cronin, K. A., Lake, A., Noone, A.-M., Henley, S. J., Eheman, C. R., Anderson, R. N., & Penberthy, L. (2015). Annual Report to the Nation on the Status of Cancer, 1975-2011, Featuring Incidence of Breast Cancer Subtypes by Race/Ethnicity, Poverty, and State. *JNCI: Journal of the National Cancer Institute*, *107*(6). https://doi.org/10.1093/jnci/djv048

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*(1), 59–69. https://doi.org/10.1007/BF00337288

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, *78*(9), 1464–1480. https://doi.org/10.1109/5.58325

Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, M. D. C., & Yusa, K. (2014). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nature Biotechnology*, *32*(3), 267–273. https://doi.org/10.1038/nbt.2800

Koressaar, T., & Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics*, *23*(10), 1289–1291. https://doi.org/10.1093/bioinformatics/btm091

Koussounadis, A., Langdon, S. P., Um, I. H., Harrison, D. J., & Smith, V. A. (2015). Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. *Scientific Reports*, *5*(1), 10775. https://doi.org/10.1038/srep10775

Kovatcheva, M., Liao, W., Klein, M. E., Robine, N., Geiger, H., Crago, A. M., Dickson, M. A., Tap, W. D., Singer, S., & Koff, A. (2017). ATRX is a regulator of therapy induced senescence in human cells. *Nature Communications*, *8*(1), 386. https://doi.org/10.1038/s41467-017-00540-5

Koyama, K., Myles, K., Smith, R., & Krozowski, Z. (2001). Expression of the 11β-hydroxysteroid dehydrogenase type II enzyme in breast tumors and modulation of activity and cell growth in PMC42 cells. *The Journal of Steroid Biochemistry and Molecular Biology*, *76*(1–5), 153–159. https://doi.org/10.1016/S0960-0760(00)00157-6

Koyama-Nasu, R., Nasu-Nishimura, Y., Todo, T., Ino, Y., Saito, N., Aburatani, H., Funato, K., Echizen, K., Sugano, H., Haruta, R., Matsui, M., Takahashi, R., Manabe, E., Oda, T., & Akiyama, T. (2013). The critical role of cyclin D2 in cell cycle progression and tumorigenicity of glioblastoma stem cells. *Oncogene*, *32*(33), 3840–3845. https://doi.org/10.1038/onc.2012.399

Krattenmacher, R. (2000). Drospirenone: Pharmacology and pharmacokinetics of a unique progestogen. *Contraception*, *62*(1), 29–38. https://doi.org/10.1016/S0010-7824(00)00133-5

Krishnamurthi, K., Verbalis, J. G., Zheng, W., Wu, Z., Clerch, L. B., & Sandberg, K. (1999). Estrogen Regulates Angiotensin AT1 Receptor Expression via Cytosolic Proteins that Bind to the 5′ Leader Sequence of the Receptor mRNA. *Endocrinology*, *140*(11), 5435–5438. https://doi.org/10.1210/endo.140.11.7242

Krozowski, Z. S., & Funder, J. W. (1983). Renal mineralocorticoid receptors and hippocampal corticosterone-binding species have identical intrinsic steroid

specificity. *Proceedings of the National Academy of Sciences*, *80*(19), 6056–6060. https://doi.org/10.1073/pnas.80.19.6056
Krtolica, A., Parrinello, S., Lockett, S., Desprez, P.-Y., & Campisi, J. (2001). Senescent fibroblasts promote epithelial cell growth and tumorigenesis: A link between cancer and aging. *Proceedings of the National Academy of Sciences*, *98*(21), 12072–12077. https://doi.org/10.1073/pnas.211053698
Kubik, J., Humeniuk, E., Adamczuk, G., Madej-Czerwonka, B., & Korga-Plewko, A. (2022). Targeting Energy Metabolism in Cancer Treatment. *International Journal of Molecular Sciences*, *23*(10), 5572. https://doi.org/10.3390/ijms23105572
Kucukural, A., Yukselen, O., Ozata, D. M., Moore, M. J., & Garber, M. (2019). DEBrowser: Interactive differential expression analysis and visualization tool for count data. *BMC Genomics*, *20*(1), 6. https://doi.org/10.1186/s12864-018-5362-x
Kuger, S., Graus, D., Brendtke, R., Günther, N., Katzer, A., Lutyj, P., Polat, B., Chatterjee, M., Sukhorukov, V. L., Flentje, M., & Djuzenova, C. S. (2013). Radiosensitization of Glioblastoma Cell Lines by the Dual PI3K and mTOR Inhibitor NVP-BEZ235 Depends on Drug-Irradiation Schedule. *Translational Oncology*, *6*(2), 169–179. https://doi.org/10.1593/tlo.12364
Kuijjer, M. L., Fagny, M., Marin, A., Quackenbush, J., & Glass, K. (2020). PUMA: PANDA Using MicroRNA Associations. *Bioinformatics*, *36*(18), 4765–4773. https://doi.org/10.1093/bioinformatics/btaa571
Kuijjer, M. L., Tung, M. G., Yuan, G., Quackenbush, J., & Glass, K. (2019). Estimating Sample-Specific Regulatory Networks. *iScience*, *14*, 226–240. https://doi.org/10.1016/j.isci.2019.03.021
Kuilman, T., Michaloglou, C., Mooi, W. J., & Peeper, D. S. (2010). The essence of senescence. *Genes & Development*, *24*(22), 2463–2479. https://doi.org/10.1101/gad.1971610
Kumar, L., & Futschik, M. E. (2007). Mfuzz: A software package for soft clustering of microarray data. *Bioinformation*, *2*(1), 5–7. https://doi.org/10.6026/97320630002005
Kumar, R. S., & Goyal, N. (2021). Estrogens as regulator of hematopoietic stem cell, immune cells and bone biology. *Life Sciences*, *269*, 119091. https://doi.org/10.1016/j.lfs.2021.119091
Kupferman, M. E., Jiffar, T., El-Naggar, A., Yilmaz, T., Zhou, G., Xie, T., Feng, L., Wang, J., Holsinger, F. C., Yu, D., & Myers, J. N. (2010). TrkB induces EMT and has a key role in invasion of head and neck squamous cell carcinoma. *Oncogene*, *29*(14), 2047–2059. https://doi.org/10.1038/onc.2009.486
Kutbay, U. (2018). Partitional Clustering. In H. Pirim (Ed.), *Recent Applications in Data Clustering*. InTech. https://doi.org/10.5772/intechopen.75836
Kwon, Y.-J., Baek, H.-S., Ye, D.-J., Shin, S., Kim, D., & Chun, Y.-J. (2016). CYP1B1 Enhances Cell Proliferation and Metastasis through Induction of EMT and Activation of Wnt/β-Catenin Signaling via Sp1 Upregulation. *PLOS ONE*, *11*(3), e0151598. https://doi.org/10.1371/journal.pone.0151598
Lachmann, A., Torre, D., Keenan, A. B., Jagodnik, K. M., Lee, H. J., Wang, L., Silverstein, M. C., & Ma'ayan, A. (2018). Massive mining of publicly available RNA-seq data from human and mouse. *Nature Communications*, *9*(1), 1366. https://doi.org/10.1038/s41467-018-03751-6
Landau, R. L., & Lugibihl, K. (1958). INHIBITION OF THE SODIUM-RETAINING INFLUENCE OF ALDOSTERONE BY PROGESTERONE*†. *The Journal of*

Clinical Endocrinology & Metabolism, *18*(11), 1237–1245. https://doi.org/10.1210/jcem-18-11-1237

Lange, C. A., Shen, T., & Horwitz, K. B. (2000). Phosphorylation of human progesterone receptors at serine-294 by mitogen-activated protein kinase signals their degradation by the 26S proteasome. *Proceedings of the National Academy of Sciences*, *97*(3), 1032–1037. https://doi.org/10.1073/pnas.97.3.1032

Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*(1), 559. https://doi.org/10.1186/1471-2105-9-559

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25. https://doi.org/10.1186/gb-2009-10-3-r25

Lapointe, J., Li, C., Higgins, J. P., Van De Rijn, M., Bair, E., Montgomery, K., Ferrari, M., Egevad, L., Rayford, W., Bergerheim, U., Ekman, P., DeMarzo, A. M., Tibshirani, R., Botstein, D., Brown, P. O., Brooks, J. D., & Pollack, J. R. (2004). Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proceedings of the National Academy of Sciences*, *101*(3), 811–816. https://doi.org/10.1073/pnas.0304146101

Lappano, R., Pisano, A., & Maggiolini, M. (2014). GPER Function in Breast Cancer: An Overview. *Frontiers in Endocrinology*, *5*, 66. https://doi.org/10.3389/fendo.2014.00066

Lashen, A. G., Toss, M. S., Mongan, N. P., Green, A. R., & Rakha, E. A. (2023). The clinical value of progesterone receptor expression in luminal breast cancer: A study of a large cohort with long-term follow-up. *Cancer*, *129*(8), 1183–1194. https://doi.org/10.1002/cncr.34655

Lau, W. M., Doucet, M., Huang, D., Weber, K. L., & Kominsky, S. L. (2013). CITED2 modulates estrogen receptor transcriptional activity in breast cancer cells. *Biochemical and Biophysical Research Communications*, *437*(2), 261–266. https://doi.org/10.1016/j.bbrc.2013.06.063

Lau, W. M., Weber, K. L., Doucet, M., Chou, Y., Brady, K., Kowalski, J., Tsai, H., Yang, J., & Kominsky, S. L. (2010). Identification of prospective factors promoting osteotropism in breast cancer: A potential role for CITED2. *International Journal of Cancer*, *126*(4), 876–884. https://doi.org/10.1002/ijc.24780

Laurent Gatto, Theo Killian. (2023). *Depmap* (R package version 1.16.0) [Computer software]. Bioconductor. https://doi.org/10.18129/B9.BIOC.DEPMAP

Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, *15*(2). https://doi.org/10.1186/GB-2014-15-2-R29

Lawn, S., Krishna, N., Pisklakova, A., Qu, X., Fenstermacher, D. A., Fournier, M., Vrionis, F. D., Tran, N., Chan, J. A., Kenchappa, R. S., & Forsyth, P. A. (2015). Neurotrophin signaling via TrkB and TrkC receptors promotes the growth of brain tumor-initiating cells. *The Journal of Biological Chemistry*, *290*(6), 3814–3824. https://doi.org/10.1074/jbc.M114.599373

Lee, K.-A., Flores, R. R., Jang, I. H., Saathoff, A., & Robbins, P. D. (2022). Immune Senescence, Immunosenescence and Aging. *Frontiers in Aging*, *3*, 900028. https://doi.org/10.3389/fragi.2022.900028

Lee, K.-H., Hwang, H.-J., Im, Y.-J., Nam, A.-R., Lee, J.-W., & Cho, J.-Y. (2023). New oncogenic functions of LINE1 retroelement as a ceRNA for tumor suppressive

microRNA miR-126 on ENPP5. *PLOS ONE*, *18*(6), e0286814. https://doi.org/10.1371/journal.pone.0286814

Lee Rodgers, J., & Nicewander, W. A. (1988). Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician*, *42*(1), 59–66. https://doi.org/10.1080/00031305.1988.10475524

Leehy, K. A., Truong, T. H., Mauro, L. J., & Lange, C. A. (2018). Progesterone receptors (PR) mediate STAT actions: PR and prolactin receptor signaling crosstalk in breast cancer models. *The Journal of Steroid Biochemistry and Molecular Biology*, *176*, 88–93. https://doi.org/10.1016/j.jsbmb.2017.04.011

Leek, J. T. (2014). svaseq: Removing batch effects and other unwanted noise from sequencing data. *Nucleic Acids Research*, *42*(21), e161. https://doi.org/10.1093/nar/gku864

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., & Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics (Oxford, England)*, *28*(6), 882–883. https://doi.org/10.1093/bioinformatics/bts034

Legler, J. M., Ries, L. A. G., Smith, M. A., Warren, J. L., Heineman, E. F., Kaplan, R. S., & Linet, M. S. (1999). Brain and Other Central Nervous System Cancers: Recent Trends in Incidence and Mortality. *JNCI Journal of the National Cancer Institute*, *91*(16), 1382–1390. https://doi.org/10.1093/jnci/91.16.1382

Lei, F.-J., Chiang, J.-Y., Chang, H.-J., Chen, D.-C., Wang, H.-L., Yang, H.-A., Wei, K.-Y., Huang, Y., Wang, C.-C., Wei, S.-T., & Hsieh, C.-H. (2023). Cellular and exosomal GPx1 are essential for controlling hydrogen peroxide balance and alleviating oxidative stress in hypoxic glioblastoma. *Redox Biology*, *65*, 102831. https://doi.org/10.1016/j.redox.2023.102831

Leinonen, R., Sugawara, H., Shumway, M., & on behalf of the International Nucleotide Sequence Database Collaboration. (2011). The Sequence Read Archive. *Nucleic Acids Research*, *39*(Database), D19–D21. https://doi.org/10.1093/nar/gkq1019

Lemmon, M. A., & Schlessinger, J. (2010). Cell Signaling by Receptor Tyrosine Kinases. *Cell*, *141*(7), 1117–1134. https://doi.org/10.1016/j.cell.2010.06.011

Leng, N., Dawson, J. A., Thomson, J. A., Ruotti, V., Rissman, A. I., Smits, B. M. G., Haag, J. D., Gould, M. N., Stewart, R. M., & Kendziorski, C. (2013). EBSeq: An empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics (Oxford, England)*, *29*(8), 1035–1043. https://doi.org/10.1093/bioinformatics/btt087

Leo, J. C. L., Guo, C., Woon, C. T., Aw, S. E., & Lin, V. C. L. (2004). Glucocorticoid and Mineralocorticoid Cross-Talk with Progesterone Receptor to Induce Focal Adhesion and Growth Inhibition in Breast Cancer Cells. *Endocrinology*, *145*(3), 1314–1321. https://doi.org/10.1210/en.2003-0732

Lewis, D. R., Olex, A. L., Lundy, S. R., Turkett, W. H., Fetrow, J. S., & Muday, G. K. (2013). A Kinetic Analysis of the Auxin Transcriptome Reveals Cell Wall Remodeling Proteins That Modulate Lateral Root Development in *Arabidopsis*. *The Plant Cell*, *25*(9), 3329–3346. https://doi.org/10.1105/tpc.113.114868

Li, D., Wang, L., Fei, F., Hou, Y., Luo, J., Wei-Chen, Zeng, R., Wu, J., Lu, J., Di, G., Ou, Z., Xia, Q., Shen, Z., & Shao, Z. (2006). Identification of breast cancer metastasis-associated proteins in an isogenic tumor metastasis model using two-dimensional gel electrophoresis and liquid chromatography-ion trap-mass spectrometry. *PROTEOMICS*, *6*(11), 3352–3368. https://doi.org/10.1002/pmic.200500617

Li, F., Zhu, W., & Gonzalez, F. J. (2017). Potential role of CYP1B1 in the development and treatment of metabolic diseases. *Pharmacology & Therapeutics*, *178*, 18–30. https://doi.org/10.1016/j.pharmthera.2017.03.007

Li, G.-W., Burkhardt, D., Gross, C., & Weissman, J. S. (2014). Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources. *Cell*, *157*(3), 624–635. https://doi.org/10.1016/j.cell.2014.02.033

Li, H., & Durbin, R. (2009a). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., & Durbin, R. (2009b). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, *26*(5), 589–595. https://doi.org/10.1093/bioinformatics/btp698

Li, J., & Tibshirani, R. (2013). Finding consistent patterns: A nonparametric approach for identifying differential expression in RNA-Seq data. *Statistical Methods in Medical Research*, *22*(5), 519–536. https://doi.org/10.1177/0962280211428386

Li, J., Witten, D. M., Johnstone, I. M., & Tibshirani, R. (2012). Normalization, testing, and false discovery rate estimation for RNA-sequencing data. *Biostatistics (Oxford, England)*, *13*(3), 523–538. https://doi.org/10.1093/biostatistics/kxr031

Li, L., Lee, K.-M., Han, W., Choi, J.-Y., Lee, J.-Y., Kang, G. H., Park, S. K., Noh, D.-Y., Yoo, K.-Y., & Kang, D. (2010). Estrogen and progesterone receptor status affect genome-wide DNA methylation profile in breast cancer. *Human Molecular Genetics*, *19*(21), 4273–4277. https://doi.org/10.1093/hmg/ddq351

Li, L., Lou, Z., & Wang, L. (2011). The role of FKBP5 in cancer aetiology and chemoresistance. *British Journal of Cancer*, *104*(1), 19–23. https://doi.org/10.1038/sj.bjc.6606014

Li, P., Ponnala, L., Gandotra, N., Wang, L., Si, Y., Tausta, S. L., Kebrom, T. H., Provart, N., Patel, R., Myers, C. R., Reidel, E. J., Turgeon, R., Liu, P., Sun, Q., Nelson, T., & Brutnell, T. P. (2010). The developmental dynamics of the maize leaf transcriptome. *Nature Genetics*, *42*(12), 1060–1067. https://doi.org/10.1038/ng.703

Li, S., Ma, L., Song, Y., Zheng, J., Cai, Y., Xu, H., Chen, P., Xu, W., Huang, Y., Qi, T., Li, C., Chu, K., Lan, Y., Xu, L., & Zhou, J. (2021). Effects of hormone replacement therapy on glucose and lipid metabolism in peri- and postmenopausal women with a history of menstrual disorders. *BMC Endocrine Disorders*, *21*(1), 121. https://doi.org/10.1186/s12902-021-00784-9

Li, W., Zhang, Y., Wang, Y., Rong, Z., Liu, C., Miao, H., Chen, H., He, Y., He, W., & Chen, L. (2021). Candidate gene prioritization for chronic obstructive pulmonary disease using expression information in protein–protein interaction networks. *BMC Pulmonary Medicine*, *21*(1), 280. https://doi.org/10.1186/s12890-021-01646-9

Li, X., & Wang, C.-Y. (2021). From bulk, single-cell to spatial RNA sequencing. *International Journal of Oral Science*, *13*(1), 36. https://doi.org/10.1038/s41368-021-00146-0

Li, Z., Chang, Z., Chiao, L. J., Kang, Y., Xia, Q., Zhu, C., Fleming, J. B., Evans, D. B., & Chiao, P. J. (2009). TrkBT1 induces liver metastasis of pancreatic cancer cells by sequestering Rho GDP dissociation inhibitor and promoting RhoA activation.

Cancer Research, *69*(19), 7851–7859. https://doi.org/10.1158/0008-5472.CAN-08-4002

Liang, C., Zhou, J., Wang, Y., Sun, Y., Zhou, J., Shao, L., Zhang, Z., Yan, W., Liu, Z., & Dong, Y. (2024). Essential genes analysis reveals small ribosomal subunit protein eS28 may be a prognostic factor and potential vulnerability in osteosarcoma. *Journal of Bone Oncology*, *44*, 100517. https://doi.org/10.1016/j.jbo.2023.100517

Liang, J., & Shang, Y. (2013). Estrogen and Cancer. *Annual Review of Physiology*, *75*(1), 225–240. https://doi.org/10.1146/annurev-physiol-030212-183708

Liang, S., Chen, Z., Jiang, G., Zhou, Y., Liu, Q., Su, Q., Wei, W., Du, J., & Wang, H. (2017). Activation of GPER suppresses migration and angiogenesis of triple negative breast cancer via inhibition of NF-κB/IL-6 signals. *Cancer Letters*, *386*, 12–23. https://doi.org/10.1016/j.canlet.2016.11.003

Liao, X., Li, M., Hu, K., Wu, F.-X., Gao, X., & Wang, J. (2021). A sensitive repeat identification framework based on short and long reads. *Nucleic Acids Research*, *49*(17), e100–e100. https://doi.org/10.1093/nar/gkab563

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Systems*, *1*(6), 417–425. https://doi.org/10.1016/j.cels.2015.12.004

Lieben, L. (2016). Redefining gene essentiality. *Nature Reviews Genetics*, *17*(2), 66–66. https://doi.org/10.1038/nrg.2015.23

Lin, G. (2023). *reactable: Interactive Data Tables for R* (R package version 0.4.4) [Computer software]. https://CRAN.R-project.org/package=reactable

Lin, J. J.-C., Eppinga, R. D., Warren, K. S., & McCrae, K. R. (2008). Human Tropomyosin Isoforms in the Regulation of Cytoskeleton Functions. In P. Gunning (Ed.), *Tropomyosin* (Vol. 644, pp. 201–222). Springer New York. https://doi.org/10.1007/978-0-387-85766-4_16

Litviakov, N., Tsyganov, M., Larionova, I., Ibragimova, M., Deryusheva, I., Kazantseva, P., Slonimskaya, E., Frolova, I., Choinzonov, E., Cherdyntseva, N., & Kzhyshkowska, J. (2018). Expression of M2 macrophage markers YKL-39 and CCL18 in breast cancer is associated with the effect of neoadjuvant chemotherapy. *Cancer Chemotherapy and Pharmacology*, *82*(1), 99–109. https://doi.org/10.1007/s00280-018-3594-8

Liu, B., Meng, Q., Gao, X., Sun, H., Xu, Z., Wang, Y., & Zhou, H. (2023). Lipid and glucose metabolism in senescence. *Frontiers in Nutrition*, *10*, 1157352. https://doi.org/10.3389/fnut.2023.1157352

Liu, J., Liang, W., Xu, Y., & Zhong, S. (2023). Prognostic value and chemotherapy response prediction of a proliferation essential gene signature in colon cancer. *Bioscience Reports*, *43*(7), BSR20230733. https://doi.org/10.1042/BSR20230733

Liu, J., Xiao, Q., Xiao, J., Niu, C., Li, Y., Zhang, X., Zhou, Z., Shu, G., & Yin, G. (2022). Wnt/β-catenin signalling: Function, biological mechanisms, and therapeutic opportunities. *Signal Transduction and Targeted Therapy*, *7*(1), 3. https://doi.org/10.1038/s41392-021-00762-6

Liu, P., Tang, H., Song, C., Wang, J., Chen, B., Huang, X., Pei, X., & Liu, L. (2018). SOX2 Promotes Cell Proliferation and Metastasis in Triple Negative Breast

Cancer. *Frontiers in Pharmacology*, *9*, 942. https://doi.org/10.3389/fphar.2018.00942

Liu, T., Larionova, I., Litviakov, N., Riabov, V., Zavyalova, M., Tsyganov, M., Buldakov, M., Song, B., Moganti, K., Kazantseva, P., Slonimskaya, E., Kremmer, E., Flatley, A., Klüter, H., Cherdyntseva, N., & Kzhyshkowska, J. (2018). Tumor-associated macrophages in human breast cancer produce new monocyte attracting and pro-angiogenic factor YKL-39 indicative for increased metastasis after neoadjuvant chemotherapy. *OncoImmunology*, *7*(6), e1436922. https://doi.org/10.1080/2162402X.2018.1436922

Liu, X., Krishnan, A., & Mondry, A. (2005). An Entropy-based gene selection method for cancer classification using microarray data. *BMC Bioinformatics*, *6*(1), 76. https://doi.org/10.1186/1471-2105-6-76

Liu, X., Wei, L., Dong, Q., Liu, L., Zhang, M. Q., Xie, Z., & Wang, X. (2019). A large-scale CRISPR screen and identification of essential genes in cellular senescence bypass. *Aging*, *11*(12), 4011–4031. https://doi.org/10.18632/aging.102034

Liu, Y., Gupta, G. D., Barnabas, D. D., Agircan, F. G., Mehmood, S., Wu, D., Coyaud, E., Johnson, C. M., McLaughlin, S. H., Andreeva, A., Freund, S. M. V., Robinson, C. V., Cheung, S. W. T., Raught, B., Pelletier, L., & Van Breugel, M. (2018). Direct binding of CEP85 to STIL ensures robust PLK4 activation and efficient centriole assembly. *Nature Communications*, *9*(1), 1731. https://doi.org/10.1038/s41467-018-04122-x

Liu, Y., Kim, J., Philip, R., Sridhar, V., Chandrashekhar, M., Moffat, J., Van Breugel, M., & Pelletier, L. (2020). Direct interaction between CEP85 and STIL mediates PLk4-driven directed cell migration. *Journal of Cell Science*, jcs.238352. https://doi.org/10.1242/jcs.238352

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods*, *25*(4), 402–408. https://doi.org/10.1006/meth.2001.1262

Lo, C.-C., & Chain, P. S. G. (2014). Rapid evaluation and quality control of next generation sequencing data with FaQCs. *BMC Bioinformatics*, *15*(1), 366. https://doi.org/10.1186/s12859-014-0366-2

Logarajah, S., Hunter, P., Kraman, M., Steele, D., Lakhani, S., Bobrow, L., Venkitaraman, A., & Wagner, S. (2003). BCL-6 is expressed in breast cancer and prevents mammary epithelial differentiation. *Oncogene*, *22*(36), 5572–5578. https://doi.org/10.1038/sj.onc.1206689

López, G. Y., Perry, A., Harding, B., Li, M., & Santi, M. (2019). CDKN2A/B Loss Is Associated with Anaplastic Transformation in a Case of NTRK2 Fusion-positive Pilocytic Astrocytoma. *Neuropathology and Applied Neurobiology*, *45*(2), 174–178. https://doi.org/10.1111/nan.12503

López-Cortés, A., Paz-y-Miño, C., Guerrero, S., Cabrera-Andrade, A., Barigye, S. J., Munteanu, C. R., González-Díaz, H., Pazos, A., Pérez-Castillo, Y., & Tejera, E. (2020). OncoOmics approaches to reveal essential genes in breast cancer: A panoramic view from pathogenesis to precision medicine. *Scientific Reports*, *10*(1), 5285. https://doi.org/10.1038/s41598-020-62279-2

Lorbek, G., Perše, M., Jeruc, J., Juvan, P., Gutierrez-Mariscal, F. M., Lewinska, M., Gebhardt, R., Keber, R., Horvat, S., Björkhem, I., & Rozman, D. (2015). Lessons from Hepatocyte-Specific Cyp51 Knockout Mice: Impaired Cholesterol Synthesis Leads to Oval Cell-Driven Liver Injury. *Scientific Reports*, *5*(1), 8777. https://doi.org/10.1038/srep08777

Louis, D. N., Ohgaki, H., Wiestler, O. D., Cavenee, W. K., Burger, P. C., Jouvet, A., Scheithauer, B. W., & Kleihues, P. (2007). The 2007 WHO Classification of Tumours of the Central Nervous System. *Acta Neuropathologica*, *114*(2), 97–109. https://doi.org/10.1007/s00401-007-0243-4

Louis, D. N., Perry, A., Reifenberger, G., Von Deimling, A., Figarella-Branger, D., Cavenee, W. K., Ohgaki, H., Wiestler, O. D., Kleihues, P., & Ellison, D. W. (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System: A summary. *Acta Neuropathologica*, *131*(6), 803–820. https://doi.org/10.1007/s00401-016-1545-1

Louw-du Toit, R., Perkins, M. S., Hapgood, J. P., & Africander, D. (2017). Comparing the androgenic and estrogenic properties of progestins used in contraception and hormone therapy. *Biochemical and Biophysical Research Communications*, *491*(1), 140–146. https://doi.org/10.1016/j.bbrc.2017.07.063

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lowe, R., Shirley, N., Bleackley, M., Dolan, S., & Shafee, T. (2017). Transcriptomics technologies. *PLOS Computational Biology*, *13*(5), e1005457. https://doi.org/10.1371/journal.pcbi.1005457

Luberg, K., Wong, J., Weickert, C. S., & Timmusk, T. (2010). Human TrkB gene: Novel alternative transcripts, protein isoforms and expression pattern in the prefrontal cerebral cortex during postnatal development. *Journal of Neurochemistry*, *113*(4), 952–964. https://doi.org/10.1111/j.1471-4159.2010.06662.x

Luo, H., Gao, F., & Lin, Y. (2015). Evolutionary conservation analysis between the essential and nonessential genes in bacterial genomes. *Scientific Reports*, *5*(1), 13210. https://doi.org/10.1038/srep13210

Ly, T., Ahmad, Y., Shlien, A., Soroka, D., Mills, A., Emanuele, M. J., Stratton, M. R., & Lamond, A. I. (2014). A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. *eLife*, *3*, e01630. https://doi.org/10.7554/eLife.01630

Ma, S., Pang, C., Song, L., Guo, F., & Sun, H. (2015). Activating transcription factor 3 is overexpressed in human glioma and its knockdown in glioblastoma cells causes growth inhibition both in vitro and in vivo. *International Journal of Molecular Medicine*, *35*(6), 1561–1573. https://doi.org/10.3892/ijmm.2015.2173

MacLeod, G., Bozek, D. A., Rajakulendran, N., Monteiro, V., Ahmadi, M., Steinhart, Z., Kushida, M. M., Yu, H., Coutinho, F. J., Cavalli, F. M. G., Restall, I., Hao, X., Hart, T., Luchman, H. A., Weiss, S., Dirks, P. B., & Angers, S. (2019). Genome-Wide CRISPR-Cas9 Screens Expose Genetic Vulnerabilities and Mechanisms of Temozolomide Sensitivity in Glioblastoma Stem Cells. *Cell Reports*, *27*(3), 971-986.e9. https://doi.org/10.1016/j.celrep.2019.03.047

Macova, M., Armando, I., Zhou, J., Baiardi, G., Tyurmin, D., Larrayoz-Roldan, I. M., & Saavedra, J. M. (2008). Estrogen Reduces Aldosterone, Upregulates Adrenal Angiotensin II AT_2 Receptors and Normalizes Adrenomedullary Fra-2 in Ovariectomized Rats. *Neuroendocrinology*, *88*(4), 276–286. https://doi.org/10.1159/000150977

MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, *1*, 281–297.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2023). *cluster: Cluster Analysis Basics and Extensions* (R package version 2.1.6) [Computer software]. Bioconductor. https://CRAN.R-project.org/package=cluster

Mah, V., Marquez, D., Alavi, M., Maresh, E. L., Zhang, L., Yoon, N., Horvath, S., Bagryanova, L., Fishbein, M. C., Chia, D., Pietras, R., & Goodglick, L. (2011). Expression levels of estrogen receptor beta in conjunction with aromatase predict survival in non-small cell lung cancer. *Lung Cancer*, *74*(2), 318–325. https://doi.org/10.1016/j.lungcan.2011.03.009

Maier, T., Güell, M., & Serrano, L. (2009). Correlation of mRNA and protein in complex biological samples. *FEBS Letters*, *583*(24), 3966–3973. https://doi.org/10.1016/j.febslet.2009.10.036

Mair, B., Tomic, J., Masud, S. N., Tonge, P., Weiss, A., Usaj, M., Tong, A. H. Y., Kwan, J. J., Brown, K. R., Titus, E., Atkins, M., Chan, K. S. K., Munsie, L., Habsid, A., Han, H., Kennedy, M., Cohen, B., Keller, G., & Moffat, J. (2019). Essential Gene Profiles for Human Pluripotent Stem Cells Identify Uncharacterized Genes and Substrate Dependencies. *Cell Reports*, *27*(2), 599-615.e12. https://doi.org/10.1016/j.celrep.2019.02.041

Mal, R., Magner, A., David, J., Datta, J., Vallabhaneni, M., Kassem, M., Manouchehri, J., Willingham, N., Stover, D., Vandeusen, J., Sardesai, S., Williams, N., Wesolowski, R., Lustberg, M., Ganju, R. K., Ramaswamy, B., & Cherian, M. A. (2020). Estrogen Receptor Beta (ERβ): A Ligand Activated Tumor Suppressor. *Frontiers in Oncology*, *10*, 587386. https://doi.org/10.3389/fonc.2020.587386

Mangé, A., Coyaud, E., Desmetz, C., Laurent, E., Béganton, B., Coopman, P., Raught, B., & Solassol, J. (2019). FKBP4 connects mTORC2 and PI3K to activate the PDK1/Akt-dependent cell proliferation signaling in breast cancer. *Theranostics*, *9*(23), 7003–7015. https://doi.org/10.7150/thno.35561

Mangelsdorf, D. J., Thummel, C., Beato, M., Herrlich, P., Schütz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P., & Evans, R. M. (1995). The nuclear receptor superfamily: The second decade. *Cell*, *83*(6), 835–839. https://doi.org/10.1016/0092-8674(95)90199-X

Marchbanks, P. A., McDonald, J. A., Wilson, H. G., Folger, S. G., Mandel, M. G., Daling, J. R., Bernstein, L., Malone, K. E., Ursin, G., Strom, B. L., Norman, S. A., Wingo, P. A., Burkman, R. T., Berlin, J. A., Simon, M. S., Spirtas, R., & Weiss, L. K. (2002). Oral Contraceptives and the Risk of Breast Cancer. *New England Journal of Medicine*, *346*(26), 2025–2032. https://doi.org/10.1056/NEJMoa013202

Marcotte, R., Lacelle, C., & Wang, E. (2004). Senescent fibroblasts resist apoptosis by downregulating caspase-3. *Mechanisms of Ageing and Development*, *125*(10–11), 777–783. https://doi.org/10.1016/j.mad.2004.07.007

Marie, S. K. N., & Shinjo, S. M. O. (2011). Metabolism and Brain Cancer. *Clinics*, *66*, 33–43. https://doi.org/10.1590/S1807-59322011001300005

Marjoribanks, J., Farquhar, C., Roberts, H., & Lethaby, A. (2012). Long term hormone therapy for perimenopausal and postmenopausal women. In The Cochrane Collaboration (Ed.), *Cochrane Database of Systematic Reviews* (p. CD004143.pub4). John Wiley & Sons, Ltd. https://doi.org/10.1002/14651858.CD004143.pub4

Marquez-Garban, D. C., Mah, V., Alavi, M., Maresh, E. L., Chen, H.-W., Bagryanova, L., Horvath, S., Chia, D., Garon, E., Goodglick, L., & Pietras, R. J. (2011). Progesterone and estrogen receptor expression and activity in human non-small

cell lung cancer. *Steroids*, S0039128X11001607. https://doi.org/10.1016/j.steroids.2011.04.015

Martínez, V., Cano, C., & Blanco, A. (2014). ProphNet: A generic prioritization method through propagation of information. *BMC Bioinformatics*, *15*(S1), S5. https://doi.org/10.1186/1471-2105-15-S1-S5

Martinez-Lage, M., Puig-Serra, P., Menendez, P., Torres-Ruiz, R., & Rodriguez-Perales, S. (2018). CRISPR/Cas9 for Cancer Therapy: Hopes and Challenges. *Biomedicines*, *6*(4), 105. https://doi.org/10.3390/biomedicines6040105

Mauvais-Jarvis, F. (2011). Estrogen and androgen receptors: Regulators of fuel homeostasis and emerging targets for diabetes and obesity. *Trends in Endocrinology & Metabolism*, *22*(1), 24–33. https://doi.org/10.1016/j.tem.2010.10.002

Mayca Pozo, F., Geng, X., Tamagno, I., Jackson, M. W., Heimsath, E. G., Hammer, J. A., Cheney, R. E., & Zhang, Y. (2021). MYO10 drives genomic instability and inflammation in cancer. *Science Advances*, *7*(38), eabg6908. https://doi.org/10.1126/sciadv.abg6908

McGowan, E. M., & Clarke, C. L. (1999). Effect of Overexpression of Progesterone Receptor A on Endogenous Progestin-Sensitive Endpoints in Breast Cancer Cells. *Molecular Endocrinology*, *13*(10), 1657–1671. https://doi.org/10.1210/mend.13.10.0356

McGowan, E. M., Saad, S., Bendall, L. J., Bradstock, K. F., & Clarke, C. L. (2004). Effect of Progesterone Receptor A Predominance on Breast Cancer Cell Migration into Bone Marrow Fibroblasts. *Breast Cancer Research and Treatment*, *83*(3), 211–220. https://doi.org/10.1023/B:BREA.0000014041.58977.80

McKay, J. A., Melvin, W. T., Ah-See, A. K., Ewen, S. W. B., Greenlee, W. F., Marcus, C. B., Burke, M. D., & Murray, G. I. (1995). Expression of cytochrome P450 CYP1B1 in breast cancer. *FEBS Letters*, *374*(2), 270–272. https://doi.org/10.1016/0014-5793(95)01126-Y

Mei, S., Qin, Q., Wu, Q., Sun, H., Zheng, R., Zang, C., Zhu, M., Wu, J., Shi, X., Taing, L., Liu, T., Brown, M., Meyer, C. A., & Liu, X. S. (2017). Cistrome Data Browser: A data portal for ChIP-Seq and chromatin accessibility data in human and mouse. *Nucleic Acids Research*, *45*(D1), D658–D662. https://doi.org/10.1093/nar/gkw983

Meyers, R. M., Bryan, J. G., McFarland, J. M., Weir, B. A., Sizemore, A. E., Xu, H., Dharia, N. V., Montgomery, P. G., Cowley, G. S., Pantel, S., Goodale, A., Lee, Y., Ali, L. D., Jiang, G., Lubonja, R., Harrington, W. F., Strickland, M., Wu, T., Hawes, D. C., … Tsherniak, A. (2017). Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nature Genetics*, *49*(12), 1779–1784. https://doi.org/10.1038/ng.3984

Meyyappan, M., Wong, H., Hull, C., & Riabowol, K. T. (1998). Increased expression of cyclin D2 during multiple states of growth arrest in primary and established cells. *Molecular and Cellular Biology*, *18*(6), 3163–3172. https://doi.org/10.1128/MCB.18.6.3163

Michailidou, K., Beesley, J., Lindstrom, S., Canisius, S., Dennis, J., Lush, M. J., Maranian, M. J., Bolla, M. K., Wang, Q., Shah, M., Perkins, B. J., Czene, K., Eriksson, M., Darabi, H., Brand, J. S., Bojesen, S. E., Nordestgaard, B. G., Flyger, H., Nielsen, S. F., … Easton, D. F. (2015). Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci

for breast cancer. *Nature Genetics*, *47*(4), 373–380. https://doi.org/10.1038/ng.3242

Michels, K. A., Pfeiffer, R. M., Brinton, L. A., & Trabert, B. (2018). Modification of the Associations Between Duration of Oral Contraceptive Use and Ovarian, Endometrial, Breast, and Colorectal Cancers. *JAMA Oncology*, *4*(4), 516. https://doi.org/10.1001/jamaoncol.2017.4942

Michna, H., Schneider, M. R., Nishino, Y., & El Etreby, M. F. (1989). Antitumor activity of the antiprogestins ZK 98.299 and RU 38.486 in hormone dependent rat and mouse mammary tumors: Mechanistic studies. *Breast Cancer Research and Treatment*, *14*(3), 275–288. https://doi.org/10.1007/BF01806299

Mijiti, M., Maimaiti, A., Chen, X., Tuersun, M., Dilixiati, M., Dilixiati, Y., Zhu, G., Wu, H., Li, Y., Turhon, M., Abulaiti, A., Maimaitiaili, N., Yiming, N., Kasimu, M., & Wang, Y. (2023). CRISPR-cas9 screening identified lethal genes enriched in Hippo kinase pathway and of predictive significance in primary low-grade glioma. *Molecular Medicine*, *29*(1), 64. https://doi.org/10.1186/s10020-023-00652-3

Miller, L., & Hughes, J. P. (2003). Continuous combination oral contraceptive pills to eliminate withdrawal bleeding: A randomized trial. *Obstetrics & Gynecology*, *101*(4), 653–661. https://doi.org/10.1016/S0029-7844(03)00014-0

Miller, W. L. (2017). Steroidogenesis: Unanswered Questions. *Trends in Endocrinology & Metabolism*, *28*(11), 771–793. https://doi.org/10.1016/j.tem.2017.09.002

Minemura, H., Takagi, K., Sato, A., Takahashi, H., Miki, Y., Shibahara, Y., Watanabe, M., Ishida, T., Sasano, H., & Suzuki, T. (2016). CITED2 in breast carcinoma as a potent prognostic predictor associated with proliferation, migration and chemoresistance. *Cancer Science*, *107*(12), 1898–1908. https://doi.org/10.1111/cas.13081

Mishell, D. R. (2012). Contraception. In *Goldman's Cecil Medicine* (pp. 1552–1555). Elsevier. https://doi.org/10.1016/B978-1-4377-1604-7.00246-3

Mishra, A., Bonello, M., Byron, A., Langdon, S. P., & Sims, A. H. (2020). Evaluation of Gene Expression Data From Cybrids and Tumours Highlights Elevated *NDRG1* -Driven Proliferation in Triple-Negative Breast Cancer. *Breast Cancer: Basic and Clinical Research*, *14*, 117822342093444. https://doi.org/10.1177/1178223420934447

Mizukami, Y. (2010). In Vivo Functions of GPR30/GPER-1, a Membrane Receptor for Estrogen: From Discovery to Functions In Vivo. *Endocrine Journal*, *57*(2), 101–107. https://doi.org/10.1507/endocrj.K09E-332

Moasser, M. M. (2007). The oncogene HER2: Its signaling and transforming functions and its role in human cancer pathogenesis. *Oncogene*, *26*(45), 6469–6487. https://doi.org/10.1038/sj.onc.1210477

Mohammed, A., Hamdan, A., & Homoud, A. (2019). Histopathological profile of brain tumors: A 12-year retrospective study from Madinah, Saudi Arabia. *Asian Journal of Neurosurgery*, *14*(04), 1106–1111. https://doi.org/10.4103/ajns.AJNS_185_19

Mohammed, H., Russell, I. A., Stark, R., Rueda, O. M., Hickey, T. E., Tarulli, G. A., Serandour, A. A., Birrell, S. N., Bruna, A., Saadi, A., Menon, S., Hadfield, J., Pugh, M., Raj, G. V., Brown, G. D., D'Santos, C., Robinson, J. L. L., Silva, G., Launchbury, R., … Carroll, J. S. (2015). Progesterone receptor modulates ERα action in breast cancer. *Nature*, *523*(7560), 313–317. https://doi.org/10.1038/nature14583

Mommsen, T. P., Vijayan, M. M., & Moon, T. W. (1999). Cortisol in teleosts: Dynamics, mechanisms of action, and metabolic regulation. *Reviews in Fish Biology and Fisheries*, *9*(3), 211–268. https://doi.org/10.1023/A:1008924418720

Moody, L., Mantha, S., Chen, H., & Pan, Y.-X. (2019). Computational methods to identify bimodal gene expression and facilitate personalized treatment in cancer patients. *Journal of Biomedical Informatics*, *100*, 100001. https://doi.org/10.1016/j.yjbinx.2018.100001

Mori, H., Saeki, K., Chang, G., Wang, J., Wu, X., Hsu, P.-Y., Kanaya, N., Wang, X., Somlo, G., Nakamura, M., Bild, A., & Chen, S. (2021). Influence of Estrogen Treatment on ESR1+ and ESR1− Cells in ER+ Breast Cancer: Insights from Single-Cell Analysis of Patient-Derived Xenograft Models. *Cancers*, *13*(24), 6375. https://doi.org/10.3390/cancers13246375

Morishige, M., Hashimoto, S., Ogawa, E., Toda, Y., Kotani, H., Hirose, M., Wei, S., Hashimoto, A., Yamada, A., Yano, H., Mazaki, Y., Kodama, H., Nio, Y., Manabe, T., Wada, H., Kobayashi, H., & Sabe, H. (2008). GEP100 links epidermal growth factor receptor signalling to Arf6 activation to induce breast cancer invasion. *Nature Cell Biology*, *10*(1), 85–92. https://doi.org/10.1038/ncb1672

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, *5*(7), 621–628. https://doi.org/10.1038/nmeth.1226

Mote, P. A., Gompel, A., Howe, C., Hilton, H. N., Sestak, I., Cuzick, J., Dowsett, M., Hugol, D., Forgez, P., Byth, K., Graham, J. D., & Clarke, C. L. (2015). Progesterone receptor A predominance is a discriminator of benefit from endocrine therapy in the ATAC trial. *Breast Cancer Research and Treatment*, *151*(2), 309–318. https://doi.org/10.1007/s10549-015-3397-0

Mote, P. A., Graham, J. D., & Clarke, C. L. (2007). Progesterone receptor isoforms in normal and malignant breast. *Ernst Schering Foundation Symposium Proceedings*, *1*, 77–107.

Moustakas, I. I., Katsarou, A., Legaki, A.-I., Pyrina, I., Ntostoglou, K., Papatheodoridi, A.-M., Gercken, B., Pateras, I. S., Gorgoulis, V. G., Koutsilieris, M., Chavakis, T., & Chatzigeorgiou, A. (2021). Hepatic Senescence Accompanies the Development of NAFLD in Non-Aged Mice Independently of Obesity. *International Journal of Molecular Sciences*, *22*(7), 3446. https://doi.org/10.3390/ijms22073446

Mueck, A. O., Ruan, X., Seeger, H., Fehm, T., & Neubauer, H. (2014). Genomic and non-genomic actions of progestogens in the breast. *The Journal of Steroid Biochemistry and Molecular Biology*, *142*, 62–67. https://doi.org/10.1016/j.jsbmb.2013.08.011

Muhn, P., Krattenmacher, R., Beier, S., Elger, W., & Schillinger, E. (1995a). Drospirenone: A novel progestogen with antimineralocorticoid and antiandrogenic activity. *Contraception*, *51*(2), 99–110. https://doi.org/10.1016/0010-7824(94)00015-O

Muhn, P., Krattenmacher, R., Beier, S., Elger, W., & Schillinger, E. (1995b). Drospirenone: A novel progestogen with antimineralocorticoid and antiandrogenic activity: Pharmacological characterization in animal models. *Contraception*, *51*(2), 99–110. https://doi.org/10.1016/0010-7824(94)00015-O

Muir, W. W. (2015). Drug Interactions, Analgesic Protocols and Their Consequences, and Analgesic Drug Antagonism. In *Handbook of Veterinary Pain Management* (pp. 335–355). Elsevier. https://doi.org/10.1016/B978-0-323-08935-7.00016-8

Mukherjee, A. G., Gopalakrishnan, A. V., Jayaraj, R., Ganesan, R., Renu, K., Vellingiri, B., Dey, A., & Parveen, M. (2023). Recent advances in understanding brain cancer metabolomics: A review. *Medical Oncology*, *40*(8), 220. https://doi.org/10.1007/s12032-023-02109-3

Muraki, N., Yamada, M., Doki, H., Nakai, R., Komeda, K., Goto, D., Kawabe, N., Matsuoka, K., Matsushima, M., Kawabe, T., Tanaka, I., Morise, M., Shay, J. W., Minna, J. D., & Sato, M. (2022). Resistance to mutant KRAS-induced senescence in an hTERT/Cdk4-immortalized normal human bronchial epithelial cell line. *Experimental Cell Research*, *414*(1), 113053. https://doi.org/10.1016/j.yexcr.2022.113053

Murphy, E. (2011). Estrogen Signaling and Cardiovascular Disease. *Circulation Research*, *109*(6), 687–696. https://doi.org/10.1161/CIRCRESAHA.110.236687

Murtagh, F., & Legendre, P. (2014). Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *Journal of Classification*, *31*(3), 274–295. https://doi.org/10.1007/s00357-014-9161-z

Muscat, G. E. O., Eriksson, N. A., Byth, K., Loi, S., Graham, D., Jindal, S., Davis, M. J., Clyne, C., Funder, J. W., Simpson, E. R., Ragan, M. A., Kuczek, E., Fuller, P. J., Tilley, W. D., Leedman, P. J., & Clarke, C. L. (2013). Research Resource: Nuclear Receptors as Transcriptome: Discriminant and Prognostic Value in Breast Cancer. *Molecular Endocrinology*, *27*(2), 350–365. https://doi.org/10.1210/me.2012-1265

Musgrove, E. A., & Sutherland, R. L. (2009). Biological determinants of endocrine resistance in breast cancer. *Nature Reviews. Cancer*, *9*(9), 631–643. https://doi.org/10.1038/nrc2713

Muthu Lakshmi Bavithra, C., Murugan, M., Pavithran, S., & Naveena, K. (2023). Enthralling genetic regulatory mechanisms meddling insecticide resistance development in insects: Role of transcriptional and post-transcriptional events. *Frontiers in Molecular Biosciences*, *10*, 1257859. https://doi.org/10.3389/fmolb.2023.1257859

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., & Snyder, M. (2008). The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. *Science*, *320*(5881), 1344–1349. https://doi.org/10.1126/science.1158441

Nakada, M., Kita, D., Watanabe, T., Hayashi, Y., Teng, L., Pyko, I. V., & Hamada, J.-I. (2011). Aberrant Signaling Pathways in Glioma. *Cancers*, *3*(3), 3242–3278. https://doi.org/10.3390/cancers3033242

Nakagawara, A., Liu, X. G., Ikegaki, N., White, P. S., Yamashiro, D. J., Nycum, L. M., Biegel, J. A., & Brodeur, G. M. (1995). Cloning and chromosomal localization of the human TRK-B tyrosine kinase receptor gene (NTRK2). *Genomics*, *25*(2), 538–546. https://doi.org/10.1016/0888-7543(95)80055-q

Nayak, T., Wang, L.-J., Ning, M., Rubannelsonkumar, G., Jin, E., Zheng, S., Houghton, P. J., Huang, Y., Chiu, Y.-C., & Chen, Y. (2023). DepLink: An R Shiny app to systematically link genetic and pharmacologic dependencies of cancer. *Bioinformatics Advances*, *3*(1), vbad076. https://doi.org/10.1093/bioadv/vbad076

Need, E. F., Selth, L. A., Trotta, A. P., Leach, D. A., Giorgio, L., O'Loughlin, M. A., Smith, E., Gill, P. G., Ingman, W. V., Graham, J. D., & Buchanan, G. (2015).

The unique transcriptional response produced by concurrent estrogen and progesterone treatment in breast cancer cells results in upregulation of growth factor pathways and switching from a Luminal A to a Basal-like subtype. *BMC Cancer*, *15*(1), 791. https://doi.org/10.1186/s12885-015-1819-3

Nesterov, V., Dahlmann, A., Krueger, B., Bertog, M., Loffing, J., & Korbmacher, C. (2012). Aldosterone-dependent and -independent regulation of the epithelial sodium channel (ENaC) in mouse distal nephron. *American Journal of Physiology-Renal Physiology*, *303*(9), F1289–F1299. https://doi.org/10.1152/ajprenal.00247.2012

Network, T. C. G. A. R. (2013). Corrigendum: Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, *494*(7438), 506. https://doi.org/10.1038/nature11903

Ni, J., Xie, S., Ramkissoon, S. H., Luu, V., Sun, Y., Bandopadhayay, P., Beroukhim, R., Roberts, T. M., Stiles, C. D., Segal, R. A., Ligon, K. L., Hahn, W. C., & Zhao, J. J. (2017). Tyrosine receptor kinase B is a drug target in astrocytomas. *Neuro-Oncology*, *19*(1), 22–30. https://doi.org/10.1093/neuonc/now139

NIH State-of-the-Science Panel*. (2005). National Institutes of Health State-of-the-Science Conference Statement: Management of Menopause-Related Symptoms. *Annals of Internal Medicine*, *142*(12_Part_1), 1003. https://doi.org/10.7326/0003-4819-142-12_Part_1-200506210-00117

Nobusawa, S., Watanabe, T., Kleihues, P., & Ohgaki, H. (2009). *IDH1* Mutations as Molecular Signature and Predictive Factor of Secondary Glioblastomas. *Clinical Cancer Research*, *15*(19), 6002–6007. https://doi.org/10.1158/1078-0432.CCR-09-0715

Nusinow, D. P., Szpyt, J., Ghandi, M., Rose, C. M., McDonald, E. R., Kalocsay, M., Jané-Valbuena, J., Gelfand, E., Schweppe, D. K., Jedrychowski, M., Golji, J., Porter, D. A., Rejtar, T., Wang, Y. K., Kryukov, G. V., Stegmeier, F., Erickson, B. K., Garraway, L. A., Sellers, W. R., & Gygi, S. P. (2020). Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell*, *180*(2), 387-402.e16. https://doi.org/10.1016/j.cell.2019.12.023

Oelkers, W. (2004). Drospirenone, a progestogen with antimineralocorticoid properties: A short review. *Molecular and Cellular Endocrinology*, *217*(1–2), 255–261. https://doi.org/10.1016/j.mce.2003.10.030

Oelkers, W., Berger, V., Bolik, A., Bähr, V., Hazard, B., Beier, S., Elger, W., & Heithecker, A. (1991). Dihydrospirorenone, a New Progestogen with Antimineralocorticoid Activity: Effects on Ovulation, Electrolyte Excretion, and the Renin-Aldosterone System in Normal Women. *The Journal of Clinical Endocrinology & Metabolism*, *73*(4), 837–842. https://doi.org/10.1210/jcem-73-4-837

Oelkers, W., Schöneshöfer, M., & Blümel, A. (1974). Effects of Progesterone and Four Synthetic Progestagens on Sodium Balance and the Renin-Aldosterone System in Man*†. *The Journal of Clinical Endocrinology & Metabolism*, *39*(5), 882–890. https://doi.org/10.1210/jcem-39-5-882

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., & Kanehisa, M. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, *27*(1), 29–34. https://doi.org/10.1093/nar/27.1.29

Ohira, K., Homma, K. J., Hirai, H., Nakamura, S., & Hayashi, M. (2006). TrkB-T1 regulates the RhoA signaling and actin cytoskeleton in glioma cells. *Biochemical and Biophysical Research Communications*, *342*(3), 867–874. https://doi.org/10.1016/j.bbrc.2006.02.033

Ohira, K., Kumanogoh, H., Sahara, Y., Homma, K. J., Hirai, H., Nakamura, S., & Hayashi, M. (2005). A truncated tropomyosin-related kinase B receptor, T1, regulates glial cell morphology via Rho GDP dissociation inhibitor 1. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *25*(6), 1343–1353. https://doi.org/10.1523/JNEUROSCI.4436-04.2005

Ohta, K., Chiba, Y., Kaise, A., & Endo, Y. (2015). Structure–activity relationship study of diphenylamine-based estrogen receptor (ER) antagonists. *Bioorganic & Medicinal Chemistry*, *23*(4), 861–867. https://doi.org/10.1016/j.bmc.2014.12.022

Ola, B., Afnan, M., Papaioannou, S., Sharif, K., Björndahl, L., & Coomarasamy, A. (2003). Accuracy of sperm-cervical mucus penetration tests in evaluating sperm motility in semen: A systematic quantitative review. *Human Reproduction (Oxford, England)*, *18*(5), 1037–1046. https://doi.org/10.1093/humrep/deg209

Olde, B., & Leeb-Lundberg, L. M. F. (2009). GPR30/GPER1: Searching for a role in estrogen physiology. *Trends in Endocrinology & Metabolism*, *20*(8), 409–416. https://doi.org/10.1016/j.tem.2009.04.006

Olopade, O. I., Adeyanju, M. O., Safa, A. R., Hagos, F., Mick, R., Thompson, C. B., & Recant, W. M. (1997). Overexpression of BCL-x protein in primary breast cancer is associated with high tumor grade and nodal metastases. *The Cancer Journal from Scientific American*, *3*(4), 230–237.

Oltean, S., & Bates, D. O. (2014). Hallmarks of alternative splicing in cancer. *Oncogene*, *33*(46), 5311–5318. https://doi.org/10.1038/onc.2013.533

Önskog, J., Freyhult, E., Landfors, M., Rydén, P., & Hvidsten, T. R. (2011). Classification of microarrays; synergistic effects between normalization, gene selection and machine learning. *BMC Bioinformatics*, *12*(1), 390. https://doi.org/10.1186/1471-2105-12-390

Ørntoft, T. F., Thykjaer, T., Waldman, F. M., Wolf, H., & Celis, J. E. (2002). Genome-wide Study of Gene Copy Numbers, Transcripts, and Protein Levels in Pairs of Non-invasive and Invasive Human Transitional Cell Carcinomas. *Molecular & Cellular Proteomics*, *1*(1), 37–45. https://doi.org/10.1074/mcp.M100019-MCP200

Ortega, H., Li, H., Suruki, R., Albers, F., Gordon, D., & Yancey, S. (2014). Cluster Analysis and Characterization of Response to Mepolizumab. A Step Closer to Personalized Medicine for Patients with Severe Asthma. *Annals of the American Thoracic Society*, *11*(7), 1011–1017. https://doi.org/10.1513/AnnalsATS.201312-454OC

Osabe, T., Shimizu, K., & Kadota, K. (2021). Differential expression analysis using a model-based gene clustering algorithm for RNA-seq data. *BMC Bioinformatics*, *22*(1), 511. https://doi.org/10.1186/s12859-021-04438-4

Oshlack, A., & Wakefield, M. J. (2009). Transcript length bias in RNA-seq data confounds systems biology. *Biology Direct*, *4*, 14. https://doi.org/10.1186/1745-6150-4-14

Oytam, Y., Sobhanmanesh, F., Duesing, K., Bowden, J. C., Osmond-McLeod, M., & Ross, J. (2016). Risk-conscious correction of batch effects: Maximising information extraction from high-throughput genomic datasets. *BMC Bioinformatics*, *17*(1), 332. https://doi.org/10.1186/s12859-016-1212-5

Oz, N., Vayndorf, E. M., Tsuchiya, M., McLean, S., Turcios-Hernandez, L., Pitt, J. N., Blue, B. W., Muir, M., Kiflezghi, M. G., Tyshkovskiy, A., Mendenhall, A., Kaeberlein, M., & Kaya, A. (2022). Evidence that conserved essential genes are

enriched for pro-longevity factors. *GeroScience*, *44*(4), 1995–2006. https://doi.org/10.1007/s11357-022-00604-5

Pacini, C., Dempster, J. M., Boyle, I., Gonçalves, E., Najgebauer, H., Karakoc, E., Van Der Meer, D., Barthorpe, A., Lightfoot, H., Jaaks, P., McFarland, J. M., Garnett, M. J., Tsherniak, A., & Iorio, F. (2020). *Integrated cross-study datasets of genetic dependencies in cancer* [Preprint]. Bioinformatics. https://doi.org/10.1101/2020.05.22.110247

Palacios, S., Foidart, J.-M., & Genazzani, A. R. (2006). Advances in hormone replacement therapy with drospirenone, a unique progestogen with aldosterone receptor antagonism. *Maturitas*, *55*(4), 297–307. https://doi.org/10.1016/j.maturitas.2006.07.009

Palmisano, B. T., Zhu, L., Eckel, R. H., & Stafford, J. M. (2018). Sex differences in lipid and lipoprotein metabolism. *Molecular Metabolism*, *15*, 45–55. https://doi.org/10.1016/j.molmet.2018.05.008

Pan, C. C., Maeso-Díaz, R., Lewis, T. R., Xiang, K., Tan, L., Liang, Y., Wang, L., Yang, F., Yin, T., Wang, C., Du, K., Huang, D., Oh, S. H., Wang, E., Lim, B. J. W., Chong, M., Alexander, P. B., Yao, X., Arshavsky, V. Y., … Wang, X.-F. (2023). Antagonizing the irreversible thrombomodulin-initiated proteolytic signaling alleviates age-related liver fibrosis via senescent cell killing. *Cell Research*, *33*(7), 516–532. https://doi.org/10.1038/s41422-023-00820-4

Parker, C. G., Galmozzi, A., Wang, Y., Correia, B. E., Sasaki, K., Joslyn, C. M., Kim, A. S., Cavallaro, C. L., Lawrence, R. M., Johnson, S. R., Narvaiza, I., Saez, E., & Cravatt, B. F. (2017). Ligand and Target Discovery by Fragment-Based Screening in Human Cells. *Cell*, *168*(3), 527-541.e29. https://doi.org/10.1016/j.cell.2016.12.029

Parodi, A., Rudzińska, M., Deviatkin, A., Soond, S., Baldin, A., & Zamyatnin, A. (2019). Established and Emerging Strategies for Drug Delivery Across the Blood-Brain Barrier in Brain Cancer. *Pharmaceutics*, *11*(5), 245. https://doi.org/10.3390/pharmaceutics11050245

Parsania, C., Chen, R., Sethiya, P., Miao, Z., Dong, L., & Wong, K. H. (2023). FungiExpresZ: An intuitive package for fungal gene expression data analysis, visualization and discovery. *Briefings in Bioinformatics*, *24*(2), bbad051. https://doi.org/10.1093/bib/bbad051

Parsey, K. S., & Pong, A. (2000). An open-label, multicenter study to evaluate Yasmin, a low-dose combination oral contraceptive containing drospirenone, a new progestogen. *Contraception*, *61*(2), 105–111. https://doi.org/10.1016/S0010-7824(00)00083-4

Parsyan, A., Cruickshank, J., Hodgson, K., Wakeham, D., Pellizzari, S., Bhat, V., & Cescon, D. W. (2021). Anticancer effects of radiation therapy combined with Polo-Like Kinase 4 (PLK4) inhibitor CFI-400945 in triple negative breast cancer. *The Breast*, *58*, 6–9. https://doi.org/10.1016/j.breast.2021.03.011

Patibandla, S., Heaton, J., & Kyaw, H. (2023). Spironolactone. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK554421/

Paton, D.M. (2020). Drospirenone: A novel progestogen used as an oral contraceptivewithout an estrogen component. *Drugs of Today*, *56*(5), 321. https://doi.org/10.1358/dot.2020.56.5.3134788

Pattwell, S. S., Arora, S., Cimino, P. J., Ozawa, T., Szulzewsky, F., Hoellerbauer, P., Bonifert, T., Hoffstrom, B. G., Boiani, N. E., Bolouri, H., Correnti, C. E., Oldrini, B., Silber, J. R., Squatrito, M., Paddison, P. J., & Holland, E. C. (2020). A kinase-deficient NTRK2 splice variant predominates in glioma and amplifies

several oncogenic signaling pathways. *Nature Communications*, *11*(1), 2977. https://doi.org/10.1038/s41467-020-16786-5

Patwardhan, R. S., Rai, A., Sharma, D., Sandur, S. K., & Patwardhan, S. (2024). Txnrd1 as a prognosticator for recurrence, metastasis and response to neoadjuvant chemotherapy and radiotherapy in breast cancer patients. *Heliyon*, e27011. https://doi.org/10.1016/j.heliyon.2024.e27011

Payvar, F., DeFranco, D., Firestone, G. L., Edgar, B., Wrange, Ö., Okret, S., Gustafsson, J.-Å., & Yamamoto, K. R. (1983). Sequence-specific binding of glucocorticoid receptor to MTV DNA at sites within and upstream of the transcribed region. *Cell*, *35*(2), 381–392. https://doi.org/10.1016/0092-8674(83)90171-X

Pearson, W. R., & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences*, *85*(8), 2444–2448. https://doi.org/10.1073/pnas.85.8.2444

Pedroza, D. A., Subramani, R., & Lakshmanaswamy, R. (2020). Classical and Non-Classical Progesterone Signaling in Breast Cancers. *Cancers*, *12*(9), 2440. https://doi.org/10.3390/cancers12092440

Pei, H., Li, L., Fridley, B. L., Jenkins, G. D., Kalari, K. R., Lingle, W., Petersen, G., Lou, Z., & Wang, L. (2009). FKBP51 Affects Cancer Cell Response to Chemotherapy by Negatively Regulating Akt. *Cancer Cell*, *16*(3), 259–266. https://doi.org/10.1016/j.ccr.2009.07.016

Pekmezci, M., & Perry, A. (2013). Neuropathology of brain metastases. *Surgical Neurology International*, *4*(5), 245. https://doi.org/10.4103/2152-7806.111302

Perampalam, P., & Dick, F. A. (2020). BEAVR: A browser-based tool for the exploration and visualization of RNA-seq data. *BMC Bioinformatics*, *21*(1), 221. https://doi.org/10.1186/s12859-020-03549-8

Perkins, M. S., Louw-du Toit, R., & Africander, D. (2018). Hormone therapy and breast cancer: Emerging steroid receptor mechanisms. *Journal of Molecular Endocrinology*, *61*(4), R133–R160. https://doi.org/10.1530/JME-18-0094

Perkins, M. S., Louw-du Toit, R., Jackson, H., Simons, M., & Africander, D. (2022). Upregulation of an estrogen receptor-regulated gene by first generation progestins requires both the progesterone receptor and estrogen receptor alpha. *Frontiers in Endocrinology*, *13*, 959396. https://doi.org/10.3389/fendo.2022.959396

Perrier, V., & Meyer, F. (2020). *fresh: Create Custom "Bootstrap" Themes to Use in "Shiny"* (R package version 0.2.0) [Computer software]. https://CRAN.R-project.org/package=fresh

Perrier, V., Meyer, F., & Granjon, D. (2023). *shinyWidgets: Custom Inputs Widgets for Shiny* (R package version 0.7.6) [Computer software]. https://CRAN.R-project.org/package=shinyWidgets

Peters, V. A., Joesting, J. J., & Freund, G. G. (2013). IL-1 receptor 2 (IL-1R2) and its role in immune regulation. *Brain, Behavior, and Immunity*, *32*, 1–8. https://doi.org/10.1016/j.bbi.2012.11.006

Piva, M., Domenici, G., Iriondo, O., Rábano, M., Simões, B. M., Comaills, V., Barredo, I., López-Ruiz, J. A., Zabalza, I., Kypta, R., & Vivanco, M. D. M. (2014). Sox2 promotes tamoxifen resistance in breast cancer cells. *EMBO Molecular Medicine*, *6*(1), 66–79. https://doi.org/10.1002/emmm.201303411

Plaimas, K., Eils, R., & König, R. (2010). Identifying essential genes in bacterial metabolic networks with machine learning methods. *BMC Systems Biology*, *4*(1), 56. https://doi.org/10.1186/1752-0509-4-56

Podfigurna, A., Meczekalski, B., Petraglia, F., & Luisi, S. (2020). Clinical, hormonal and metabolic parameters in women with PCOS with different combined oral contraceptives (containing chlormadinone acetate versus drospirenone). *Journal of Endocrinological Investigation*, *43*(4), 483–492. https://doi.org/10.1007/s40618-019-01133-3

Pollow, K., Juchem, M., Elger, W., Jacobi, N., Hoffmann, G., & Möbus, V. (1992). Dihydrospirorenone (ZK30595): A novel synthetic progestagen-characterization of binding to different receptor proteins. *Contraception*, *46*(6), 561–574. https://doi.org/10.1016/0010-7824(92)90121-9

Popescu, V.-B., Sánchez-Martín, J. Á., Schacherer, D., Safadoust, S., Majidi, N., Andronescu, A., Nedea, A., Ion, D., Mititelu, E., Czeizler, E., & Petre, I. (2021). NetControl4BioMed: A web-based platform for controllability analysis of protein–protein interaction networks. *Bioinformatics*, *37*(21), 3976–3978. https://doi.org/10.1093/bioinformatics/btab570

Pratap, U. P., Tidwell, M., Balinda, H. U., Clanton, N. A., Yang, X., Viswanadhapalli, S., Sareddy, G. R., Liang, D., Xie, H., Chen, Y., Lai, Z., Tekmal, R. R., McHardy, S. F., Brenner, A. J., & Vadlamudi, R. K. (2023). Preclinical Development of Brain Permeable ERβ Agonist for the Treatment of Glioblastoma. *Molecular Cancer Therapeutics*, *22*(11), 1248–1260. https://doi.org/10.1158/1535-7163.MCT-23-0031

Prince, E. W., Balakrishnan, I., Shah, M., Mulcahy Levy, J. M., Griesinger, A. M., Alimova, I., Harris, P. S., Birks, D. K., Donson, A. M., Davidson, N., Remke, M., Taylor, M. D., Handler, M. H., Foreman, N. K., Venkataraman, S., & Vibhakar, R. (2016). Checkpoint kinase 1 expression is an adverse prognostic marker and therapeutic target in MYC-driven medulloblastoma. *Oncotarget*, *7*(33), 53881–53894. https://doi.org/10.18632/oncotarget.10692

Prolo, L. M., Li, A., Owen, S. F., Parker, J. J., Foshay, K., Nitta, R. T., Morgens, D. W., Bolin, S., Wilson, C. M., Vega L, J. C. M., Luo, E. J., Nwagbo, G., Waziri, A., Li, G., Reimer, R. J., Bassik, M. C., & Grant, G. A. (2019). Targeted genomic CRISPR-Cas9 screen identifies MAP4K4 as essential for glioblastoma invasion. *Scientific Reports*, *9*(1), 14020. https://doi.org/10.1038/s41598-019-50160-w

Prossnitz, E. R., & Maggiolini, M. (2009). Mechanisms of estrogen signaling and gene expression via GPR30. *Molecular and Cellular Endocrinology*, *308*(1–2), 32–38. https://doi.org/10.1016/j.mce.2009.03.026

Pucci, S., & Bettuzzi, S. (2009). Chapter 3 The Shifting Balance Between CLU Forms During Tumor Progression. In *Advances in Cancer Research* (Vol. 104, pp. 25–32). Elsevier. https://doi.org/10.1016/S0065-230X(09)04003-2

Quatredeniers, M., Serafin, A. S., Benmerah, A., Rausell, A., Saunier, S., & Viau, A. (2023). Meta-analysis of single-cell and single-nucleus transcriptomics reveals kidney cell type consensus signatures. *Scientific Data*, *10*(1), 361. https://doi.org/10.1038/s41597-023-02209-9

Queisser, N., Oteiza, P. I., Link, S., Hey, V., Stopper, H., & Schupp, N. (2014). Aldosterone activates transcription factor Nrf2 in kidney cells both in vitro and in vivo. *Antioxidants & Redox Signaling*, *21*(15), 2126–2142. https://doi.org/10.1089/ars.2013.5565

Quinkler, M., Meyer, B., Bumke-Vogt, C., Grossmann, C., Gruber, U., Oelkers, W., Diederich, S., & Bahr, V. (2002). Agonistic and antagonistic properties of progesterone metabolites at the human mineralocorticoid receptor. *European Journal of Endocrinology*, 789–799. https://doi.org/10.1530/eje.0.1460789

R Development Core Team. (2023). *R: A language and environment for statistical computing. R Foundation for Statistical Computing,* [Computer software]. https://www.R-project.org

Rabbitt, E. H., Gittoes, N. J. L., Stewart, P. M., & Hewison, M. (2003). 11beta-hydroxysteroid dehydrogenases, cell proliferation and malignancy. *The Journal of Steroid Biochemistry and Molecular Biology*, *85*(2–5), 415–421. https://doi.org/10.1016/s0960-0760(03)00224-3

Randall, E. C., Lopez, B. G. C., Peng, S., Regan, M. S., Abdelmoula, W. M., Basu, S. S., Santagata, S., Yoon, H., Haigis, M. C., Agar, J. N., Tran, N. L., Elmquist, W. F., White, F. M., Sarkaria, J. N., & Agar, N. Y. R. (2020). Localized Metabolomic Gradients in Patient-Derived Xenograft Models of Glioblastoma. *Cancer Research*, *80*(6), 1258–1267. https://doi.org/10.1158/0008-5472.CAN-19-0638

Rangel, N., Fortunati, N., Osella-Abate, S., Annaratone, L., Isella, C., Catalano, M. G., Rinella, L., Metovic, J., Boldorini, R., Balmativola, D., Ferrando, P., Marano, F., Cassoni, P., Sapino, A., & Castellano, I. (2018). FOXA1 and AR in invasive breast cancer: New findings on their co-expression and impact on prognosis in ER-positive patients. *BMC Cancer*, *18*(1), 703. https://doi.org/10.1186/s12885-018-4624-y

Rankin, G. O. (2008). Aldosterone Antagonists. In *xPharm: The Comprehensive Pharmacology Reference* (pp. 1–5). Elsevier. https://doi.org/10.1016/B978-008055232-3.63843-7

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., & Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, *159*(7), 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021

Rapkin, A. J., & Winer, S. A. (2007). Drospirenone: A novel progestin. *Expert Opinion on Pharmacotherapy*, *8*(7), 989–999. https://doi.org/10.1517/14656566.8.7.989

Reichardt, L. F. (2006). Neurotrophin-regulated signalling pathways. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *361*(1473), 1545–1564. https://doi.org/10.1098/rstb.2006.1894

Ren, Z., Wang, W., & Li, J. (2016). Identifying molecular subtypes in human colon cancer using gene expression and DNA methylation microarray data. *International Journal of Oncology*, *48*(2), 690–702. https://doi.org/10.3892/ijo.2015.3263

Revankar, C. M., Cimino, D. F., Sklar, L. A., Arterburn, J. B., & Prossnitz, E. R. (2005). A Transmembrane Intracellular Estrogen Receptor Mediates Rapid Cell Signaling. *Science*, *307*(5715), 1625–1630. https://doi.org/10.1126/science.1106943

Reyes, A. L. P., Silva, T. C., Coetzee, S. G., Plummer, J. T., Davis, B. D., Chen, S., Hazelett, D. J., Lawrenson, K., Berman, B. P., Gayther, S. A., & Jones, M. R. (2019). GENAVi: A shiny web application for gene expression normalization, analysis and visualization. *BMC Genomics*, *20*(1), 745. https://doi.org/10.1186/s12864-019-6073-7

Rigiracciolo, D. C., Scarpelli, A., Lappano, R., Pisano, A., Santolla, M. F., Avino, S., De Marco, P., Bussolati, B., Maggiolini, M., & De Francesco, E. M. (2016). GPER is involved in the stimulatory effects of aldosterone in breast cancer cells and breast tumor-derived endothelial cells. *Oncotarget*, *7*(1), 94–111. https://doi.org/10.18632/oncotarget.6475

Risso, D., Schwartz, K., Sherlock, G., & Dudoit, S. (2011). GC-content normalization for RNA-Seq data. *BMC Bioinformatics*, *12*, 480. https://doi.org/10.1186/1471-2105-12-480

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47–e47. https://doi.org/10.1093/nar/gkv007

Rizzo, M., & Szekely, G. (2022). *Energy: E-Statistics: Multivariate Inference via the Energy of Data* (R package version 1.7-11) [Computer software]. https://CRAN.R-project.org/package=energy

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), 24–26. https://doi.org/10.1038/nbt.1754

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, *11*(3), R25. https://doi.org/10.1186/gb-2010-11-3-r25

Rocha, R., Stier, C. T., Kifor, I., Ochoa-Maya, M. R., Rennke, H. G., Williams, G. H., & Adler, G. K. (2000). Aldosterone: A Mediator of Myocardial Necrosis and Renal Arteriopathy*. *Endocrinology*, *141*(10), 3871–3878. https://doi.org/10.1210/endo.141.10.7711

Rodier, F., & Campisi, J. (2011). Four faces of cellular senescence. *Journal of Cell Biology*, *192*(4), 547–556. https://doi.org/10.1083/jcb.201009094

Rodon, J., Dienstmann, R., Serra, V., & Tabernero, J. (2013). Development of PI3K inhibitors: Lessons learned from early clinical trials. *Nature Reviews. Clinical Oncology*, *10*(3), 143–153. https://doi.org/10.1038/nrclinonc.2013.10

Roger, L., Tomas, F., & Gire, V. (2021). Mechanisms and Regulation of Cellular Senescence. *International Journal of Molecular Sciences*, *22*(23), 13173. https://doi.org/10.3390/ijms222313173

Rogerson, F. M., Brennan, F. E., & Fuller, P. J. (2004). Mineralocorticoid receptor binding, structure and function. *Molecular and Cellular Endocrinology*, *217*(1–2), 203–212. https://doi.org/10.1016/j.mce.2003.10.021

Rojas, P. A., May, M., Sequeira, G. R., Elia, A., Alvarez, M., Martínez, P., Gonzalez, P., Hewitt, S., He, X., Perou, C. M., Molinolo, A., Gibbons, L., Abba, M. C., Gass, H., & Lanari, C. (2017). Progesterone Receptor Isoform Ratio: A Breast Cancer Prognostic and Predictive Factor for Antiprogestin Responsiveness. *JNCI: Journal of the National Cancer Institute*, *109*(7). https://doi.org/10.1093/jnci/djw317

Rose, C. R., Blum, R., Pichler, B., Lepier, A., Kafitz, K. W., & Konnerth, A. (2003). Truncated TrkB-T1 mediates neurotrophin-evoked calcium signalling in glia cells. *Nature*, *426*(6962), 74–78. https://doi.org/10.1038/nature01983

Rosenbaum, P., Schmidt, W., Helmerhorst, F. M., Wuttke, W., Rossmanith, W., Freundl, F., Thomas, K., Grillo, M., Wolf, A., & Heithecker, R. (2000). Inhibition of ovulation by a novel progestogen (drospirenone) alone or in combination with ethinylestradiol. *The European Journal of Contraception & Reproductive Health Care*, *5*(1), 16–24. https://doi.org/10.1080/13625180008500376

Ross, D. T., & Perou, C. M. (2001). A Comparison of Gene Expression Signatures from Breast Tumors and Breast Tissue Derived Cell Lines. *Disease Markers*, *17*(2), 99–109. https://doi.org/10.1155/2001/850531

Ross, R. K. (2000). Effect of Hormone Replacement Therapy on Breast Cancer Risk: Estrogen Versus Estrogen Plus Progestin. *Journal of the National Cancer Institute*, *92*(4), 328–332. https://doi.org/10.1093/jnci/92.4.328

Rossier, B. C., Baker, M. E., & Studer, R. A. (2015). Epithelial Sodium Transport and Its Control by Aldosterone: The Story of Our Internal Environment Revisited. *Physiological Reviews*, *95*(1), 297–340. https://doi.org/10.1152/physrev.00011.2014

Roth, H. E., Bhinderwala, F., Franco, R., Zhou, Y., & Powers, R. (2021). DNAJA1 Dysregulates Metabolism Promoting an Antiapoptotic Phenotype in Pancreatic Ductal Adenocarcinoma. *Journal of Proteome Research*, *20*(8), 3925–3939. https://doi.org/10.1021/acs.jproteome.1c00233

Roux, M. (2015). *A comparative study of divisive hierarchical clustering algorithms*. https://doi.org/10.48550/ARXIV.1506.08977

Roy, P. G., Pratt, N., Purdie, C. A., Baker, L., Ashfield, A., Quinlan, P., & Thompson, A. M. (2010). High CCND1 amplification identifies a group of poor prognosis women with estrogen receptor positive breast cancer. *International Journal of Cancer*, *127*(2), 355–360. https://doi.org/10.1002/ijc.25034

Rozeboom, A. M., Akil, H., & Seasholtz, A. F. (2007). Mineralocorticoid receptor overexpression in forebrain decreases anxiety-like behavior and alters the stress response in mice. *Proceedings of the National Academy of Sciences*, *104*(11), 4688–4693. https://doi.org/10.1073/pnas.0606067104

Rupprecht, R., Reul, J. M. H. M., Van Steensel, B., Spengler, D., Söder, M., Berning, B., Holsboer, F., & Damm, K. (1993). Pharmacological and functional characterization of human mineralocorticoid and glucocorticoid receptor ligands. *European Journal of Pharmacology: Molecular Pharmacology*, *247*(2), 145–154. https://doi.org/10.1016/0922-4106(93)90072-H

Rutkowsky, J. M., Knotts, T. A., Ono-Moore, K. D., McCoin, C. S., Huang, S., Schneider, D., Singh, S., Adams, S. H., & Hwang, D. H. (2014). Acylcarnitines activate proinflammatory signaling pathways. *American Journal of Physiology-Endocrinology and Metabolism*, *306*(12), E1378–E1387. https://doi.org/10.1152/ajpendo.00656.2013

Saigusa, S., Tanaka, K., Toiyama, Y., Yokoe, T., Okugawa, Y., Ioue, Y., Miki, C., & Kusunoki, M. (2009). Correlation of CD133, OCT4, and SOX2 in Rectal Cancer and Their Association with Distant Recurrence After Chemoradiotherapy. *Annals of Surgical Oncology*, *16*(12), 3488–3498. https://doi.org/10.1245/s10434-009-0617-z

Saik, O. V., Demenkov, P. S., Ivanisenko, T. V., Bragina, E. Yu., Freidin, M. B., Dosenko, V. E., Zolotareva, O. I., Choynzonov, E. L., Hofestaedt, R., & Ivanisenko, V. A. (2018). Search for New Candidate Genes Involved in the Comorbidity of Asthma and Hypertension Based on Automatic Analysis of Scientific Literature. *Journal of Integrative Bioinformatics*, *15*(4), 20180054. https://doi.org/10.1515/jib-2018-0054

Sali, A., & Attali, D. (2020). *shinycssloaders: Add Loading Animations to a "shiny" Output While It's Recalculating* (R package version 1.0.0) [Computer software]. https://CRAN.R-project.org/package=shinycssloaders

Samavat, H., & Kurzer, M. S. (2015). Estrogen metabolism and breast cancer. *Cancer Letters*, *356*(2), 231–243. https://doi.org/10.1016/j.canlet.2014.04.018

Sanders, Y. Y., Liu, H., Zhang, X., Hecker, L., Bernard, K., Desai, L., Liu, G., & Thannickal, V. J. (2013). Histone Modifications in Senescence-Associated Resistance to Apoptosis by Oxidative Stress. *Redox Biology*, *1*(1), 8–16. https://doi.org/10.1016/j.redox.2012.11.004

Santen, R. J. (2003). Risk of breast cancer with progestins: Critical assessment of current data. *Steroids*, *68*(10–13), 953–964. https://doi.org/10.1016/S0039-128X(03)00138-7

SAS Institute Inc. (1990). *SAS/STAT User's Guide: GLM-VARCOMP* (4th Edn Cary, Vol. 2). SAS institute Incorporated.

Sasaki, M., Kumazaki, T., Takano, H., Nishiyama, M., & Mitsui, Y. (2001). Senescent cells are resistant to death despite low Bcl-2 level. *Mechanisms of Ageing and Development*, *122*(15), 1695–1706. https://doi.org/10.1016/S0047-6374(01)00281-0

Sasano, H., Frost, A. R., Saitoh, R., Matsunaga, G., Nagura, H., Krozowski, Z. S., & Silverberg, S. G. (1997). Localization of mineralocorticoid receptor and 11 beta-hydroxysteroid dehydrogenase type II in human breast and its disorders. *Anticancer Research*, *17*(3C), 2001–2007.

Sato, R., Nakano, T., Hosonaga, M., Sampetrean, O., Harigai, R., Sasaki, T., Koya, I., Okano, H., Kudoh, J., Saya, H., & Arima, Y. (2017). RNA Sequencing Analysis Reveals Interactions between Breast Cancer or Melanoma Cells and the Tissue Microenvironment during Brain Metastasis. *BioMed Research International*, *2017*, 1–10. https://doi.org/10.1155/2017/8032910

Scarpin, K. M., Graham, J. D., Mote, P. A., & Clarke, C. L. (2009). Progesterone Action in Human Tissues: Regulation by Progesterone Receptor (PR) Isoform Expression, Nuclear Positioning and Coregulator Expression. *Nuclear Receptor Signaling*, *7*(1), nrs.07009. https://doi.org/10.1621/nrs.07009

Schena, M., Shalon, D., Davis, R. W., & Brown, P. O. (1995). Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science*, *270*(5235), 467–470. https://doi.org/10.1126/science.270.5235.467

Schürmann, R., Holler, T., & Benda, N. (2004). Estradiol and drospirenone for climacteric symptoms in postmenopausal women: A double-blind, randomized, placebo-controlled study of the safety and efficacy of three dose regimens. *Climacteric*, *7*(2), 189–196. https://doi.org/10.1080/13697130410001713698

Scott, J. H., Menouar, M. A., & Dunn, R. J. (2023). Physiology, Aldosterone. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK470339/

Seachrist, D. D., Anstine, L. J., & Keri, R. A. (2021). FOXA1: A Pioneer of Nuclear Receptor Action in Breast Cancer. *Cancers*, *13*(20), 5205. https://doi.org/10.3390/cancers13205205

Seal, D. B., Saha, S., Mukherjee, P., Chatterjee, M., Mukherjee, A., & Dey, K. N. (2016). Gene ranking: An entropy & decision tree based approach. *2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 1–5. https://doi.org/10.1109/UEMCON.2016.7777837

Seeger, H., Ruan, X., Neubauer, H., & Mueck, A. O. (2011). Effect of drospirenone on proliferation of human benign and cancerous epithelial breast cells. *Hormone Molecular Biology and Clinical Investigation*, *6*(2), 211–214. https://doi.org/10.1515/HMBCI.2011.008

Seixas, M. J., Domingues, R. R., & Antunes, A. (2023). Decoding the Transcriptome of Sharks, Rays, and Chimaeras: Insights into Their Physiology, Morphology,

Evolution, and Biomedical Applications. *Fishes*, *8*(5), 271. https://doi.org/10.3390/fishes8050271

Sene, A., Khan, A. A., Cox, D., Nakamura, R. E. I., Santeford, A., Kim, B. M., Sidhu, R., Onken, M. D., Harbour, J. W., Hagbi-Levi, S., Chowers, I., Edwards, P. A., Baldan, A., Parks, J. S., Ory, D. S., & Apte, R. S. (2013). Impaired Cholesterol Efflux in Senescent Macrophages Promotes Age-Related Macular Degeneration. *Cell Metabolism*, *17*(4), 549–561. https://doi.org/10.1016/j.cmet.2013.03.009

Sever, R., & Glass, C. K. (2013). Signaling by Nuclear Receptors. *Cold Spring Harbor Perspectives in Biology*, *5*(3), a016709–a016709. https://doi.org/10.1101/cshperspect.a016709

Shamseddin, M., De Martino, F., Constantin, C., Scabia, V., Lancelot, A., Laszlo, C., Ayyannan, A., Battista, L., Raffoul, W., Gailloud-Matthieu, M., Bucher, P., Fiche, M., Ambrosini, G., Sflomos, G., & Brisken, C. (2021). Contraceptive progestins with androgenic properties stimulate breast epithelial cell proliferation. *EMBO Molecular Medicine*, *13*(7), e14314. https://doi.org/10.15252/emmm.202114314

Shankavaram, U. T., Reinhold, W. C., Nishizuka, S., Major, S., Morita, D., Chary, K. K., Reimers, M. A., Scherf, U., Kahn, A., Dolginow, D., Cossman, J., Kaldjian, E. P., Scudiero, D. A., Petricoin, E., Liotta, L., Lee, J. K., & Weinstein, J. N. (2007). Transcript and protein expression profiles of the NCI-60 cancer cell panel: An integromic microarray study. *Molecular Cancer Therapeutics*, *6*(3), 820–832. https://doi.org/10.1158/1535-7163.MCT-06-0650

Sharma, S., Dincer, C., Weidemüller, P., Wright, G. J., & Petsalaki, E. (2020). CEN - tools: An integrative platform to identify the contexts of essential genes. *Molecular Systems Biology*, *16*(10), e9698. https://doi.org/10.15252/msb.20209698

Shaw, N. D., Histed, S. N., Srouji, S. S., Yang, J., Lee, H., & Hall, J. E. (2010). Estrogen Negative Feedback on Gonadotropin Secretion: Evidence for a Direct Pituitary Effect in Women. *The Journal of Clinical Endocrinology & Metabolism*, *95*(4), 1955–1961. https://doi.org/10.1210/jc.2009-2108

Shchelkunova, T. A., & Morozov, I. A. (2015). [Molecular basis and tissue specificity of the progestins action]. *Molekuliarnaia Biologiia*, *49*(5), 728–748. https://doi.org/10.7868/S0026898415050158

Shelton, D. N., Chang, E., Whittier, P. S., Choi, D., & Funk, W. D. (1999). Microarray analysis of replicative senescence. *Current Biology*, *9*(17), 939–945. https://doi.org/10.1016/S0960-9822(99)80420-5

Shen, R., Li, P., Li, B., Zhang, B., Feng, L., & Cheng, S. (2020). Identification of Distinct Immune Subtypes in Colorectal Cancer Based on the Stromal Compartment. *Frontiers in Oncology*, *9*, 1497. https://doi.org/10.3389/fonc.2019.01497

Shi, X., Gekas, C., Verduzco, D., Petiwala, S., Jeffries, C., Lu, C., Murphy, E., Anton, T., Vo, A. H., Xiao, Z., Narayanan, P., Barnes, J. M., Roy, S., Ramathal, C., Flister, M. J., & Dezso, Z. (2022). *TCGA $_{DEPMAP}$ – Mapping Translational Dependencies and Synthetic Lethalities within The Cancer Genome Atlas* [Preprint]. Cancer Biology. https://doi.org/10.1101/2022.03.24.485544

Shibata, S. (2017). 30 YEARS OF THE MINERALOCORTICOID RECEPTOR: Mineralocorticoid receptor and NaCl transport mechanisms in the renal distal nephron. *Journal of Endocrinology*, *234*(1), T35–T47. https://doi.org/10.1530/JOE-16-0669

Shigemizu, D., Mori, T., Akiyama, S., Higaki, S., Watanabe, H., Sakurai, T., Niida, S., & Ozaki, K. (2020). Identification of potential blood biomarkers for early diagnosis of Alzheimer's disease through RNA sequencing analysis. *Alzheimer's Research & Therapy*, *12*(1), 87. https://doi.org/10.1186/s13195-020-00654-x

Shimada, K., Bachman, J. A., Muhlich, J. L., & Mitchison, T. J. (2021). shinyDepMap, a tool to identify targetable cancer genes and their functional connections from Cancer Dependency Map data. *eLife*, *10*, e57116. https://doi.org/10.7554/eLife.57116

Shimosawa, T. (2013). Salt, the renin–angiotensin–aldosterone system and resistant hypertension. *Hypertension Research*, *36*(8), 657–660. https://doi.org/10.1038/hr.2013.69

Shohat, S., & Shifman, S. (2019). Genes essential for embryonic stem cells are associated with neurodevelopmental disorders. *Genome Research*, *29*(11), 1910–1918. https://doi.org/10.1101/gr.250019.119

Shulgin, A. A., Lebedev, T. D., Prassolov, V. S., & Spirin, P. V. (2021). Plasmolipin and Its Role in Cell Processes. *Molecular Biology*, *55*(6), 773–785. https://doi.org/10.1134/S0026893321050113

Shull, J. (1997). Ovary-intact, but not ovariectomized female ACI rats treated with 17beta-estradiol rapidly develop mammary carcinoma. *Carcinogenesis*, *18*(8), 1595–1601. https://doi.org/10.1093/carcin/18.8.1595

Si, Y., Liu, P., Li, P., & Brutnell, T. P. (2014). Model-based clustering for RNA-seq data. *Bioinformatics*, *30*(2), 197–205. https://doi.org/10.1093/bioinformatics/btt632

Siegel, R. L., Miller, K. D., Wagle, N. S., & Jemal, A. (2023). Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians*, *73*(1), 17–48. https://doi.org/10.3322/caac.21763

Sievert, C. (2020). *Interactive Web-Based Data Visualization with R, plotly, and shiny* (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9780429447273

Sinars, C. R., Cheung-Flynn, J., Rimerman, R. A., Scammell, J. G., Smith, D. F., & Clardy, J. (2003). Structure of the large FK506-binding protein FKBP51, an Hsp90-binding protein and a component of steroid receptor complexes. *Proceedings of the National Academy of Sciences*, *100*(3), 868–873. https://doi.org/10.1073/pnas.0231020100

Singh, B., & Eyras, E. (2017). The role of alternative splicing in cancer. *Transcription*, *8*(2), 91–98. https://doi.org/10.1080/21541264.2016.1268245

Sitruk-Ware, R. (2004). Pharmacological profile of progestins. *Maturitas*, *47*(4), 277–283. https://doi.org/10.1016/j.maturitas.2004.01.001

Sitruk-Ware, R., Bonsack, B., Brinton, R., Schumacher, M., Kumar, N., Lee, J.-Y., Castelli, V., Corey, S., Coats, A., Sadanandan, N., Gonzales-Portillo, B., Heyck, M., Shear, A., Blaise, C., Zhang, H., Sheyner, M., García-Sánchez, J., Navarro, L., El-Etr, M., … Borlongan, C. V. (2021). Progress in progestin-based therapies for neurological disorders. *Neuroscience & Biobehavioral Reviews*, *122*, 38–65. https://doi.org/10.1016/j.neubiorev.2020.12.007

Slamon, D. J., Godolphin, W., Jones, L. A., Holt, J. A., Wong, S. G., Keith, D. E., Levin, W. J., Stuart, S. G., Udove, J., Ullrich, A., & Press, M. F. (1989). Studies of the HER-2/ *neu* Proto-Oncogene in Human Breast and Ovarian Cancer. *Science*, *244*(4905), 707–712. https://doi.org/10.1126/science.2470152

Song, P., Duan, J., Ding, J., Liu, J., Fang, Z., Xu, H., Li, Z., Du, W., Xu, M., Ling, Y., He, F., Tao, K., & Wang, L. (2023). Cellular senescence primes liver fibrosis

regression through Notch-EZH2. *MedComm*, *4*(5), e346. https://doi.org/10.1002/mco2.346

Şorop, V. B., Borugă, V. M., Pînzaru, I. A., Barac, I. R., Utescu, C., Maghiari, A. L., Baderca, F., Bălan, L., Şorop-Florea, M., Dumitraşcu, V., Anastasiu, D. M., Simu, S., Radu, D., & Suciu, O. (2020). Hormone treatment and UVB exposure influences on female mice regarding skin physiological parameters, biochemical parameters and organ histology. *Romanian Journal of Morphology and Embryology = Revue Roumaine De Morphologie Et Embryologie*, *61*(3), 879–887. https://doi.org/10.47162/RJME.61.3.26

Sosa, M. S., Lopez-Haber, C., Yang, C., Wang, H., Lemmon, M. A., Busillo, J. M., Luo, J., Benovic, J. L., Klein-Szanto, A., Yagi, H., Gutkind, J. S., Parsons, R. E., & Kazanietz, M. G. (2010). Identification of the Rac-GEF P-Rex1 as an Essential Mediator of ErbB Signaling in Breast Cancer. *Molecular Cell*, *40*(6), 877–892. https://doi.org/10.1016/j.molcel.2010.11.029

Sotiriou, C., Neo, S.-Y., McShane, L. M., Korn, E. L., Long, P. M., Jazaeri, A., Martiat, P., Fox, S. B., Harris, A. L., & Liu, E. T. (2003). Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proceedings of the National Academy of Sciences*, *100*(18), 10393–10398. https://doi.org/10.1073/pnas.1732912100

Stanczyk, F. Z. (2002). Pharmacokinetics and Potency of Progestins used for Hormone Replacement Therapy and Contraception. *Reviews in Endocrine and Metabolic Disorders*, *3*(3), 211–224. https://doi.org/10.1023/A:1020072325818

Stanczyk, F. Z. (2007). Structure –Function Relationships, Pharmacokinetics, and Potency of Orally and Parenterally Administered Progestogens. In *Treatment of the Postmenopausal Woman* (pp. 779–798). Elsevier. https://doi.org/10.1016/B978-012369443-0/50067-3

Stanelle, J. (2002). Gene expression changes in response to E2F1 activation. *Nucleic Acids Research*, *30*(8), 1859–1867. https://doi.org/10.1093/nar/30.8.1859

Stensjøen, A. L., Solheim, O., Kvistad, K. A., Håberg, A. K., Salvesen, Ø., & Berntsen, E. M. (2015). Growth dynamics of untreated glioblastomas in vivo. *Neuro-Oncology*, *17*(10), 1402–1411. https://doi.org/10.1093/neuonc/nov029

Stoilov, P., Castren, E., & Stamm, S. (2002). Analysis of the human TrkB gene genomic organization reveals novel TrkB isoforms, unusual gene length, and splicing mechanism. *Biochemical and Biophysical Research Communications*, *290*(3), 1054–1065. https://doi.org/10.1006/bbrc.2001.6301

Su, L., Liu, G., Bai, T., Meng, X., & Ma, Q. (2018). MGOGP: A gene module-based heuristic algorithm for cancer-related gene prioritization. *BMC Bioinformatics*, *19*(1), 215. https://doi.org/10.1186/s12859-018-2216-0

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

Sultan, A. S., Brim, H., & Sherif, Z. A. (2008). Co-overexpression of Janus kinase 2 and signal transducer and activator of transcription 5a promotes differentiation of mammary cancer cells through reversal of epithelial–mesenchymal transition. *Cancer Science*, *99*(2), 272–279. https://doi.org/10.1111/j.1349-7006.2007.00685.x

Sun, X., Wang, Z., Chen, X., & Shen, K. (2021). CRISPR-cas9 Screening Identified Lethal Genes Enriched in Cell Cycle Pathway and of Prognosis Significance in Breast Cancer. *Frontiers in Cell and Developmental Biology*, *9*, 646774. https://doi.org/10.3389/fcell.2021.646774

Sunayama, J., Matsuda, K.-I., Sato, A., Tachibana, K., Suzuki, K., Narita, Y., Shibui, S., Sakurada, K., Kayama, T., Tomiyama, A., & Kitanaka, C. (2010). Crosstalk between the PI3K/mTOR and MEK/ERK pathways involved in the maintenance of self-renewal and tumorigenicity of glioblastoma stem-like cells. *Stem Cells (Dayton, Ohio)*, *28*(11), 1930–1939. https://doi.org/10.1002/stem.521

Sung, B. H., Choe, D., Kim, S. C., & Cho, B.-K. (2016). Construction of a minimal genome as a chassis for synthetic biology. *Essays in Biochemistry*, *60*(4), 337–346. https://doi.org/10.1042/EBC20160024

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

Szafran, H., & Smielak-Korombel, W. (1998). [The role of estrogens in hormonal regulation of lipid metabolism in women]. *Przeglad Lekarski*, *55*(5), 266–270.

Szopa, W., Burley, T. A., Kramer-Marek, G., & Kaspera, W. (2017). Diagnostic and Therapeutic Biomarkers in Glioblastoma: Current Status and Future Perspectives. *BioMed Research International*, *2017*, 1–13. https://doi.org/10.1155/2017/8013575

Szostakowska, M., Trębińska-Stryjewska, A., Grzybowska, E. A., & Fabisiewicz, A. (2019). Resistance to endocrine therapy in breast cancer: Molecular mechanisms and future goals. *Breast Cancer Research and Treatment*, *173*(3), 489–497. https://doi.org/10.1007/s10549-018-5023-4

Taraborrelli, S. (2015). Physiology, production and action of progesterone. *Acta Obstetricia et Gynecologica Scandinavica*, *94*(S161), 8–16. https://doi.org/10.1111/aogs.12771

Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A., & Conesa, A. (2011). Differential expression in RNA-seq: A matter of depth. *Genome Research*, *21*(12), 2213–2223. https://doi.org/10.1101/gr.124321.111

Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., Boutselakis, H., Cole, C. G., Creatore, C., Dawson, E., Fish, P., Harsha, B., Hathaway, C., Jupe, S. C., Kok, C. Y., Noble, K., Ponting, L., Ramshaw, C. C., Rye, C. E., … Forbes, S. A. (2019). COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research*, *47*(D1), D941–D947. https://doi.org/10.1093/nar/gky1015

Tella, S. H., & Gallagher, J. C. (2014). Prevention and treatment of postmenopausal osteoporosis. *The Journal of Steroid Biochemistry and Molecular Biology*, *142*, 155–170. https://doi.org/10.1016/j.jsbmb.2013.09.008

Tepper, C. G., Seldin, M. F., & Mudryj, M. (2000). Fas-Mediated Apoptosis of Proliferating, Transiently Growth-Arrested, and Senescent Normal Human Fibroblasts. *Experimental Cell Research*, *260*(1), 9–19. https://doi.org/10.1006/excr.2000.4990

Tesch, G. H., & Young, M. J. (2017). Mineralocorticoid Receptor Signaling as a Therapeutic Target for Renal and Cardiac Fibrosis. *Frontiers in Pharmacology*, *8*, 313. https://doi.org/10.3389/fphar.2017.00313

Tharakan, R., Lepont, P., Singleton, D., Kumar, R., & Khan, S. (2008). Phosphorylation of estrogen receptor alpha, serine residue 305 enhances activity. *Molecular and Cellular Endocrinology*, *295*(1–2), 70–78. https://doi.org/10.1016/j.mce.2008.07.018

Thomas, W., & Harvey, B. (2019). Aldosterone Regulation of Protein Kinase Signaling Pathways and Renal Na $^{+}$ Transport by Non-genomic Mechanisms. In B. Harvey & F. Jaisser (Eds.), *Aldosterone-Mineralocorticoid Receptor—Cell Biology to Translational Medicine*. IntechOpen. https://doi.org/10.5772/intechopen.87238

Thornblad, T. A., Elliott, K. S., Jowett, J., & Visscher, P. M. (2007). Prioritization of Positional Candidate Genes Using Multiple Web-Based Software Tools. *Twin Research and Human Genetics*, *10*(6), 861–870. https://doi.org/10.1375/twin.10.6.861

Tian, J.-M., Ran, B., Zhang, C.-L., Yan, D.-M., & Li, X.-H. (2018). Estrogen and progesterone promote breast cancer cell proliferation by inducing cyclin G1 expression. *Brazilian Journal of Medical and Biological Research = Revista Brasileira De Pesquisas Medicas E Biologicas*, *51*(3), 1–7. https://doi.org/10.1590/1414-431X20175612

Tian, Q., Stepaniants, S. B., Mao, M., Weng, L., Feetham, M. C., Doyle, M. J., Yi, E. C., Dai, H., Thorsson, V., Eng, J., Goodlett, D., Berger, J. P., Gunter, B., Linseley, P. S., Stoughton, R. B., Aebersold, R., Collins, S. J., Hanlon, W. A., & Hood, L. E. (2004). Integrated Genomic and Proteomic Analyses of Gene Expression in Mammalian Cells. *Molecular & Cellular Proteomics*, *3*(10), 960–969. https://doi.org/10.1074/mcp.M400055-MCP200

Tong, F., Lu, G., Zang, J., Hao, D., Xu, W., Chen, J., Ding, Q., & Xiong, H. (2023). FKBP5 associated CD8 T cell infiltration is a novel prognostic biomarker in luminal B breast cancer. *Journal of International Medical Research*, *51*(11), 03000605231211771. https://doi.org/10.1177/03000605231211771

Touat, M., Idbaih, A., Sanson, M., & Ligon, K. L. (2017). Glioblastoma targeted therapy: Updated approaches from recent biological insights. *Annals of Oncology*, *28*(7), 1457–1472. https://doi.org/10.1093/annonc/mdx106

Tran, A. N., Dussaq, A. M., Kennell, T., Willey, C. D., & Hjelmeland, A. B. (2019). HPAanalyze: An R package that facilitates the retrieval and analysis of the Human Protein Atlas data. *BMC Bioinformatics*, *20*(1), 463. https://doi.org/10.1186/s12859-019-3059-z

Tranchevent, L.-C., Ardeshirdavani, A., ElShal, S., Alcaide, D., Aerts, J., Auboeuf, D., & Moreau, Y. (2016). Candidate gene prioritization with Endeavour. *Nucleic Acids Research*, *44*(W1), W117–W121. https://doi.org/10.1093/nar/gkw365

Tranchevent, L.-C., Barriot, R., Yu, S., Van Vooren, S., Van Loo, P., Coessens, B., De Moor, B., Aerts, S., & Moreau, Y. (2008). ENDEAVOUR update: A web resource for gene prioritization in multiple species. *Nucleic Acids Research*, *36*(Web Server), W377–W384. https://doi.org/10.1093/nar/gkn325

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., & Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*, *31*(1), 46–53. https://doi.org/10.1038/nbt.2450

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics (Oxford, England)*, *25*(9), 1105–1111. https://doi.org/10.1093/bioinformatics/btp120

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., & Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform

switching during cell differentiation. *Nature Biotechnology*, *28*(5), 511–515. https://doi.org/10.1038/nbt.1621

Treeck, O., Schüler-Toprak, S., & Ortmann, O. (2020). Estrogen Actions in Triple-Negative Breast Cancer. *Cells*, *9*(11), 2358. https://doi.org/10.3390/cells9112358

Tremblay, A., Tremblay, G. B., Labrie, F., & Giguère, V. (1999). Ligand-Independent Recruitment of SRC-1 to Estrogen Receptor β through Phosphorylation of Activation Function AF-1. *Molecular Cell*, *3*(4), 513–519. https://doi.org/10.1016/S1097-2765(00)80479-7

Trougakos, I. P., & Gonos, E. S. (2002). Clusterin/Apolipoprotein J in human aging and cancer. *The International Journal of Biochemistry & Cell Biology*, *34*(11), 1430–1448. https://doi.org/10.1016/S1357-2725(02)00041-9

Tsherniak, A., Vazquez, F., Montgomery, P. G., Weir, B. A., Kryukov, G., Cowley, G. S., Gill, S., Harrington, W. F., Pantel, S., Krill-Burger, J. M., Meyers, R. M., Ali, L., Goodale, A., Lee, Y., Jiang, G., Hsiao, J., Gerath, W. F. J., Howell, S., Merkel, E., … Hahn, W. C. (2017). Defining a Cancer Dependency Map. *Cell*, *170*(3), 564-576.e16. https://doi.org/10.1016/j.cell.2017.06.010

Tsitsiridis, G., Steinkamp, R., Giurgiu, M., Brauner, B., Fobo, G., Frishman, G., Montrone, C., & Ruepp, A. (2023). CORUM: The comprehensive resource of mammalian protein complexes–2022. *Nucleic Acids Research*, *51*(D1), D539–D545. https://doi.org/10.1093/nar/gkac1015

Tsuboi, K., Nagatomo, T., Gohno, T., Higuchi, T., Sasaki, S., Fujiki, N., Kurosumi, M., Takei, H., Yamaguchi, Y., Niwa, T., & Hayashi, S. (2017). Single CpG site methylation controls estrogen receptor gene transcription and correlates with hormone therapy resistance. *The Journal of Steroid Biochemistry and Molecular Biology*, *171*, 209–217. https://doi.org/10.1016/j.jsbmb.2017.04.001

Tung, L., Mohamed, M. K., Hoeffler, J. P., Takimoto, G. S., & Horwitz, K. B. (1993). Antagonist-occupied human progesterone B-receptors activate transcription without binding to progesterone response elements and are dominantly inhibited by A-receptors. *Molecular Endocrinology*, *7*(10), 1256–1265. https://doi.org/10.1210/mend.7.10.8123133

Turner, F. S., Clutterbuck, D. R., & Semple, C. A. (2003). POCUS: Mining genomic sequence annotation to predict disease genes. *Genome Biology*, *4*(11), R75. https://doi.org/10.1186/gb-2003-4-11-r75

Tyshkovskiy, A., Ma, S., Shindyapina, A. V., Tikhonov, S., Lee, S.-G., Bozaykut, P., Castro, J. P., Seluanov, A., Schork, N. J., Gorbunova, V., Dmitriev, S. E., Miller, R. A., & Gladyshev, V. N. (2023). Distinct longevity mechanisms across and within species and their association with aging. *Cell*, *186*(13), 2929-2949.e20. https://doi.org/10.1016/j.cell.2023.05.002

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—New capabilities and interfaces. *Nucleic Acids Research*, *40*(15), e115–e115. https://doi.org/10.1093/nar/gks596

Van Agthoven, T., Sieuwerts, A. M., Veldscholte, J., Meijer-van Gelder, M. E., Smid, M., Brinkman, A., Den Dekker, A. T., Leroy, I. M., Van IJcken, W. F. J., Sleijfer, S., Foekens, J. A., & Dorssers, L. C. J. (2009). CITED2 and NCOR2 in anti-oestrogen resistance and progression of breast cancer. *British Journal of Cancer*, *101*(11), 1824–1832. https://doi.org/10.1038/sj.bjc.6605423

Van Bragt, M. P., Hu, X., Xie, Y., & Li, Z. (2014). RUNX1, a transcription factor mutated in breast cancer, controls the fate of ER-positive mammary luminal cells. *eLife*, *3*, e03881. https://doi.org/10.7554/eLife.03881

Van 'T Veer, L. J., Dai, H., Van De Vijver, M. J., He, Y. D., Hart, A. A. M., Mao, M., Peterse, H. L., Van Der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., & Friend, S. H. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, *415*(6871), 530–536. https://doi.org/10.1038/415530a

Van Verk, M. C., Hickman, R., Pieterse, C. M. J., & Van Wees, S. C. M. (2013). RNA-Seq: Revelation of the messengers. *Trends in Plant Science*, *18*(4), 175–179. https://doi.org/10.1016/j.tplants.2013.02.001

Vassiliou, A. G., Athanasiou, N., Vassiliadi, D. A., Jahaj, E., Keskinidou, C., Kotanidou, A., & Dimopoulou, I. (2021). Glucocorticoid and mineralocorticoid receptor expression in critical illness: A narrative review. *World Journal of Critical Care Medicine*, *10*(4), 102–111. https://doi.org/10.5492/wjccm.v10.i4.102

Vazquez, F., & Boehm, J. S. (2020). The Cancer Dependency Map enables drug mechanism-of-action investigations. *Molecular Systems Biology*, *16*(7), e9757. https://doi.org/10.15252/msb.20209757

Velculescu, V. E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M. A., Bassett, D. E., Hieter, P., Vogelstein, B., & Kinzler, K. W. (1997). Characterization of the Yeast Transcriptome. *Cell*, *88*(2), 243–251. https://doi.org/10.1016/S0092-8674(00)81845-0

Vera, J. C., Wheat, C. W., Fescemyer, H. W., Frilander, M. J., Crawford, D. L., Hanski, I., & Marden, J. H. (2008). Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Molecular Ecology*, *17*(7), 1636–1647. https://doi.org/10.1111/j.1365-294X.2008.03666.x

Verhaak, R. G. W., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., Miller, C. R., Ding, L., Golub, T., Mesirov, J. P., Alexe, G., Lawrence, M., O'Kelly, M., Tamayo, P., Weir, B. A., Gabriel, S., Winckler, W., Gupta, S., Jakkula, L., … Hayes, D. N. (2010). Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*, *17*(1), 98–110. https://doi.org/10.1016/j.ccr.2009.12.020

Vidman, L., Källberg, D., & Rydén, P. (2019). Cluster analysis on high dimensional RNA-seq data with applications to cancer research—An evaluation study. *PLOS ONE*, *14*(12), e0219102. https://doi.org/10.1371/journal.pone.0219102

Vivanco, I., & Sawyers, C. L. (2002). The phosphatidylinositol 3-Kinase AKT pathway in human cancer. *Nature Reviews. Cancer*, *2*(7), 489–501. https://doi.org/10.1038/nrc839

Vogel, C., & Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics*, *13*(4), 227–232. https://doi.org/10.1038/nrg3185

Voogt, J. L. (1978). Control of hormone release during lactation. *Clinics in Obstetrics and Gynaecology*, *5*(2), 435–455.

Voutsadakis, I. (2016). Epithelial-Mesenchymal Transition (EMT) and Regulation of EMT Factors by Steroid Nuclear Receptors in Breast Cancer: A Review and in Silico Investigation. *Journal of Clinical Medicine*, *5*(1), 11. https://doi.org/10.3390/jcm5010011

Vural-Ozdeniz, M., Calisir, K., Acar, R., Yavuz, A., Ozgur, M. M., Dalgıc, E., & Konu, O. (2024). CAP-RNAseq: An integrated pipeline for functional annotation and prioritization of co-expression clusters. *Briefings in Bioinformatics*, *25*(2), bbad536. https://doi.org/10.1093/bib/bbad536

Walker, S. R., Liu, S., Xiang, M., Nicolais, M., Hatzi, K., Giannopoulou, E., Elemento, O., Cerchietti, L., Melnick, A., & Frank, D. A. (2015). The transcriptional modulator BCL6 as a molecular target for breast cancer therapy. *Oncogene*, *34*(9), 1073–1082. https://doi.org/10.1038/onc.2014.61

Wambach, G., & Higgins, J. R. (1978). Antimineralocorticoid Action of Progesterone in the Rat: Correlation of the Effect on Electrolyte Excretion and Interaction with Renal Mineralocorticoid Receptors*. *Endocrinology*, *102*(6), 1686–1693. https://doi.org/10.1210/endo-102-6-1686

Wan, Y., McDaniel, K., Wu, N., Ramos-Lorenzo, S., Glaser, T., Venter, J., Francis, H., Kennedy, L., Sato, K., Zhou, T., Kyritsi, K., Huang, Q., Annable, T., Wu, C., Glaser, S., Alpini, G., & Meng, F. (2017). Regulation of Cellular Senescence by miR-34a in Alcoholic Liver Injury. *The American Journal of Pathology*, *187*(12), 2788–2798. https://doi.org/10.1016/j.ajpath.2017.08.027

Wang, C., Li, C., Cai, J.-H., Sheu, P., Tsai, J., Wu, M.-Y., Li, C.-J., & Hou, M.-F. (2019). Identification of Prognostic Candidate Genes in Breast Cancer by Integrated Bioinformatic Analysis. *Journal of Clinical Medicine*, *8*(8), 1160. https://doi.org/10.3390/jcm8081160

Wang, C.-Y., Xuan, D. T. M., Ye, P.-H., Li, C.-Y., Anuraga, G., Ta, H. D. K., Lai, M.-D., & Hsu, H.-P. (2023). Synergistic suppressive effects on triple-negative breast cancer by the combination of JTC-801 and sodium oxamate. *American Journal of Cancer Research*, *13*(10), 4661–4677.

Wang, D. (2008). Discrepancy between mRNA and protein abundance: Insight from information retrieval process in computers. *Computational Biology and Chemistry*, *32*(6), 462–468. https://doi.org/10.1016/j.compbiolchem.2008.07.014

Wang, D., Wang, C., Wang, L., & Chen, Y. (2019). A comprehensive review in improving delivery of small-molecule chemotherapeutic agents overcoming the blood-brain/brain tumor barriers for glioblastoma treatment. *Drug Delivery*, *26*(1), 551–565. https://doi.org/10.1080/10717544.2019.1616235

Wang, E. (1995). Senescent human fibroblasts resist programmed cell death, and failure to suppress bcl2 is involved. *Cancer Research*, *55*(11), 2284–2292.

Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F., & Jaenisch, R. (2013). One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell*, *153*(4), 910–918. https://doi.org/10.1016/j.cell.2013.04.025

Wang, J., Wen, S., Symmans, W. F., Pusztai, L., & Coombes, K. R. (2009). The Bimodality Index: A criterion for Discovering and Ranking Bimodal Signatures from Cancer Gene Expression Profiling Data. *Cancer Informatics*, *7*, CIN.S2846. https://doi.org/10.4137/CIN.S2846

Wang, L., Brugge, J. S., & Janes, K. A. (2011). Intersection of FOXO- and RUNX1-mediated gene expression programs in single breast epithelial cells during morphogenesis and tumor progression. *Proceedings of the National Academy of Sciences*, *108*(40). https://doi.org/10.1073/pnas.1103423108

Wang, L., Zhou, L., Li, M., Zhao, J., Liu, Y., Chen, Y., Qin, X., Wang, S., Chen, H., Piao, Y., Xiang, R., Li, J., & Shi, Y. (2022). Genome-wide CRISPR/Cas9 knockout screening uncovers ZNF319 as a novel tumor suppressor critical for breast cancer metastasis. *Biochemical and Biophysical Research Communications*, *589*, 107–115. https://doi.org/10.1016/j.bbrc.2021.12.023

Wang, M., & Tang, J. (2018). RelB facilitates cell migration and invasion in breast cancer via MMP1 upregulation. *Annals of Oncology*, *29*, ix19. https://doi.org/10.1093/annonc/mdy428.018

Wang, M., Zhang, Y., Xu, Z., Qian, P., Sun, W., Wang, X., Jian, Z., Xia, T., Xu, Y., & Tang, J. (2020). RelB sustains endocrine resistant malignancy: An insight of noncanonical NF-κB pathway into breast Cancer progression. *Cell Communication and Signaling*, *18*(1), 128. https://doi.org/10.1186/s12964-020-00613-x

Wang, S., Hauskrecht, M., & Visweswaran, S. (2009). Gene prioritization using a probabilistic knowledge model. *2009 IEEE International Conference on Bioinformatics and Biomedicine Workshop*, 272–278. https://doi.org/10.1109/BIBMW.2009.5332107

Wang, T., Birsoy, K., Hughes, N. W., Krupczak, K. M., Post, Y., Wei, J. J., Lander, E. S., & Sabatini, D. M. (2015). Identification and characterization of essential genes in the human genome. *Science*, *350*(6264), 1096–1101. https://doi.org/10.1126/science.aac7041

Wang, T., Wei, J. J., Sabatini, D. M., & Lander, E. S. (2014). Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science*, *343*(6166), 80–84. https://doi.org/10.1126/science.1246981

Wang, Y. E., Kutnetsov, L., Partensky, A., Farid, J., & Quackenbush, J. (2017). WebMeV: A Cloud Platform for Analyzing and Visualizing Cancer Genomic Data. *Cancer Research*, *77*(21), e11–e14. https://doi.org/10.1158/0008-5472.CAN-17-0802

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*(1), 57–63. https://doi.org/10.1038/nrg2484

Wang, Z., Lucas, F. A. S., Qiu, P., & Liu, Y. (2014). Improving the sensitivity of sample clustering by leveraging gene co-expression networks in variable selection. *BMC Bioinformatics*, *15*(1), 153. https://doi.org/10.1186/1471-2105-15-153

Ward, A. V., Matthews, S. B., Fettig, L. M., Riley, D., Finlay-Schultz, J., Paul, K. V., Jackman, M., Kabos, P., MacLean, P. S., & Sartorius, C. A. (2022). Estrogens and Progestins Cooperatively Shift Breast Cancer Cell Metabolism. *Cancers*, *14*(7), 1776. https://doi.org/10.3390/cancers14071776

Warming, L., Ravn, P., Nielsen, T., & Christiansen, C. (2004). Safety and efficacy of drospirenone used in a continuous combination with 17beta-estradiol for prevention of postmenopausal osteoporosis. *Climacteric: The Journal of the International Menopause Society*, *7*(1), 103–111. https://doi.org/10.1080/13697130310001651535

Washburn, M. P., Koller, A., Oshiro, G., Ulaszek, R. R., Plouffe, D., Deciu, C., Winzeler, E., & Yates, J. R. (2003). Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences*, *100*(6), 3107–3112. https://doi.org/10.1073/pnas.0634629100

Wee, P., & Wang, Z. (2017). Epidermal Growth Factor Receptor Cell Proliferation Signaling Pathways. *Cancers*, *9*(5), 52. https://doi.org/10.3390/cancers9050052

Wei, L. L., Gonzalez-Aller, C., Wood, W. M., Miller, L. A., & Horwitz, K. B. (1990). 5′-Heterogeneity in Human Progesterone Receptor Transcripts Predicts a New Amino-Terminal Truncated “C”-Receptor and Unique A-Receptor Messages.

Molecular Endocrinology, *4*(12), 1833–1840. https://doi.org/10.1210/mend-4-12-1833

Wei, L. L., Hawkins, P., Baker, C., Norris, B., Sheridan, P. L., & Quinn, P. G. (1996). An amino-terminal truncated progesterone receptor isoform, PRc, enhances progestin-induced transcriptional activity. *Molecular Endocrinology*, *10*(11), 1379–1387. https://doi.org/10.1210/mend.10.11.8923464

Weikum, E. R., Liu, X., & Ortlund, E. A. (2018). The nuclear receptor superfamily: A structural perspective. *Protein Science*, *27*(11), 1876–1892. https://doi.org/10.1002/pro.3496

Wenger, A., Karlsson, I., Kling, T., & Carén, H. (2023). CRISPR-Cas9 knockout screen identifies novel treatment targets in childhood high-grade glioma. *Clinical Epigenetics*, *15*(1), 80. https://doi.org/10.1186/s13148-023-01498-6

West, D. C., Kocherginsky, M., Tonsing-Carter, E. Y., Dolcen, D. N., Hosfield, D. J., Lastra, R. R., Sinnwell, J. P., Thompson, K. J., Bowie, K. R., Harkless, R. V., Skor, M. N., Pierce, C. F., Styke, S. C., Kim, C. R., De Wet, L., Greene, G. L., Boughey, J. C., Goetz, M. P., Kalari, K. R., … Conzen, S. D. (2018). Discovery of a Glucocorticoid Receptor (GR) Activity Signature Using Selective GR Antagonism in ER-Negative Breast Cancer. *Clinical Cancer Research*, *24*(14), 3433–3446. https://doi.org/10.1158/1078-0432.CCR-17-2793

Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software*, *40*(1), 1–29. https://doi.org/10.18637/jss.v040.i01

Wickham, H. (2016). *Ggplot2*. Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation* (R package version 1.1.2) [Computer software]. https://CRAN.R-project.org/package=dplyr

Wiegratz, I., & Kuhl, H. (2004). Progestogen therapies: Differences in clinical effects? *Trends in Endocrinology and Metabolism*, *15*(6), 277–285. https://doi.org/10.1016/j.tem.2004.06.006

Wijayasiri, P., Astbury, S., Kaye, P., Oakley, F., Alexander, G. J., Kendall, T. J., & Aravinthan, A. D. (2022). Role of Hepatocyte Senescence in the Activation of Hepatic Stellate Cells and Liver Fibrosis Progression. *Cells*, *11*(14), 2221. https://doi.org/10.3390/cells11142221

Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C. J., Rogers, J., & Bähler, J. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*, *453*(7199), 1239–1243. https://doi.org/10.1038/nature07002

Willebrand, R., & Kleinewietfeld, M. (2018). The role of salt for immune cell function and disease. *Immunology*, *154*(3), 346–353. https://doi.org/10.1111/imm.12915

Wilson, A. C., Carlson, S. S., & White, T. J. (1977). Biochemical Evolution. *Annual Review of Biochemistry*, *46*(1), 573–639. https://doi.org/10.1146/annurev.bi.46.070177.003041

Wolf, S., Melo, D., Garske, K. M., Pallares, L. F., Lea, A. J., & Ayroles, J. F. (2023). Characterizing the landscape of gene expression variance in humans. *PLOS Genetics*, *19*(7), e1010833. https://doi.org/10.1371/journal.pgen.1010833

World Health Organization. (2023). *Global breast cancer initiative implementation framework: Assessing, strengthening and scaling-up of services for the early detection and management of breast cancer. Executive summary.* https://www.who.int/publications/i/item/9789240067134

World Health Organization: Geneva, Switzerland. (2022). *Cancer*. https://www.who.int/news-room/fact-sheets/detail/cancer

Wu, C., Zhu, J., & Zhang, X. (2012). Integrating gene expression and protein-protein interaction network to prioritize cancer-associated genes. *BMC Bioinformatics*, *13*(1), 182. https://doi.org/10.1186/1471-2105-13-182

Wu, G., Diaz, A. K., Paugh, B. S., Rankin, S. L., Ju, B., Li, Y., Zhu, X., Qu, C., Chen, X., Zhang, J., Easton, J., Edmonson, M., Ma, X., Lu, C., Nagahawatte, P., Hedlund, E., Rusch, M., Pounds, S., Lin, T., … Baker, S. J. (2014). The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma. *Nature Genetics*, *46*(5), 444–450. https://doi.org/10.1038/ng.2938

Wu, K., Xu, H., Tian, Y., Yuan, X., Wu, H., Liu, Q., & Pestell, R. (2015). The role of CD44 in epithelial–mesenchymal transition and cancer development. *OncoTargets and Therapy*, 3783. https://doi.org/10.2147/OTT.S95470

Wu, P., Sun, R., Fahira, A., Chen, Y., Jiangzhou, H., Wang, K., Yang, Q., Dai, Y., Pan, D., Shi, Y., & Wang, Z. (2023). DROEG: A method for cancer drug response prediction based on omics and essential genes integration. *Briefings in Bioinformatics*, *24*(2), bbad003. https://doi.org/10.1093/bib/bbad003

Wu, P.-T., Su, W.-R., Li, C.-L., Hsieh, J.-L., Ma, C.-H., Wu, C.-L., Kuo, L.-C., Jou, I.-M., & Chen, S.-Y. (2019). Inhibition of CD44 induces apoptosis, inflammation, and matrix metalloproteinase expression in tendinopathy. *Journal of Biological Chemistry*, *294*(52), 20177–20184. https://doi.org/10.1074/jbc.RA119.009675

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*, *2*(3), 100141. https://doi.org/10.1016/j.xinn.2021.100141

Wu, W., Warner, M., Wang, L., He, W.-W., Zhao, R., Guan, X., Botero, C., Huang, B., Ion, C., Coombes, C., & Gustafsson, J.-A. (2021). Drivers and suppressors of triple-negative breast cancer. *Proceedings of the National Academy of Sciences*, *118*(33), e2104162118. https://doi.org/10.1073/pnas.2104162118

Xie, Y., Cheng, J., & Tan, X. (2023). *DT: A Wrapper of the JavaScript Library "DataTables"* (R package version 0.28) [Computer software]. https://CRAN.R-project.org/package=DT

Xing, W., & Archer, T. K. (1998). Upstream stimulatory factors mediate estrogen receptor activation of the cathepsin D promoter. *Molecular Endocrinology (Baltimore, Md.)*, *12*(9), 1310–1321. https://doi.org/10.1210/mend.12.9.0159

Xiong, H., Chen, Z., Zheng, W., Sun, J., Fu, Q., Teng, R., Chen, J., Xie, S., Wang, L., Yu, X.-F., & Zhou, J. (2020). FKBP4 is a malignant indicator in luminal A subtype of breast cancer. *Journal of Cancer*, *11*(7), 1727–1736. https://doi.org/10.7150/jca.40982

Xu, R., Zhang, Y., Gu, L., Zheng, J., Cui, J., Dong, J., & Du, J. (2015). Arf6 regulates EGF-induced internalization of E-cadherin in breast cancer cells. *Cancer Cell International*, *15*(1), 11. https://doi.org/10.1186/s12935-015-0159-3

Xu, X., Murdoch, F. E., Curran, E. M., Welshons, W. V., & Fritsch, M. K. (2004). Transcription factor accessibility and histone acetylation of the progesterone receptor gene differs between parental MCF-7 cells and a subline that has lost progesterone receptor expression. *Gene*, *328*, 143–151. https://doi.org/10.1016/j.gene.2003.12.003

Yaari, G., Bolen, C. R., Thakar, J., & Kleinstein, S. H. (2013). Quantitative set analysis for gene expression: A method to quantify gene set differential expression

including gene-gene correlations. *Nucleic Acids Research*, *41*(18), e170–e170. https://doi.org/10.1093/nar/gkt660

Yager, J. D. (2000). Chapter 3: Endogenous Estrogens as Carcinogens Through Metabolic Activation. *JNCI Monographs*, *2000*(27), 67–73. https://doi.org/10.1093/oxfordjournals.jncimonographs.a024245

Yang, J., & Fuller, P. J. (2012). Interactions of the mineralocorticoid receptor – Within and without. *Molecular and Cellular Endocrinology*, *350*(2), 196–205. https://doi.org/10.1016/j.mce.2011.07.001

Yang, P., Yang, Z., Dong, Y., Yang, L., Peng, S., Yuan, L., Hu, X., Chen, S., Tang, H., Yang, X., Fan, D., Zhao, H., & Bao, G. (2023). Clusterin is a biomarker of breast cancer prognosis and correlated with immune microenvironment. *Translational Cancer Research*, *12*(1), 31–45. https://doi.org/10.21037/tcr-22-1882

Yang, S., Zheng, J., Xiao, X., Xu, T., Tang, W., Zhu, H., Yang, L., Zheng, S., Dong, K., Zhou, G., & Wang, Y. (2015). SOX2 promotes tumorigenicity and inhibits the differentiation of I-type neuroblastoma cells. *International Journal of Oncology*, *46*(1), 317–323. https://doi.org/10.3892/ijo.2014.2713

Yang, W. S., Moon, H.-G., Kim, H. S., Choi, E.-J., Yu, M.-H., Noh, D.-Y., & Lee, C. (2012). Proteomic Approach Reveals FKBP4 and S100A9 as Potential Prediction Markers of Therapeutic Response to Neoadjuvant Chemotherapy in Patients with Breast Cancer. *Journal of Proteome Research*, *11*(2), 1078–1088. https://doi.org/10.1021/pr2008187

Yankina, M. A., Saik, O. V., Ivanisenko, V. A., Demenkov, P. S., & Khusnutdinova, E. K. (2018). Evaluation of Prioritization Methods of Extrinsic Apoptotic Signaling Pathway Genes for Retrieval of the New Candidates Associated with Major Depressive Disorder. *Russian Journal of Genetics*, *54*(11), 1366–1374. https://doi.org/10.1134/S1022795418110170

Yao, X., Li, H., Chen, L., & Tan, L. P. (2022). UV-induced senescence of human dermal fibroblasts restrained by low-stiffness matrix by inhibiting NF-κB activation. *Engineered Regeneration*, *3*(4), 365–373. https://doi.org/10.1016/j.engreg.2022.08.002

Yaslianifard, S., Movahedi, M., Yaslianifard, S., & Mozhgani, S. (2023). The mirror like expression of genes involved in the FOXO signaling pathway could be effective in the pathogenesis of human lymphotropic virus type 1 (HTLV-1) through disruption of the downstream pathways. *BMC Research Notes*, *16*(1), 147. https://doi.org/10.1186/s13104-023-06423-x

Ye, C., Ho, D. J., Neri, M., Yang, C., Kulkarni, T., Randhawa, R., Henault, M., Mostacci, N., Farmer, P., Renner, S., Ihry, R., Mansur, L., Keller, C. G., McAllister, G., Hild, M., Jenkins, J., & Kaykas, A. (2018). DRUG-seq for miniaturized high-throughput transcriptome profiling in drug discovery. *Nature Communications*, *9*(1), 4307. https://doi.org/10.1038/s41467-018-06500-x

Yedier-Bayram, O., Gokbayrak, B., Kayabolen, A., Aksu, A. C., Cavga, A. D., Cingöz, A., Kala, E. Y., Karabiyik, G., Günsay, R., Esin, B., Morova, T., Uyulur, F., Syed, H., Philpott, M., Cribbs, A. P., Kung, S. H. Y., Lack, N. A., Onder, T. T., & Bagci-Onder, T. (2022). EPIKOL, a chromatin-focused CRISPR/Cas9-based screening platform, to identify cancer-specific epigenetic vulnerabilities. *Cell Death & Disease*, *13*(8), 710. https://doi.org/10.1038/s41419-022-05146-4

Yeo, G. S. H., Connie Hung, C.-C., Rochford, J., Keogh, J., Gray, J., Sivaramakrishnan, S., O'Rahilly, S., & Farooqi, I. S. (2004). A de novo mutation affecting human

TrkB associated with severe obesity and developmental delay. *Nature Neuroscience*, *7*(11), 1187–1189. https://doi.org/10.1038/nn1336

Yilmaz, A., Braverman-Gross, C., Bialer-Tsypin, A., Peretz, M., & Benvenisty, N. (2020). Mapping Gene Circuits Essential for Germ Layer Differentiation via Loss-of-Function Screens in Haploid Human Embryonic Stem Cells. *Cell Stem Cell*, *27*(4), 679-691.e6. https://doi.org/10.1016/j.stem.2020.06.023

Yilmaz, A., Peretz, M., Aharony, A., Sagi, I., & Benvenisty, N. (2018). Defining essential genes for human pluripotent stem cells by CRISPR-Cas9 screening in haploid cells. *Nature Cell Biology*, *20*(5), 610–619. https://doi.org/10.1038/s41556-018-0088-1

Yu, G., Wang, L.-G., & He, Q.-Y. (2015). ChIPseeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, *31*(14), 2382–2383. https://doi.org/10.1093/bioinformatics/btv145

Yu, L., Lin, N., Ye, Y., Zhou, S., Xu, Y., Chen, J., Zhuang, W., & Wang, Q. (2024). Prognostic and chemotherapeutic response prediction by proliferation essential gene signature: Investigating POLE2 in bladder cancer progression and cisplatin resistance. *Journal of Cancer*, *15*(6), 1734–1749. https://doi.org/10.7150/jca.93023

Yu, M., Shin, H.-S., Lee, H. K., Ryu, D.-R., Kim, S.-J., Choi, K.-B., & Kang, D.-H. (2015). Effect of aldosterone on epithelial-to-mesenchymal transition of human peritoneal mesothelial cells. *Kidney Research and Clinical Practice*, *34*(2), 83–92. https://doi.org/10.1016/j.krcp.2015.03.005

Yu, S., Kim, T., Yoo, K. H., & Kang, K. (2017). The T47D cell line is an ideal experimental model to elucidate the progesterone-specific effects of a luminal A subtype of breast cancer. *Biochemical and Biophysical Research Communications*, *486*(3), 752–758. https://doi.org/10.1016/j.bbrc.2017.03.114

Yu, X., Yu, G., & Wang, J. (2017). Clustering cancer gene expression data by projective clustering ensemble. *PLOS ONE*, *12*(2), e0171429. https://doi.org/10.1371/journal.pone.0171429

Yuan, B., Cheng, L., Gupta, K., Chiang, H.-C., Gupta, H. B., Sareddy, G. R., Wang, D., Lathrop, K., Elledge, R., Wang, P., McHardy, S., Vadlamudi, R., Curiel, T. J., Hu, Y., Ye, Q., & Li, R. (2106). Tyrosine phosphorylation regulates ERβ ubiquitination, protein turnover, and inhibition of breast cancer. *Oncotarget*, *7*(27), 42585–42597. https://doi.org/10.18632/oncotarget.10018

Yun, D., Wang, X., Wang, W., Ren, X., Li, J., Wang, X., Liang, J., Liu, J., Fan, J., Ren, X., Zhang, H., Shang, G., Sun, J., Chen, L., Li, T., Zhang, C., Yu, S., & Yang, X. (2022). A Novel Prognostic Signature Based on Glioma Essential Ferroptosis-Related Genes Predicts Clinical Outcomes and Indicates Treatment in Glioma. *Frontiers in Oncology*, *12*, 897702. https://doi.org/10.3389/fonc.2022.897702

Zhai, J., Tang, Y., Yuan, H., Wang, L., Shang, H., & Ma, C. (2016). A Meta-Analysis Based Method for Prioritizing Candidate Genes Involved in a Pre-specific Function. *Frontiers in Plant Science*, *7*. https://doi.org/10.3389/fpls.2016.01914

Zhan, T., & Boutros, M. (2016). Towards a compendium of essential genes – From model organisms to synthetic lethality in cancer cells. *Critical Reviews in Biochemistry and Molecular Biology*, *51*(2), 74–85. https://doi.org/10.3109/10409238.2015.1117053

Zhang, L., Qiang, J., Yang, X., Wang, D., Rehman, A. U., He, X., Chen, W., Sheng, D., Zhou, L., Jiang, Y., Li, T., Du, Y., Feng, J., Hu, X., Zhang, J., Hu, X., Shao, Z., & Liu, S. (2020). IL1R2 Blockade Suppresses Breast Tumorigenesis and

Progression by Impairing USP15-Dependent BMI1 Stability. *Advanced Science*, *7*(1), 1901728. https://doi.org/10.1002/advs.201901728

Zhang, W., Quevedo, J., & Fries, G. R. (2021). Essential genes from genome-wide screenings as a resource for neuropsychiatric disorders gene discovery. *Translational Psychiatry*, *11*(1), 317. https://doi.org/10.1038/s41398-021-01447-y

Zhang, X., Zhao, M., Huang, A., Fei, Z., Zhang, W., & Wang, X. (2005). The effect of cyclin D expression on cell proliferation in human gliomas. *Journal of Clinical Neuroscience*, *12*(2), 166–168. https://doi.org/10.1016/j.jocn.2004.03.036

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137. https://doi.org/10.1186/gb-2008-9-9-r137

Zhang, Y., Pan, Y., Bian, Z., Chen, P., Zhu, S., Gu, H., Guo, L., & Hu, C. (2016). Ceramide Production Mediates Aldosterone-Induced Human Umbilical Vein Endothelial Cell (HUVEC) Damages. *PLOS ONE*, *11*(1), e0146944. https://doi.org/10.1371/journal.pone.0146944

Zhang, Y.-H., Huang, T., Chen, L., Xu, Y., Hu, Y., Hu, L.-D., Cai, Y., & Kong, X. (2017). Identifying and analyzing different cancer subtypes using RNA-seq data of blood platelets. *Oncotarget*, *8*(50), 87494–87511. https://doi.org/10.18632/oncotarget.20903

Zhao, Q., Zheng, G., Yang, K., Ao, Y., Su, X., Li, Y., & Lv, X. (2016). The clock gene *PER1* plays an important role in regulating the clock gene network in human oral squamous cell carcinoma cells. *Oncotarget*, *7*(43), 70290–70302. https://doi.org/10.18632/oncotarget.11844

Zheng, R., Wan, C., Mei, S., Qin, Q., Wu, Q., Sun, H., Chen, C.-H., Brown, M., Zhang, X., Meyer, C. A., & Liu, X. S. (2019). Cistrome Data Browser: Expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Research*, *47*(D1), D729–D735. https://doi.org/10.1093/nar/gky1094

Zhou, M., Sareddy, G. R., Li, M., Liu, J., Luo, Y., Venkata, P. P., Viswanadhapalli, S., Tekmal, R. R., Brenner, A., & Vadlamudi, R. K. (2019). Estrogen receptor beta enhances chemotherapy response of GBM cells by down regulating DNA damage response pathways. *Scientific Reports*, *9*(1), 6124. https://doi.org/10.1038/s41598-019-42313-8

Zhou, Y., Kuang, Y., Wang, C., Yu, Y., Pan, L., & Hu, X. (2024). Impact of KRAS mutation on the tumor microenvironment in colorectal cancer. *International Journal of Biological Sciences*, *20*(5), 1947–1964. https://doi.org/10.7150/ijbs.88779

Zhu, Y., Bond, J., & Thomas, P. (2003). Identification, classification, and partial characterization of genes in humans and other vertebrates homologous to a fish membrane progestin receptor. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(5), 2237–2242. https://doi.org/10.1073/pnas.0436133100

Ziegler, D. V., Czarnecka-Herok, J., Vernier, M., Scholtes, C., Camprubi, C., Huna, A., Massemin, A., Griveau, A., Machon, C., Guitton, J., Rieusset, J., Vigneron, A. M., Giguère, V., Martin, N., & Bernard, D. (2024). Cholesterol biosynthetic pathway induces cellular senescence through ERRα. *Npj Aging*, *10*(1), 5. https://doi.org/10.1038/s41514-023-00128-y

Ziemann, M., Kaspi, A., & El-Osta, A. (2019). Digital expression explorer 2: A repository of uniformly processed RNA sequencing data. *GigaScience*, *8*(4). https://doi.org/10.1093/gigascience/giz022

Zong, S., Jiao, Y., Liu, X., Mu, W., Yuan, X., Qu, Y., Xia, Y., Liu, S., Sun, H., Wang, L., Cui, B., Liu, X., Li, P., & Zhao, Y. (2021). FKBP4 integrates FKBP4/Hsp90/IKK with FKBP4/Hsp70/RelA complex to promote lung adenocarcinoma progression via IKK/NF-κB signaling. *Cell Death & Disease*, *12*(6), 602. https://doi.org/10.1038/s41419-021-03857-8

Zullo, J., Matsumoto, K., Xavier, S., Ratliff, B., & Goligorsky, M. S. (2015). The cell secretome, a mediator of cell-to-cell communication. *Prostaglandins & Other Lipid Mediators*, *120*, 17–20. https://doi.org/10.1016/j.prostaglandins.2015.03.012

COPY RIGHT PERMISSIONS

Figure 2. Mechanisms of cellular senescence and its consequences

(This image was created by using BioRender.com using basic account).

License and Usage:

		Academic		
Use Case	Free	Academic Individual License	Academic Lab License	Academic Institutional License
Publishing uses				
Journal publication		✅	✅	✅
Journal Cover		✅	✅	✅
Published thesis		✅	✅	✅
Not-for-profit website		✅	✅	✅
Textbook figures, 5 or less		✅	✅	✅
Textbook figures, greater than 5 figures, or less than and equal to 15			✅	✅
Textbook figures, more than 15	Contact us	-------	-------	-------
Personal YouTube / social media (no profit)	✅	✅	✅	✅
Unpublished thesis	✅	✅	✅	✅
Conference poster (academic user)	✅	✅	✅	✅
Grant application	✅	✅	✅	✅
Academic Poster	✅	✅	✅	✅
Internal Meetings (lab or team)	✅	✅	✅	✅
Conference Presentation (Academic Audience)	✅	✅	✅	✅

Figure 4. Mechanisms for estrogen signaling. In the graph, estrogen, or E2 (orange circle E), binds to ERα/ERβ and GPER1 to exert its genomic and non-genomic effects. Orange highlights the genomic impact, which is caused by the E2-receptor complex binding to EREs as it enters the nucleus. The non-genomic impact is displayed in blue. E2 binds to ERs in the membrane-like GPER1 and controls expression by adjusting the activation of associated enzymes or the opening of ion channels. E; estrogen or E2, ERα; estrogen receptor alpha, ERβ; estrogen receptor beta, GPER1; G protein-coupled estrogen receptor 1, ERE; estrogen response elements.

(This image was reproducible without the need of copy right permission and required citation and it was obtained from Chen, P., Li, B., & Ou-Yang, L. (2022). Role of estrogen receptors in health and disease. *Frontiers in endocrinology*, 13, 839005. https://doi.org/10.3389/fendo.2022.839005).

Journal List > Front Endocrinol (Lausanne) > PMC9433670

As a library, NLM provides access to scientific literature. Inclusion in an NLM database does not imply endorsement of, or agreement with, the contents by NLM or the National Institutes of Health.
Learn more: PMC Disclaimer | PMC Copyright Notice



Front Endocrinol (Lausanne). 2022; 13: 839005. PMCID: PMC9433670
Published online 2022 Aug 18. doi: 10.3389/fendo.2022.839005 PMID: 36060947

Role of estrogen receptors in health and disease

Peng Chen, Bo Li,* and Ling Ou-Yang

Author information ▸ Article notes ▾ Copyright and License information PMC Disclaimer

Copyright © 2022 Chen, Li and Ou-Yang

This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Journal Article

Figure 5. Classical and non-classical signaling of progesterone. A) Classical signalling of progesterone is mediatied by the diffusion of hormone through membrane and binding to receptor inside the cell. B) Rapid response in non-classical mechanism is mediated through the activatiton of c-Src, MAPK and AKT. C) The non-classical signaling can be initated by the binding of progesterone to mPRs and PGRMC1.

(This image was reproducible without the need of copy right permission and required citation and it was obtained from Pedroza, D. A., Subramani, R., & Lakshmanaswamy, R. (2020). Classical and Non-Classical Progesterone Signaling in Breast Cancers. Cancers, 12(9), 2440. https://doi.org/10.3390/cancers12092440).

Journal List > Cancers (Basel) > PMC7563480

As a library, NLM provides access to scientific literature. Inclusion in an NLM database does not imply endorsement of, or agreement with, the contents by NLM or the National Institutes of Health.
Learn more: PMC Disclaimer | PMC Copyright Notice



Cancers (Basel). 2020 Sep; 12(9): 2440. PMCID: PMC7563480
Published online 2020 Aug 27. doi: 10.3390/cancers12092440 PMID: 32867363

Journal Article

Classical and Non-Classical Progesterone Signaling in Breast Cancers

Diego A. Pedroza,[1] Ramadevi Subramani,[1,2] and Rajkumar Lakshmanaswamy[1,2,*]

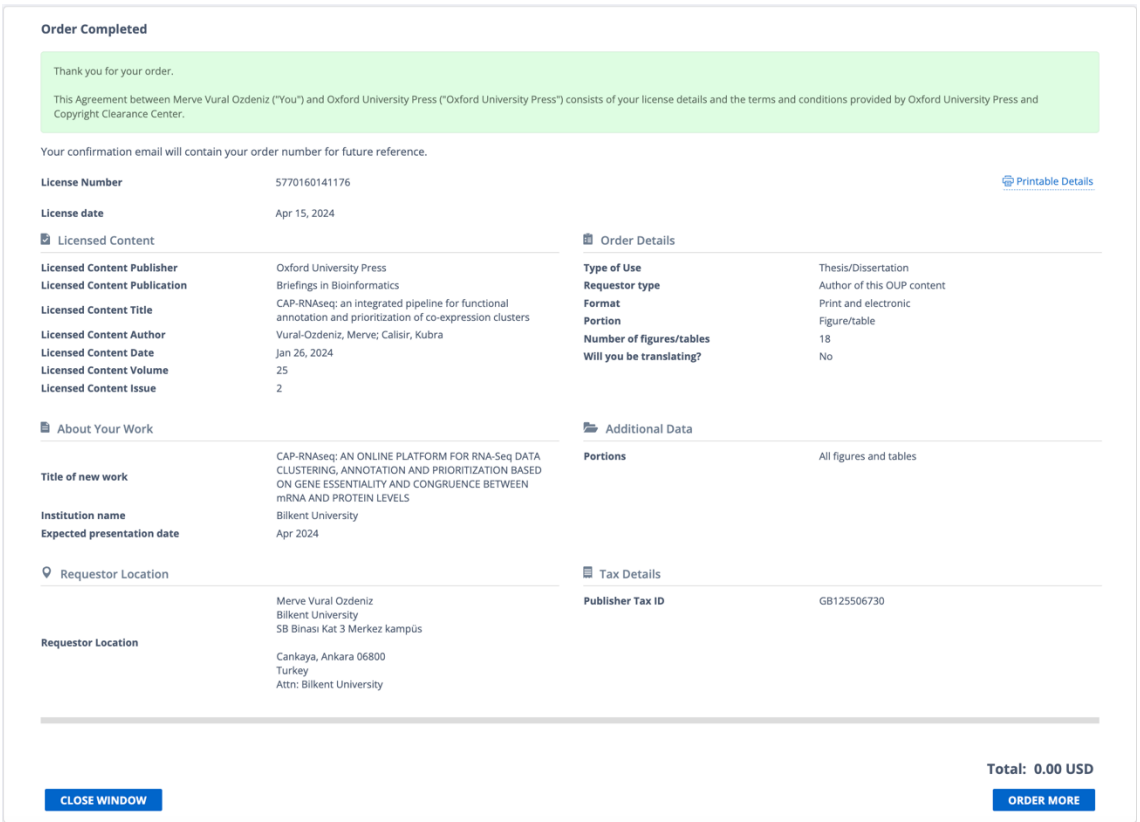
Author information Article notes Copyright and License information PMC Disclaimer

Copyright © 2020 by the authors.

Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

Figures 6, 7, 8, 9, 17, 19, 23 and 24 and Tables 13 and 14 were used under the copy right permission of Oxford University Press journal.

(The screenshots and tables were adapted from Vural-Ozdeniz, M., Calisir, K., Acar, R., Yavuz, A., Ozgur, M. M., Dalgıc, E., & Konu, O. (2024). CAP-RNAseq: an integrated pipeline for functional annotation and prioritization of co-expression clusters. Briefings in bioinformatics, 25(2), bbad536. https://doi.org/10.1093/bib/bbad536)



Order Completed

Thank you for your order.

This Agreement between Merve Vural Ozdeniz ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

Printable Details

License Number	5770160141176
License date	Apr 15, 2024

Licensed Content

Licensed Content Publisher	Oxford University Press
Licensed Content Publication	Briefings in Bioinformatics
Licensed Content Title	CAP-RNAseq: an integrated pipeline for functional annotation and prioritization of co-expression clusters
Licensed Content Author	Vural-Ozdeniz, Merve; Calisir, Kubra
Licensed Content Date	Jan 26, 2024
Licensed Content Volume	25
Licensed Content Issue	2

Order Details

Type of Use	Thesis/Dissertation
Requestor type	Author of this OUP content
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	18
Will you be translating?	No

About Your Work

Title of new work	CAP-RNAseq: AN ONLINE PLATFORM FOR RNA-Seq DATA CLUSTERING, ANNOTATION AND PRIORITIZATION BASED ON GENE ESSENTIALITY AND CONGRUENCE BETWEEN mRNA AND PROTEIN LEVELS
Institution name	Bilkent University
Expected presentation date	Apr 2024

Additional Data

Portions	All figures and tables

Requestor Location

Requestor Location	Merve Vural Ozdeniz Bilkent University SB Binası Kat 3 Merkez kampüs Cankaya, Ankara 06800 Turkey Attn: Bilkent University

Tax Details

Publisher Tax ID	GB125506730

Total: 0.00 USD

CLOSE WINDOW

ORDER MORE

Figures 10-15, 18, 20-22, 25-28, 30-50 and 52-61, and Tables 12, 14, 15, 16, 18, 20-27 and 30 were retrieved from http://konulabapps.bilkent.edu.tr:3838/CAPRNAseq/ published as Vural-Ozdeniz, M., Calisir, K., Acar, R., Yavuz, A., Ozgur, M. M., Dalgıc, E., & Konu, O. (2024). CAP-RNAseq: an integrated pipeline for functional annotation and prioritization of co-expression clusters. Briefings in bioinformatics, 25(2), bbad536. https://doi.org/10.1093/bib/bbad536)