

TP53 META: A webtool to visualize effects of TP53
modulators and mutations on expression profiles with a
focus on breast cancer

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF ENGINEERING AND
SCIENCE
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR
THE DEGREE OF
MASTER OF SCIENCE
IN
MOLECULAR BIOLOGY AND GENETICS

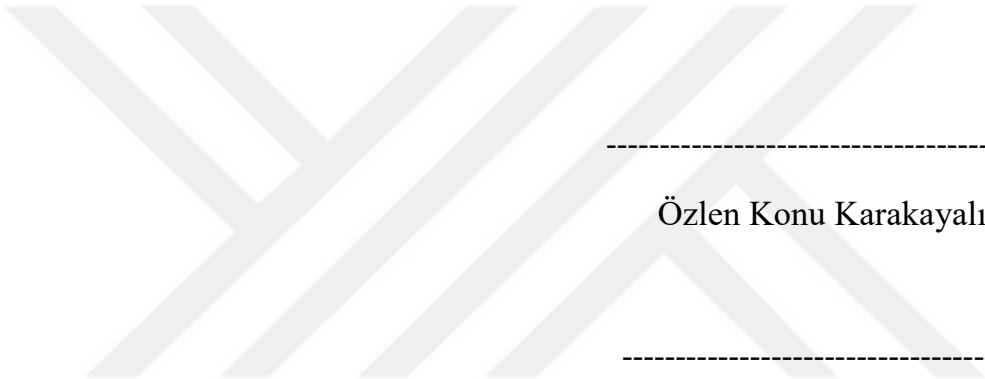
By
Abdul Moiz Aftab
May 2024

TP53 META: A webtool to visualize effects of TP53 modulators and mutations on expression profiles with a focus on breast cancer

By Abdul Moiz Aftab

May 2024

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



Özlen Konu Karakayalı (Advisor)

Ilyas Chachoua

Ceren Sucularlı

Approved for the Graduate School of Engineering and Science

Orhan Arıkan

Director of the Graduate School of Engineering and Science

ABSTRACT

TP53 META, A webtool to visualize effects of TP53 modulators and mutations on expression profiles with a focus on breast cancer

Abdul Moiz Aftab

M.Sc. in Molecular Biology and Genetics

Advisor: Özlen Konu Karakayalı

May 2024

Compilation and comparison of multiple datasets that relate to a particular treatment can help us understand more about a scientific question of interest. Transcription factor protein 53 (*TP53*), also known as the “Guardian of the genome”, is one of the most mutated transcription factors in sporadic Breast Cancer. Even though there are multiple online databases that compile, annotate and list p53 targets, these gene lists can vary significantly and no web-tool is available to visualize them and perform a meta-analysis of their expressions. Herein, I have developed an R Shiny based tool called TP53 META that allows for comparison and visualization of up to four separate, two group RNA-seq datasets. TP53 META is able to normalize and perform differential expression analysis via limma or DESeq2 on different datasets, simultaneously. Moreover, it allows users to perform clustering, gene-set enrichment analyses, transcription factor-target enrichment, as well as visualization by using disease-gene and treatment-gene bipartite networks that can be generated for further understanding of individual or combination treatments. It is also possible to limit the use of TP53 META modules to a subset of the transcriptome, e.g., specific targets of p53 such as DREAM targets. TP53 META also calculates a meta-p-value using different available meta-analysis statistical methods and incorporates several public as well as in house datasets, most of which are of breast cancer cell lines, in which *TP53* is modulated by siRNA treatments, inducers such as nutlin, doxorubicin, *CHRNA5* depletion, or stable mutations in the heterozygous or

homozygous background. Public datasets were acquired from GEO while for in house datasets, I have obtained the fastq files and converted to counts and integrated them on TP53 META web-tool and tested the following hypotheses to demonstrate the functionality of the web-tool:

- 1) Does *CHRNA5* depletion, a known *TP53* inducer, result in further downregulation of DREAM targets in MCF7 with the wildtype *TP53* or heterozygous *TP53* mutant MCF7 cells?;
- 2) Do *TP53* inducers induce *TP53* targets equally in wild type and mutant cancer cells?

I also applied existing and novel methods to determine synergistic/antagonistic relationship of genes between the two treatments, e.g., doxorubicin alone or doxorubicin together with *CHRNA5* depletion. Results showed that wildtype homozygous *TP53* overexpression depletes DREAM targets more than heterozygous mutant MCF7 cells where a copy of mutant *TP53* (R175H, R273H) exists. I have also found that *CHRNA5* depletion in combination with doxorubicin can further regulate certain *TP53* targets.

ÖZET

TP53 META: TP53 modölatörleri ve mutasyonlarının meme kanseri odağında gen ifade profillerine etkilerini görselleştiren bir web aracı

Abdul Moiz Aftab

Moleküler Biyoloji ve Genetik Yüksek Lisans Programı

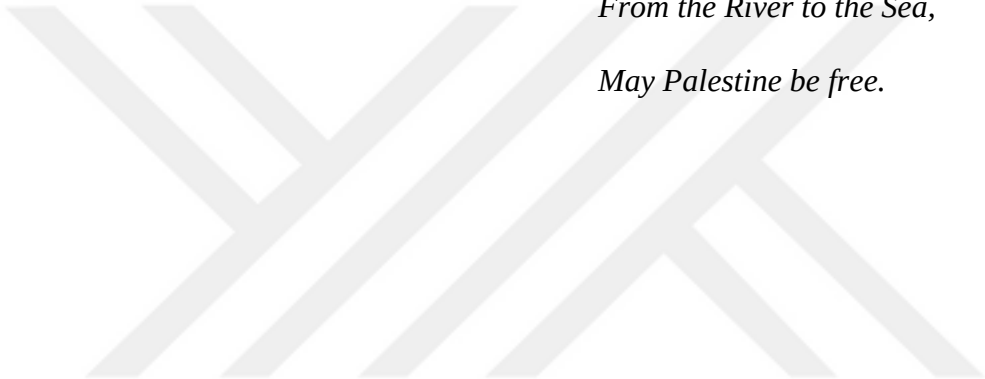
Tez Danışmanı: Özlen Konu Karakayalı

Mayıs, 2024

Belirli bir tedavi ile ilişkili veri setlerinin derlenmesi ve karşılaştırılması, bilimsel bir soruya ilişkin anlayışımıza katkıda bulunabilir. Transkripsiyon faktör protein 53 (TP53), diğer adıyla “Genomun Gardiyanı”, sporadik meme kanserinde en çok mutasyona uğrayan transkripsiyon faktörlerinden biridir. TP53 hedef genlerini derleyen, açıklayan ve listeleyen birçok web aracı bulunsa da, bu gen listeleri önemli derecede farklılık gösterebilmekte ve bu listelerin görselleştirilmesini ve meta analizinin yapılmasını sağlayan bir web aracı bulunmamaktadır. Bu bağlamda, R Shiny tabanlı, TP53 META adında, iki koşullu RNA Dizileme veri setlerinden dört farklı setin eş zamanlı karşılaştırılmasına kadar çıkabilen analizlere ve bu analizlerin görselleştirilmesine olanak sağlayan bir araç geliştirdim. TP53 META, seçilen veri setlerinin normalizasyonunu ve limma veya DESeq2 ile diferansiyel ekspresyon analizini bir arada yapabilmektedir. Dahası, bu araç kullanıcıların kümeleme, gen seti zenginleştirme ve transkripsiyon faktörü-hedef zenginleştirme analizleri yapabilmelerini sağlamakta ve bunun yanı sıra ilgili tekli veya kombinasyon tedavilerine ilişkin anlayışımızı güçlendirebilecek hastalık-gen ve tedavi-gen bipartit yapı ağlarının görselleştirmesini yapabilmektedir. TP53 META'nın modülleri transkriptomun yalnızca belirli bir alt kümesine, örneğin TP53 proteininin özelleşmiş hedef genlerinden DREAM hedef genleri listesine, sınırlı tutularak da kullanılabilir. Bir diğer yandan TP53 META, mevcut meta-analiz istatistiksel

metotlarından yararlanarak meta-p değeri hesaplamaktadır ve çoğunluğu TP53 ifadesinin; siRNA tedavileri, nutlin ve doxorubicin gibi indükleyiciler ile tedavileri, *CHRNA5* depleasyonu ve stabil homozigot veya heterozigot mutasyonlar üzerine modüle olmuş olduğu meme kanseri hücre hatlarına ait verilerden oluşan hem genel kullanıma açık hem de Konu Lab'a ait veri setlerini içermektedir. Genel kullanıma açık veri setleri GEO veri tabanından alınarak, Konu Lab'a ait veri setleri de fastq okuma dosyalarından ham sayma verilerine çevrilerek TP53 META web aracına entegre edilmiştir ve akabindeki hipotezler bu web-aracının işlevselliğini gösterebilmek için test edilmiştir: 1) Literatürde TP53 indükleyicisi olarak bilinen *CHRNA5*'in depleasyonu MCF7 hücre hattında DREAM hedef genlerinin ilerleyen kaybına yabancı TP53 varlığında mı yoksa heterozigot TP53 mutasyonları varlığında mı yol açmaktadır?; 2) TP53 indükleyicileri TP53 hedef genlerini yabancı ve mutant kanser hücre hatlarında eş seviyelerde mi indüklemektedir?

Bu tezde aynı zamanda mevcut ve özgün metotları genlerin iki tedavi, örneğin tekli doxorubicin tedavisi ile *CHRNA5* depleasyonu durumundaki doxorubicin tedavisi, etkisindeki sinerjistik veya antagonistik ilişkilerinin saptanabilmesi için kullandım. Çıkan sonuçlar MCF7 hücrelerinde yabancı homozigot TP53 yüksek ifadesinin, heterozigot mutant (R175H, R273H mutasyonlarını içeren bir kopyalı) TP53'ün ifade edildiği koşula kıyasla DREAM hedef genlerini daha çok azalttığını göstermektedir. *CHRNA5* depleasyonunun doxorubicin tedavisi ile kombinasyonunda TP53 hedef genlerinin daha çok etkilendiği de gösterilmiştir.



Dedicated to the scientists of Gaza,

From the River to the Sea,

May Palestine be free.

Acknowledgements

Firstly, I would like to express my gratitude towards my advisor Assoc. Prof. Özlen Konu Karakayalı, for her guidance and her support throughout my Master's studies. I am very grateful to her for giving me the opportunity to be a part of Konu lab, first as an intern during my undergraduate years and then as a graduate student.

I sincerely thank my jury members, Asst. Prof. Ilyas Chachoua and Assoc. Prof. Ceren Sicularlı from Hacettepe University, for taking out the time to read my thesis and for their invaluable input and suggestions. I would also like to thank Ilyas hoca for the candid conversations we had in his office and for sharing his experience and knowledge with me.

I would also like to acknowledge Bilkent University, which has been like a second home to me for the past seven years. I would like to thank the University and the MBG department for providing me with a scholarship throughout my studies.

I acknowledge that the research in this thesis was funded by the Scientific and Technological Research Council of Turkey (TUBITAK) with grant number 219Z029, and that I have received a scholarship from TUBITAK 219Z029. I would like to thank the members of this project especially, Merve Vural, Çağdaş Yanyatan, Burçin İrem Arıcı and Eren kumak for their contributions and help during this project.

I am very glad to be part of the Konu lab and very grateful for all the colleagues and friends I made during my time at the lab. I would like to thank Melike Tombaz and Ronaldo Leka for teaching and helping me during the start of my MSc studies. Both of them have been great mentors for me. I would like to thank all the previous and current members of our lab, Mert,

Rana, Emel, Muazzez, Merve, Siber, Güneş, Çağdaş, Nil, Betul, Zehra, Seda, Selvin and Algi for providing the best lab environment possible and most importantly for their friendship. I would also like to thank Tunahan for his friendship and being such an understanding and kind roommate.

Finally and most importantly, I would like to thank my parents for their constant love, support and for always believing in me. Without their support and sacrifices, this would have not been possible. I would also like to thank my siblings, especially Aqdas for their constant guidance and support. Lastly, I thank Mariam Hannan for being by my side and for being my support system throughout this journey.

Table of Contents

List of Figures	xiii
List of Tables.....	xviii
List of Appendices.....	xix
1. Introduction	1
1.1 Overview of Breast Cancer and its Mechanisms of Development.....	1
1.1.1 Genetic Aberrations in Breast Cancers	1
1.1.2 Molecular Sub-types of Breast Cancer	2
1.1.3 Drug resistance in Breast Cancer	3
1.1.4 MCF7 cell line in Breast Cancer.....	5
1.2 The role of <i>TP53</i> in Breast Cancer.....	6
1.2.1 Function of <i>TP53</i> in cancer progression.....	6
1.2.2 DREAM complex and <i>TP53</i>	7
1.2.3 LIN37 and DREAM complex.....	8
1.3 <i>TP53</i> Mutations: R175H and R273H.....	9
1.3.1 Molecular Impact of R175H mutation.....	10
1.3.2 Molecular Impact of R273H mutation.....	11
1.3.3 Clinical significance of <i>TP53</i> mutations.....	12
1.4 <i>CHRNA5</i> and its role in Cancer	12
1.5 Differential effects of Doxorubicin, Nutlin, RITA and PRIMA-1.....	14
1.5.1 Mechanisms of Action.....	14
1.5.2 Comparative Analysis and Therapeutic Efficacy.....	16
1.6 Combinatorial effects of drugs and Synergism.....	17
1.7 Use of Web Applications in Transcriptomics.....	19
1.7.1 Web applications for <i>TP53</i> datasets and/or RNA-seq data analysis.....	19
1.7.2 Meta-analysis tools for transcriptomic or biological data.....	24
1.8 Statistical methods in RNA-seq	25
1.8.1 DEseq2 vs Limma	25
1.8.2 Batch Correction	26
1.8.3 Meta-analysis of Datasets	27
2. Aims and Rationale	30
3. Materials and Methods	32

3.1 RNA-seq analyses	32
3.1.1 Data acquisition	32
3.1.2 RNA-seq data processing - QC, Alignment and Quantification.....	36
3.1.2.1 Mutant samples (Data 2)	36
3.1.2.2 Doxorubicin and si <i>CHRNA5</i> treated samples (Data 3)	38
3.2 Transcriptomic data exploration.....	40
3.2.1 DGE analysis.....	40
3.2.2 Downstream functional analysis.....	41
3.3 Development of TP53-META application.....	43
3.4 Synergy calculations.....	44
4. Results	45
4.1 TP53-META shiny web server.....	45
4.1.1 Initialization and choosing contrasts	46
4.1.2 Data summary.....	47
4.1.3 Gene selection.....	49
4.1.4 Scatter plot and Heatmap.....	51
4.1.5 Functional Profile.....	54
4.1.6 DoRothEA – Transcription factor enrichment.....	60
4.1.7 Treatment-Gene networks.....	62
4.1.8 MSigDB and Enrichment clusters.....	64
4.2 Analyses of in house RNA-seq data.....	66
4.2.1 Mutant cell lines.....	66
4.2.2 Doxorubicin and si <i>CHRNA5</i> treated Mutant cell lines.....	77
4.2.2.1 Exploratory data analysis.....	77
4.2.2.2 EV contrasts.....	78
4.2.2.3 R175H contrasts.....	87
4.2.2.4 R273H contrasts.....	99
4.2.2.5 WT contrasts.....	104
4.2.2.6 Effect of combinatorial in-between the cell lines.....	112
4.2.2.7 Cell line status specific signature.....	122
4.2.2.8 Confirmation of our mutant cell line with online datasets.....	124
4.3 Synergy Method Calculations.....	126
5. Conclusion and Discussion.....	133
5.1 TP53 META web server.....	133

5.2 Batch effect correction for in-house datasets.....	134
5.3 Combinatorial effects of doxorubicin and si <i>CHRNA5</i>	135
5.3.1 Transcriptomic regulation of the isogenic Mutant cell lines.....	136
5.3.2 Transcriptomic regulation of the combinatorial treatments in the Mutant cell lines.....	139
5.3.2.1 R175H mutant.....	140
5.3.2.2 R273H mutant.....	142
5.3.2.3 Combinatorial effects across cell lines.....	143
5.3.2.4 Difference in response of the mutant cell lines to the combinatorial treatment.....	144
5.3.2.5 Cell line status specific genes.....	144
5.3.2.6 Response to chemo-resistance.....	146
5.3.2.7 Confirmation of RNA-seq results through MTT assays.....	147
5.4 <i>CHRNA5</i> as a possible <i>TP53</i> target.....	148
5.5 Synergy Calculations.....	150
5.6 Future perspectives.....	150
6. References	153
7. Appendix	168

List of Figures

Figure 1.1 The change in transcription activation/repression due to the binding of the complexes to DNA during different phases of the cell cycle.	8
Figure 1.2 Prevalence and distribution of <i>TP53</i> mutations in Breast Cancer cohorts, taken from cBioportal. The datasets included were TCGA breast cancer datasets, METABRIC datasets, and MSK datasets (n = 9898). The three residues most commonly affected by mutations are R248, R175, and R273. The types of mutations and their associated color codes include: missense mutations (green), truncating mutations (black), in-frame mutations (red), and splice variants (orange).	10
Figure 3.1 A, B Sequence counts and mean quality scores for the samples in our mRNA-seq mutant dataset (dataset 2), as analyzed by fastQC and aggregated by multiQ. C Each FastQC section is marked to indicate if results are completely normal (green), somewhat abnormal (orange), or highly unusual (red).....	37
Figure 3.2 Sequence counts and mean quality scores for all the samples in our RNA-seq doxorubicin treated dataset (dataset 3), as analyzed by fastQC and aggregated by multiQC. A, B Sequence counts and mean quality scores of the first batch of the dataset. C, D Sequence counts and mean quality scores for the second batch of the dataset.	39
Figure 3.3 A, B All the fastQC checks presented here for all the dataset 3 (doxorubicin treated mutant cell lines). Green shows that that the filter is passed, while yellow indicates moderate quality and red shows low quality.	40
Figure 4.1 Simplified workflow of the TP53-META application.	46
Figure 4.2 Initialization, summary of the datasets and the tutorial page.	46
Figure 4.3 Data selection and contrast comparison (Treatment vs Control). The user also selects the method for DGE and the meta-analysis of P-values. Here, the three different GSE datasets are chosen and the meta-analysis is chosen as Bonferroni.	47
Figure 4.4 A Volcano plot for the chosen the contrast with log2FC plotted against the -log(P-values). The red colored genes fulfil the criteria of having a p-value less than 0.05 and logFC of over 1. These cutoffs can be changed along with the label size. B Library size bar plot for all of the samples present in the dataset. C , the respective PCA for the samples of the chosen dataset.	48
Figure 4.5 A Shows the options to select genes based on the predefined gene lists or an upload option. The table for the contrasts can be seen right underneath. The user has to select genes from this table to be used further in the analysis. B Shows the p-value and the logFC filters for upregulated selection and downregulated selection of genes across datasets. Using the check boxes, the user can choose which datasets to apply these particular filters to; for example, if a p-value cutoff of 0.05 is chosen and applied to only Data 1, the genes will be filtered only based on Data 1. However, if someone wants to choose significant genes across all contrasts, all the check boxes can be chosen.	50
Figure 4.6 Scatter plot for the logFC values of the chosen contrasts. Here the selected genes (DREAM targets) were plotted only instead of all the genes.	51
Figure 4.7 DREAM targets for contrasts from different datasets plotted against each other. This allows for transcriptomic comparison of different datasets with respect to each other. ...	52
Figure 4.8 Heatmap of the chosen contrasts and the MMB-FOXM1 geneset with “Euclidean” defined as the distance method for both rows and columns; and “complete” defined as the linkage method for both rows and columns. The colors are based on the logFC values with red representing values over 0 and blue representing negative values. The genes can be visualized at the left and the samples are clustered at the top. This is an interactive plot so the user can zoom in/out to clearly see the gene names and their effect.	53

Figure 4.9 **A** GOST plot of the selected genes. This is an interactive plot and the term can be viewed by clicking on the individual circle. **B** Table for the GO enrichment analysis results. It shows the overlap of the selected geneset with the terms present and the FDR values for the terms based on the overlap. 56

Figure 4.10 **A** Dot plot for the enriched terms. **B** Emapplot, shows the enriched terms as a network with terms with overlapping genes, clustered together. **C** Cnetplot, shows the enriched terms and their associated genes, with gene nodes colored based on the users selection. The figures above indicated that, the most significant enriched terms appear to be cell cycle pathway related. This is expected since we had chosen the downregulated genes for the *Nutlin vs Control* contrast (GSE47042). 57

Figure 4.11 **A** The disease gene association based on the chosen genes and logFC of one contrast (contrast 1 chosen here) are shown as a network. **B** This shows the parameters that can be changed by the user when plotting disease-gene associations in the DisGeNet plots. **C** Shows the enriched disease terms on the y-axis and the genes on the x-axis. 59

Figure 4.12 **A** Enriched Transcription factor terms based on the Dorothea database. **B** Shows the FDR value for each of the TF enriched with the geneset. The size shows the relative size of genes and the overlap with that TF. **C** Network of the TF can be viewed based on either TF number or the FDR. The edges show the number of overlapping genes between the TF and the node size shows the number of gene targets for that particular TF. 61

Figure 4.13 Selecting a particular TF, results in a heatmap of the gene targets of the particular TF. In the annotation part of the Heatmap, values show the confidence scores classifying regulons based on their quality (from A to E or NA). To better estimate TF activities, it is recommended to select regulons from the confidence levels A, B and C.....62

Figure 4.14 Bi-partite network of the 4 contrasts. Here only 49 genes were used for analysis instead of the whole geneset of the mmb-foxm1 targets as used previously. The edges show the logFC and the contrasts are represented with different shapes of nodes while the circles represent the genes.....63

Figure 4.15 **A**, Enrichment based on the terms present in the MSigDB terms. The collection of terms can be chosen from the left. **B**, **C** Similar to the Dorothea TF enrichment heatmap and network, here the user can select the enriched terms together to show its particular targets and visualize a network for the terms. Width of edges are represented by the GeneRatio column. Hovering over the node, shows the overlapping genes between the term and the genes the user selected for the analysis. 65

Figure 4.16 **A** PCA for the mutant cell line (only first batch) before batch correction. **B** PCA for the mutant cell line after batch correction (first and second batch combined). 67

Figure 4.17 **A-D**, Scatter plot of all the genes for the WT and mutant cell lines (R175H, R273H, R248W) vs EV contrasts, plotted from the separate batches against the Combatseq results for a comparative analysis. The Combatseq contrast is plotted on the x-axis. The slope and the R^2 values can be seen for each scatter plot. These results are based on Limma-voom. Most logFC values are grouped near the x-axis, where the Combatseq results are plotted, while on the same x-axis, some genes have very inflated results for all the contrasts. 68

Figure 4.18 Heatmap for the four contrasts, R175H, R273H, R248W, WT vs EV. Here 125 genes were selected. These were selected by filtering only genes that fulfil a threshold of 0.05 for the p-value across all the contrasts. 70

Figure 4.19 **A** Enrichment analysis based on the DisGeNET database. Fibroid tumor and Early stage carcinoma are two of the terms enriched for this R175H mutant geneset. **B** Enrichment based on Gene Ontology terms show some neuronal pathways along with the TGFB signaling pathway. **C** Heatmap for the chosen enriched terms from the MSigDB tab. The R175H contrast shows a higher downregulation for these target genes compared to the other mutations. 71

Figure 4.20 **A** Heatmap for all the 126 genes filtered for the *R273H* vs *EV* contrast. Column-wise clustering shows that *R273H* is separated from *R175H* and *WT*. **B** Transcription factors enriched for this gene set. **C** Heatmap of *MYC*, *EGR1* and *E2F1* TFs. *MYC* targets seem to be slightly upregulated compared to the other contrasts. **D** Enrichment of hallmark terms based on the MSigDB database. *NFKB*, *NGF* and Estrogen related pathways show up. These terms are very similar to the ones by the *R175H* contrast. 73

Figure 4.21 Heatmap for the Hallmark Estrogen terms and *E2F* targets, from the MSigDB database. The early response estrogen targets show a very significant upregulation for the *WT* contrast compared to very little or almost insignificant regulation for the other contrasts. Similarly, *E2F* targets are more downregulated for this contrast compared to the others. 74

Figure 4.22 **A** Heatmap of the *LIN37* dependent *p53* targets, plotted across all the contrasts. *WT* vs *EV* contrast shows a downregulation of most of the targets, which is normally expected. This is due to the regulation of the *LIN37* component of the DREAM complex by *p53*. However, the mutant cell lines clearly show a dysregulation of these targets. **B** The treatment-gene network confirms the genes with no significance with the other contrasts. The edges are only shown if the p-value is significant in that particular contrast. Here, diamond represents the *WT* vs *EV* contrast. 76

Figure 4.23 **A** PCA of the doxorubicin treated samples before Combatseq. The different samples are mostly far apart from each other with their replicates clustered somewhat together. **B** PCA of the doxorubicin treated samples after Combatseq. Clusters all of the samples from the first batch all together, thereby reducing variability. 77

Figure 4.24 **A** volcano plot for the Doxorubicin contrast. **B** volcano plot for the *siCHRNA5* contrast. **C** volcano plot for the combinatorial contrast. The upregulated genes for the combinatorial contrast appear to be similar to the ones in the doxorubicin contrast. 80

Figure 4.25 Scatter plot for the gene with a significant FDR for the *siCHRNA5* treatment. **A** *siCHRNA5* treated *EV* is on the y-axis and the combinatorial treatment is on the x-axis. **B** *siCHRNA5* treated *EV* is on the y-axis and the doxorubicin treatment is on the x-axis. R^2 value with the doxorubicin treatment is comparatively low. 81

Figure 4.26 **A** Disease-gene interactions from the DisGeNET database. “Invasive carcinoma of breast”, “Hepatoblastoma” and “T-cell lymphoma” show as the top three enriched terms. **B** GO:BP terms for the upregulated genes. **C** TFE shows *TP53* as one of the significantly enriched TFs. 82

Figure 4.27 MSigDB enriched terms for the downregulated *siCHRNA5* treated downregulated genes. 83

Figure 4.28 **A** GO enrichment based on the downregulated genes of the doxorubicin treated *EV* cell line. Most enriched terms are related to cell cycle pathways or replication. **B** Heatmap of the “Hallmark *E2F* targets”, “*G2M* targets”, and “Reactome cell cycle checkpoint” genes..85

Figure 4.29 Heatmap of *LIN37* dependent *p53* targets. 86

Figure 4.30 **A** Scatter plot of combinatorial against *siCHRNA5* for cell cycle genes. The points are close to the x-axis representing the lack of regulation from the *siCHRNA5* treatment. **B** Scatter plot of combinatorial against doxorubicin for cell cycle genes. The R^2 value is 0.743, which represents the co-dependence of the contrasts. 87

Figure 4.31 **A** Doxorubicin treated *R175H* contrast is plotted on the y-axis and *siCHRNA5* treatment on the x-axis. These genes are significantly expressed in the doxorubicin contrast and appear to be not mutually dependent. **B** These significantly expressed genes in the doxorubicin treatment are highly correlated with the combinatorial treatment. 88

Figure 4.32 **A** shows the GO functional enrichment of the upregulated *R175H* doxorubicin contrast. Most enriched terms include response to oxygen. **B** Dorothea TFE showed *TP53*, *ESR1* and *JUN* as the top most enriched TF. All of these TF play important roles in cancer progression. **C** Upregulated genes from the doxorubicin treatment also show apoptosis as an

enriched term. Important apoptosis targets upregulated by doxorubicin alone and in combination include, <i>P21</i> , <i>HMOX1</i> , <i>BAX</i> , <i>MDM2</i>	89
Figure 4.33 A Heatmap for the 230 downregulated by doxorubicin in the R175H mutant cell line. Regulation of these genes in the si <i>CHRNA5</i> contrast varies significantly compared to the doxorubicin and combinatorial. B functional enrichment based on the GO database. Various replication terms are enriched. C TFE shows the E2F family of transcription factors.	91
Figure 4.34 Scatter plot of the 931 DEGs for the si <i>CHRNA5</i> treatment in R175H mutant cell line. A si <i>CHRNA5</i> is plotted on the y-axis and doxorubicin on the x-axis. B si <i>CHRNA5</i> is plotted on the y-axis and the combinatorial on the x-axis.	92
Figure 4.35 A Heatmap of hallmark p53 and MYC targets for the upregulated genes from the si <i>CHRNA5</i> treated R175H mutant cell line. B GO enrichment for these genes. “Translational processing” and “ncRNA processing” appear to be important enriched terms in regulating gene expression.	93
Figure 4.36 A GO enrichment for the downregulated genes of the R175H si <i>CHRNA5</i> treated contrast. Transcription and apoptotic enriched terms show up. B TFE shows mostly zinc finger domain factors as enriched. C Heatmap for the hallmark estrogen response, KEGG cell cycle and DNA replication terms. Column wise clustering show that these targets appear to be more correlated between the combinatorial and the doxorubicin treatments.	95
Figure 4.37 Heatmap of the DREAM targets (filtered for p-value < 0.05 for the R175H contrasts).	96
Figure 4.38 LIN37 targets significant in all three contrasts.	97
Figure 4.39 Treatment-gene network for the LIN37 targets for the R175H contrasts. The dark grey nodes are genes which are significant in all three contrasts (previous heatmap). The white colored nodes are connected to the diamond node (combinatorial treatment) and hence could be potentially synergistic genes since they not significant in the other treatments.	98
Figure 4.40 A Volcano plot for the si <i>CHRNA5</i> treatment in the R273H cell line. <i>P21</i> can again be seen as upregulated by the silencing of <i>CHRNA5</i> . B Regulation of genes in the R273H cell line with the combinatorial treatment.	99
Figure 4.41 Scatter plot of the significant genes in the si <i>CHRNA5</i> contrast. The si <i>CHRNA5</i> contrast is on the x-axis and the combinatorial is on the y-axis.	100
Figure 4.42 A Gene Ontology enriched terms for the upregulated genes in the si <i>CHRNA5</i> treated R273H cell line. B Dorothea based TFE shows <i>ELK3</i> and <i>TP53</i> as the top enriched TF. C Heatmap showing the genes as targets of the hallmark and KEGG p53 pathway and MTORC1 signaling.	101
Figure 4.43 A Heatmap for the MSigDB enriched terms. Most of these targets appear to be more downregulated in the combinatorial treatment. B Heatmap for the LIN37 dependent p53 targets. These targets also appear to be more downregulated in the combinatorial treatment, demonstrating the effectiveness of the combinatorial treatment.	103
Figure 4.44 DEGs from doxorubicin treated WT contrast. A The doxorubicin is plotted on the x-axis and the combinatorial on the y-axis. These appear to highly correlated with one another. B The doxorubicin is plotted on the x-axis and the si <i>CHRNA5</i> on the y-axis.	104
Figure 4.45 A GO enrichment for the doxorubicin treated WT upregulated genes. B TFE analysis for the same gene set. Shows TF such <i>TP53</i> , <i>ESR1</i> and <i>ESR2</i> . C , Heatmap for the “apoptosis” and “P53 signaling pathway”.	106
Figure 4.46 A GO enrichment terms for these doxorubicin treated WT downregulated genes showed various cell regulatory pathways. B TFE shows E2F family of transcription factors along with MYC and FOXM1. C Heatmap of the downregulated genes of the E2F targets and cell regulatory terms.	108
Figure 4.47 Scatter plots for the si <i>CHRNA5</i> treated DEGs. A the si <i>CHRNA5</i> is plotted on the x-axis against the combinatorial on the y-axis. B For the same genes differently expressed by	

the si <i>CHRNA5</i> , the doxorubicin against the combinatorial was plotted. This showed a very correlation.....	109
Figure 4.48 A GO enriched terms for the upregulated genes of the si <i>CHRNA5</i> WT contrast. B TFE for the same genes shows <i>TP53</i> as the most significantly enriched TF. C Heatmap of the upregulated genes based on the MSigDB database terms relating to p53.	110
Figure 4.49 A Enriched terms for the downregulated genes of the si <i>CHRNA5</i> WT contrast. These include many cell regulatory and replication terms. B TFE shows the E2F family of transcription factors. C Heatmap for the significant TFs, <i>MYC</i> , <i>E2F4</i> , <i>E2F2</i> , <i>E2F1</i> and <i>FOXM1</i> . All of these TF have a high confidence level in the Dorothea database.	111
Figure 4.50 A GO enrichment for the upregulated R175H combinatorial genes. Terms enriched are very similar to the ones enriched by the doxorubicin treatment in R175H. B TFE based on the Dorothea database. C Heatmap of the upregulated genes for the Hallmark P53 and apoptosis pathways.....	113
Figure 4.51 A GO enriched terms for the downregulated R175H mutant cell line genes. In line with the individual treatments for the R175H cell line, various cell regulatory pathways appear to be enriched. B TFE shows the E2F family of TFs. C Heatmap for the downregulated genes for the Hallmark E2F and G2M checkpoint genes. The clustering shows an interesting pattern with R273H clustered together with the WT.....	115
Figure 4.52 A GO enriched terms for the upregulated genes in the R273H cell line with combinatorial treatment. B TFE based on the dorothea database. C Heatmap for the apoptosis and p53 enriched terms from the MSigDB database. The clustering shows the response of these upregulated genes in a similar manner for the mutant cell lines in comparison to the WT overexpressed cell line.	116
Figure 4.53 A GO enrichment terms for the downregulated genes of the R273H mutant. B TFE based on the dorothea database. C Heatmap of the E2F4, E2F1, E2F2 and FOXM1 TF based on the Dorothea database. The downregulated genes here for the R273H are also affected in a similar manner to that in the WT.	118
Figure 4.54 A GO enrichment terms for the upregulated genes in the WT combinatorial treatment. B TFE for these upregulated genes. C Heatmap of the p53 and apoptosis genes.....	120
Figure 4.55 A GO enrichment terms for the downregulated genes in the WT combinatorial treatment. B TFE for these upregulated genes. C Heatmap of the KEGG cell cycle and DNA replication terms from the MSigDB database.	121
Figure 4.56 Subset of the 126 genes, described as WT specific signature.	123
Figure 4.57 Heatmap for the 90 genes, described as the mutant cell line signature.	124
Figure 4.58 A Scatter plot of DREAM targets plotted for the Nutlin treatment against the si <i>CHRNA5</i> (on the y-axis). B Scatter plot of DREAM targets plotted for the combinatorial treatment against the si <i>CHRNA5</i> (on the y-axis). C LIN37 targets plotted as a heatmap for three different experiments.	125
Figure 4.59 AUC curve for the prediction of the model. Method 3 and method 4 gave an AUC of over 0.95, which could mean that this model is overfitting.	129
Figure 4.60 PCA of the genes based on their logFC, annotated by their cluster code.	131
Figure 4.61 UMAP clustering of binary codes to determine the annotations.....	131
Figure 5.1 Downregulation of the <i>CHRNA5</i> gene based on the doxorubicin treatments and the nutlin treatment from GSE47042.....	149

List of tables

Table 1-1. Summary and primary features of some of the present Shiny applications.	22
Table 1-2. Data types and Analysis methods/tools of the available shiny applications compared to the application presented in this thesis.....	23
Table 3-1. Summary of the datasets present in the TP53-META application.	33
Table 3-2. Sequence data analysis results for the first batch of the doxorubicin (Data 3) treated mutant samples.	34
Table 3-3. Sequence data analysis results for the second batch of the doxorubicin (Data 3) treated mutant samples.	35
Table 3-4. Sequence data analysis results for the mutant samples (Data 2).	35
Table 3-5. Tools and packages reference list for data analysis and the development of the shiny webserver.....	42
Table 4-1. Design matrix for the synergy calculation in Method 3.....	128
Table 4-2. Design matrix for the synergy calculation in Method 4.....	128
Table 4-3. Number of genes for each of the annotations for these methods.	130
Table 4-4. Clustering of the binary codes.....	132

List of Appendices

Appendix A: Experimental confirmation of the combinatorial treatments and its synergistic effect to reverse resistance

Section A1. Cell culture maintenance (carried out by Burcin Aras).....	168
Section A2. mRNA isolation (carried out by Burcin Aras).....	168
Section A3. Doxorubicin and si <i>CHRNA5</i> treatment (carried out by Cagdas Yanyatan)	168
Section A4. MTT cell viability assay (carried out by Cagdas Yanyatan).....	169

Table A1 Pre-library sample QC results for first batch of doxorubicin and si <i>CHRNA5</i> treated data.....	170
--	-----

Table A2 Pre-library sample QC results for first batch of doxorubicin and si <i>CHRNA5</i> treated data.....	170
--	-----

Figure A1 Cell viability using MTT assays for the si <i>CHRNA5</i> treatment alone. Other than the EV, the depletion of <i>CHRNA5</i> leads to a decrease in cell viability of the mutant cell lines and the WT over expressed cell line.....	170
---	-----

Figure A2 Demonstrates the effects of si <i>CHRNA5</i> in the presence of doxorubicin. Here, the doxorubicin was set as the control and the effect of the combinatoiral treatment was seen to decrease across all the cell lines. The p-value for the EV did not appear to be significant, while for other cell lines it was below 0.1. The effect of the combinatorial treatment was seen to decrease cell viability across all cell lines, thus reversing resistance observed against Doxorubicin.	170
---	-----

Figure A3 MTT assay raw readouts across the cell lines with doxorubicin treatment and the combinatorial treatment.	172
---	-----

Figure A4 Two way anova showing the MTT readouts for the doxorubicin data and its combination with si <i>CHRNA5</i> . Here the results show the same response as before with a decrease in drug resistance. The results for R175H show a higher readout which was also observed throughout the RNA-seq data, with R273H having a similar transcriptomic response with the WT upon the combinatorial treatment.	172
---	-----

Figure A5 Another MMT assay carried on the same cell lines with the same treatement groups. P-values unfortunately do not show significance, however a consistent pattern of reduced cell viability was observed with the combinatorial treatment.	173
---	-----

Appendix B: Exploration of .bam files in IGV

Figure B1 Coverage for all the samples for the <i>TP53</i> gene.	174
---	-----

1. Introduction

1.1 Overview of Breast Cancer and its Mechanisms of Development

Breast cancer is one of the most prevalent malignancies affecting women worldwide (WHO, 2023). It is a multifactorial and a complex heterogeneous disease. Features varying between breast cancer subtypes include their intrinsic nuclear receptor expression, the ability of the tumor cells to invade, the oncogenic driver mutations they possess and their response to anticancer therapy (Skibinski & Kuperwasser, 2015). Despite significant improvements in cancer treatment, heterogeneity and complexity of breast cancer and its subtypes makes it a difficult challenge to come up with a sustained treatment.

1.1.1 Genetic Aberrations Breast Cancers

The epidemiology of breast cancer presents a global health challenge, with incidence rates varying significantly across different regions due to genetic, environmental, and lifestyle factors. Several risk factors contribute to the onset and progression of breast cancer, with genetic predisposition being a key element. Apart from *BRCA1* and *BRCA2* genes, which are known for their high penetrance and autosomal dominant inheritance (Miki et al., 1994; Wooster et al., 1995), sporadic mutations in cell regulatory genes, such as *TP53*, *CDH1*, *PTEN* are also common (Łukasiewicz et al., 2021). Other than the high risk of breast cancer, individuals with mutations in these genes also face a higher risk of developing ovarian cancer (Łukasiewicz et al., 2021).

The *TP53* gene, often dubbed as the “Guardian of the Genome” is one of the most important tumor suppressor genes, and has crucial roles in maintenance of cellular checkpoints and in occurrence of cancer (Lane, 1992). It is an important transcription factor that deals with cellular response to stress, DNA damage, oncogene activation, and hypoxia. p53 can lead to the arrest of cell cycle for repair of DNA, can induce apoptosis and control senescence, making it crucial

in carcinogenesis (Vousden & Lane, 2007). *TP53* is also responsible for inducing *P21* (*CDKN1A*), which is a cyclin dependent kinase inhibitor, which serves a key role in stopping the cell cycle progression (El-Deiry et al., 1993). p21 slows down the cell cycle by preventing CDK dependent inactivation of Rb, a transcriptional repressor. This leads to the suppression of genes that are activated by the E2F family of transcription factors. (Chellappan et al., 1991; Sullivan et al., 2018). We recently showed that estrogen receptor signaling and *TP53* mutation status could associate with the tumors' expression levels of *CHRNA5*, a cholinergic nicotinic alpha receptor whose depletion induced a similar expression profiles as of known *TP53* inducers such as doxorubicin (Koker et al., 2018; Shehwana et al., 2021).

In the literature, mutations in *TP53* can lead to the development of more aggressive forms of breast cancer, which further complicates the treatment and prognosis. Research shows that about 30% of breast cancers have mutations in the *TP53* gene, and mutant P53 can either be beneficial or harmful to the clinical outcome, based on the treatments provided (Shahbandi et al., 2020).

1.1.2 Molecular Sub-types of Breast Cancer

Because of its heterogeneous nature, breast cancer is primarily categorized by either hormonal receptor status or gene expression profiles. The most common sub-type includes the hormone receptor-positive cancers, which include ER+ (Estrogen receptor positive) and PR+ (Progesterone receptor positive) tumors. These are categorized as Luminal A, which generally have a better prognosis and may be treated with hormone therapy (Ahn et al., 2015; Duffy et al., 2017).

The Luminal B subgroup, in addition to being hormone receptor (ER and PR) positive, may also exhibit HER2 overexpression and tend to be more aggressive. This subgroup also has

a higher proliferation related gene expression profile, which accounts for its poor prognosis (Z. Li et al., 2016).

HER2-positive breast cancer is marked by the high expression of the ERBB2/HER2 protein, while lacking ER and PR expression. HER2 positive breast cancers also have a poor prognosis and are likely to be more aggressive. However, the use of targeted therapies such as trastuzumab has greatly enhanced outcomes for patients with this subtype (Loibl & Gianni, 2017).

Triple-negative breast cancer (TNBC) or the basal-like subgroup lacks estrogen, progesterone, and HER2 receptors, making it more difficult to treat owing to a lack of targeted therapies. Due to its lack of suitability for these IHC markers, treatments often focus on inducing DNA damage, because high expression of cell cycle genes is observed and high mutations in *TP53* and *BRCA1* are common (Lehmann et al., 2011).

Another subgroup recently categorized as claudin-low breast cancers – a subgroup of TNBC due to its high heterogeneity – is distinguished by its unique molecular and histological characteristics (Dias et al., 2017). This subtype exhibits a low expression of claudin proteins, which are crucial for cell adhesion and tight junctions that shows a potential disruption in cellular architecture. Claudin-low tumors are further distinguished by their elevated expression of epithelial-to-mesenchymal transition (EMT) markers, which suggests increased invasiveness and metastatic potential (Dias et al., 2017; Pommier et al., 2020).

1.1.3 Drug resistance in Breast Cancer

Drug resistance can be a significant problem for cancer treatment and management. Breast cancer's complexity and molecular heterogeneity lead to the development of resistance mechanisms. The molecular basis of drug resistance is multifaceted and complex, and it consists of changes in drug targets, activation of alternative signaling pathways, efflux of drugs

from cancer cells, and alterations in the tumor microenvironment that can protect tumor cells from the effects of chemotherapeutic drugs (Tufail et al., 2022).

Genetic mutations within these cancer cells are a primary reason for the development of drug resistance. Among these, mutations in the *TP53* gene are specifically important. Mutations in *TP53* can lead to the disruption cellular pathways, further causing uncontrolled cell proliferation and increased tumor survival. This leads to the development of resistance to chemotherapy and targeted therapies (Sabapathy & Lane, 2018).

Specifically, R175H and R273H in the *TP53* gene have been linked with drug resistance in breast cancer. Such mutations result in the loss of normal p53 tumor suppressor activity while conferring oncogenic functions that promote tumor progression and therapeutic resistance (Olivier et al., 2010). The R273H mutation is known to change the DNA binding affinity and specificity of p53, which affects the transcriptional regulation of genes involved in apoptosis and cell cycle control (Freed-Pastor & Prives, 2012). Similarly, the R175H mutation disrupts the conformation of the p53 protein, which impairs the protein's ability to trigger apoptosis in response to chemotherapeutic agents (Muller & Vousden, 2013).

These mutations can also affect the therapeutic response to hormone therapy in hormone receptor-positive breast cancers. Such mutations can interfere with pathways that are targeted by therapies such as tamoxifen and aromatase inhibitors, contributing to the development of resistance in ER positive breast cancers (Sabapathy & Lane, 2018). Moreover, studies show that *TP53* mutations can enhance the adaptive capacity of cancer cells, which enables them to survive when there is treatment-induced stress, further complicating the management of drug-resistant breast cancer (Brosh & Rotter, 2009).

To address the problem of drug resistance in breast cancer, particularly in the context of *TP53* mutations, a number of different approaches are required. Previously, there have been multiple attempts to tackle the problem of drug resistance, with emergence of novel molecules

or combination therapies. For example, APR-246, also known as PRIMA-1^{MET}, has been extensively studied for its capability to target *TP53* mutations by changing mutant p53 into a conformation, which is more similar to wild-type p53 (Bykov et al., 2002). This molecule has shown effectiveness against the R175H mutation, which is one of the most prevalent *TP53* hotspots in various cancers (Chiang et al., 2021). APR-246 has advanced to clinical trials and demonstrated potential to improve chemotherapy effectiveness and enhance patient outcomes when combined with platinum-based therapy (Mohell et al., 2015). Recent studies have also suggested, that combination of two different MDM2 inhibitors, nutlin and navtermadelin, may reduce cell growth by 75 percent in hypoxia (Clavero et al., 2023).

To effectively tackle the problem of drug resistance in breast cancer, especially in the context of *TP53* mutations, the development of novel therapeutic agents and combination strategies is essential. Despite promising developments in combinatorial therapies, there remains a critical need for further research and novel strategies, to fully determine the potential of personalized treatments, which can adapt to the complex heterogeneous nature of individual tumors and prevent the emergence of resistance.

1.1.4 MCF7 cell line in Breast Cancer

The MCF7 cell line is a very important model in better understanding the pathology and pharmacology of breast cancer. It was derived in the 1970s, from the pleural effusion of a human breast adenocarcinoma patient and has become one of the most commonly utilized and studied cell lines in breast cancer research (Comşa et al., 2015). One of the reasons for the prevalent use of the MCF7 cell line is because of its hormone receptor-positive status (ER and PR positive). This makes the cell line an important model for exploring and studying the estrogen receptor (ER) signaling pathways and the effects of hormonal therapies in breast cancer (Lacroix & Leclercq, 2004; Neve et al., 2006).

In addition to the cell line's ER-positive status, it exhibits a wild-type *TP53* gene, which provides researchers with the opportunity to investigate the interaction between ER signaling and the p53 pathway. The functional p53 in MCF7 cells can also help in the study of how genotoxic stress affects breast cancer's survival and proliferation, which allows researchers to analyze the mechanism of action for different chemotherapeutic agents (Soussi & Wiman, 2007). The cell line's responsiveness to DNA-damaging agents has made it a valuable tool in screening for potential anticancer drugs and in understanding the cellular mechanisms underlying drug sensitivity and resistance (Ourique et al., 2015).

1.2 The role of *TP53* in Breast Cancer

1.2.1 Function of p53 in cancer progression

p53 regulates the transcription of various genes involved in DNA repair, metabolism, and senescence, thus reinstating its role as a master regulator in cellular defense mechanisms against cancer (Riley et al., 2008). By maintaining genomic integrity, p53 protects against oncogenic mutations that could result in cancerous growth. In case of irreparable damage to DNA, p53 triggers apoptosis by transcriptionally activating pro-apoptotic genes such as *PUMA*, *BAX*, or *NOXA* (Nakano & Vousden, 2001). Additionally, p53 affects cell metabolism by driving a shift to oxidative phosphorylation from glycolysis, a metabolic pathway preferred by cancer cells. This results in potential tumor cells being deprived of energy while it would also limit the production of metabolic intermediates, which are needed for rapid cell growth (Matoba et al., 2006).

p53 is structurally made up of several functional domains. It consists of an N-terminal transactivation domain that initiates transcription, a central core DNA-binding domain that is essential for its regulatory functions, and a C-terminal oligomerization domain that facilitates the formation of the active tetramer (Cho et al., 1994; Jeffrey et al., 1995). The DNA binding

domain of the p53 protein consists of two antiparallel β -sheets, which act as a base for a flexible DNA-binding surface. This surface has a loop-sheet-helix pattern and two sizable loops that are stabilized by a zinc atom. The attachment of the zinc ion plays a crucial role in ensuring the proper folding of the DNA binding domain (Olivier et al., 2010). Hotspot mutations in p53 exist at residues that either make direct contact with DNA or support the structure of the DNA-binding surface. These mutations are defined as "contact mutations" or "structural mutations," respectively (Olivier et al., 2010).

1.2.2 DREAM Complex and *TP53*

The role of DREAM (Dimerization partner (DP), retinoblastoma (RB)-like, E2F and MuvB) complex and *TP53* is crucial for the control of cell cycle, particularly their roles in cell cycle arrest as well as the preservation of genomic stability under stress conditions (Sadasivam & DeCaprio, 2013). It is a key regulator of cell cycle-dependent gene expression, controlling cell cycle checkpoints. It represses gene activity during the quiescent G0 phase and coordinates the activation of genes necessary for the G1/S and G2/M transitions (Litovchick et al., 2007).

The dynamics of the DREAM complex vary throughout the cell cycle and that determines its role in either repression or activation of genes. Gene expression is inhibited at the beginning of the cell cycle, transitioning to activation in the later stages. The DREAM complex binds to E2F, CHR, CDE (cell cycle-dependent element), and CLE (CHR-like element) sites to suppress transcription during the G0 and early G1 phases (Engeland, 2018).

This transcriptional inhibition is relieved during the G2 phase and mitosis, which results in activation via CHR sites. The MuvB core complex, consisting of LIN9, LIN37, LIN52, LIN54, and RBBP4, is critical for this process. LIN54 specifically binds to CHR elements, facilitating the transition from repression to activation. The dissociation of E2F4-5/DP and p107/p130 from the complex, allows for the recruitment of B-MYB and FOXM1 into the

MuvB core, which shifts function of the complex from repressive to activating. The MMB (B-MYB-MuvB), FOXM1-MMB, and FOXM1-MuvB complexes then function as activators for late cell cycle genes containing functional CHR elements (Engeland, 2018; Sadasivam & DeCaprio, 2013). Dysregulation of the DREAM pathway is crucial for onset of cancer as this unbalances the finely tuned repression and activation of cell cycle genes (Sadasivam & DeCaprio, 2013).

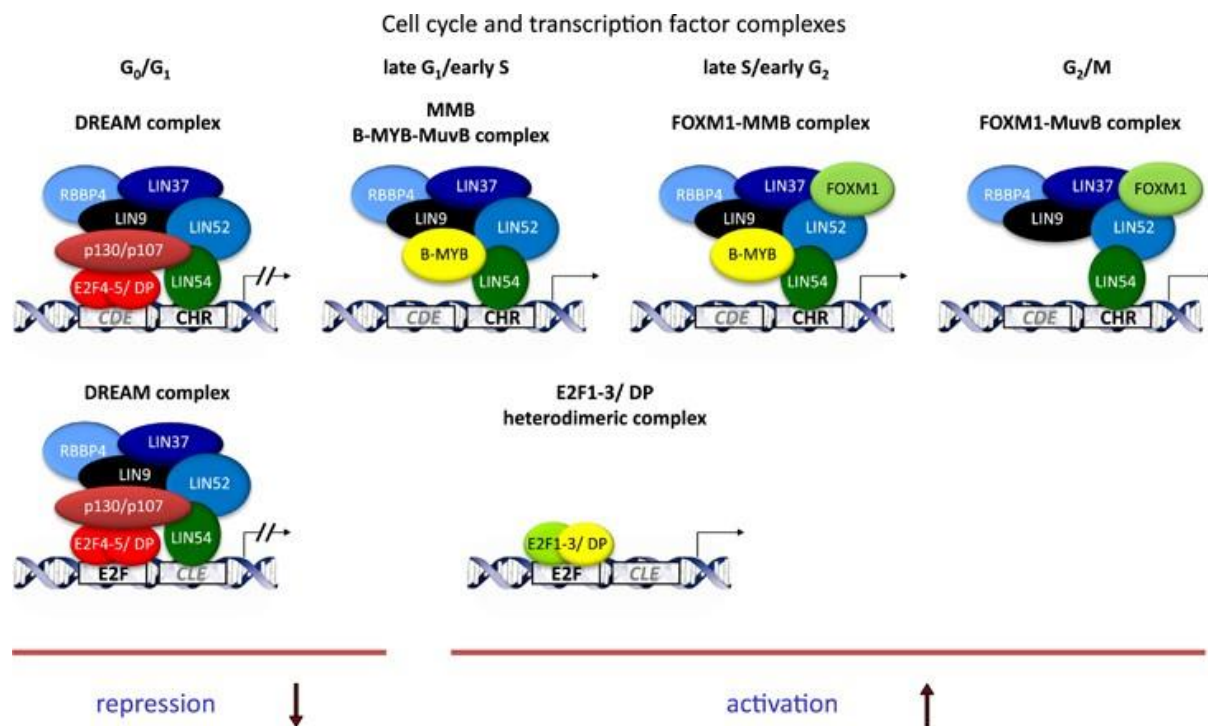


Figure 1.1 The change in transcription activation/repression due to the binding of the complexes to DNA during different phases of the cell cycle.

(This figure is reproduced under the Creative Commons license from Engeland, K. *Cell cycle arrest through indirect transcriptional repression by p53: I have a DREAM*. Cell Death Differ 25, 114–132 (2018). <https://doi.org/10.1038/cdd.2017.172>)

1.2.3 LIN37 and DREAM complex:

The LIN37 component of the DREAM complex plays a crucial role in cell cycle regulation. LIN37 as part of the DREAM complex, contributes to the repression of cell cycle-dependent genes during the G₀ and early G₁ phases, ensuring that cells only proceed to DNA replication

and division at the appropriate time (Fischer & Müller, 2017). Recent research has shown the probable role of LIN37 in cancer progression (Mages et al., 2017). This may suggest that LIN37 could function as a biomarker for cancer progression and a potential target for therapeutic intervention. Recent research by Uxa et al., confirmed the importance of LIN37 as an essential component of the DREAM complex for the p53-dependent downregulation of cell-cycle genes. Utilizing both HCT116 human colon carcinoma cell lines and mouse C2C12 cells, the researchers highlighted that the absence of LIN37 led to a significant impairment in the repression of cell-cycle genes, particularly those associated with G2/M and G1/S phases (Uxa et al., 2019). This impairment was demonstrated through experiments where LIN37^{-/-} cells showed reduced repression of critical cell-cycle regulators upon treatment with Nutlin-3a, an MDM2 inhibitor, compared to LIN37^{+/+} cells (Uxa et al., 2019).

Furthermore, the same study identified genes that were repressed by the actions of P53-DREAM but not downregulated by p53 independent of LIN37/DREAM, thereby establishing specificity in the activities of LIN37 within this regulatory network. The loss of LIN37, along with the absence of RB, completely abrogated the p53-dependent repression of cell-cycle genes, leading to the failure of the G1/S checkpoint (Uxa et al., 2019).

1.3 TP53 Mutations: R175H and R273H

Mutations in *TP53* are among the most frequent genetic changes observed in human cancers. The biological impact or effect of mutations in the p53 protein vary widely, from a simple loss-of-function to the acquisition of new oncogenic abilities. Various *in vitro* studies have shown that particular p53 mutants can gain new functions and engage in progression and development of tumors and also affect the resistance to conventional chemotherapy drugs (Tan et al., 2015). Among these mutations, the hotspot loci, R175, R248 and R273 are the most prevalent, especially in breast cancer (Y. Zhang et al., 2018). Using Breast Invasive Ductal carcinoma

TCGA datasets on cBioportal, R273H/C/L and R175H, show up as the most mutated hotspots (Cerami et al., 2012; <https://www.cbioportal.org/>).

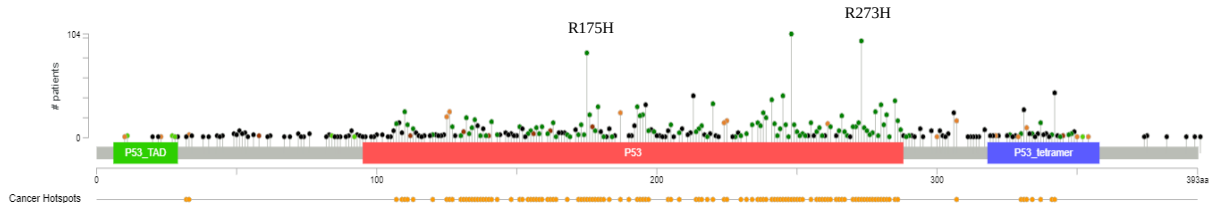


Figure 1.2 Prevalence and distribution of *TP53* mutations in Breast Cancer cohorts, taken from cBioportal. The datasets included were TCGA breast cancer datasets, METABRIC datasets, and MSK datasets (n = 9898). The three residues most commonly affected by mutations are R248, R175, and R273. The types of mutations and their associated color codes include: missense mutations (green), truncating mutations (black), in-frame mutations (red), and splice variants (orange).

1.3.1 Molecular Impact of R175H Mutation

The R175H mutation in *TP53* is a hotspot mutation that has been extensively studied for its impact on the function of the p53 protein. This mutation causes the substitution of arginine with histidine at codon 175, which is present at the zinc-binding site, close to the core DNA-binding domain (DBD) of p53, which is crucial for maintaining the protein's structural stability (Chiang et al., 2021). This mutation disrupts the original conformation of the DBD, and hence impairs the ability of p53 protein to bind specifically to its target DNA sequences and hence is also classified as a structural mutation unlike the R248 or R273 which are classified as contact mutations (Chiang et al., 2021).

This loss of DNA-binding affinity compromises the protein's ability to function as a transcription factor for several genes that control its tumor suppressor functions such as cell cycle arrest through p21, DNA repair, and apoptosis. As a result, its function in preserving genomic stability and inhibiting the growth of tumors is rendered ineffective (Chiang et al., 2021). Studies have shown that the R175H mutation not only diminishes the tumor suppressive activities of p53 but may also give new oncogenic properties to the mutant protein (Petitjean et al., 2007). p53-R175H may also bind to particular DNA sequences that vary from the wild-

type p53 response elements, hence transactivating those particular downstream pathways or target genes (Quante et al., 2012). These "gain of function" effects include the ability to promote cell proliferation, metastasis, and invasion, along with inducing genomic instability and altering cellular metabolism, which collectively aid the progression of cancer and resistance to therapy (Chiang et al., 2021).

1.3.2 Molecular Impact of R273H Mutation

Unlike the R175H mutation in p53, the R273H is a contact mutation. It results from the substitution of arginine with histidine at position 273, which is vital for the protein's ability to bind DNA and activate the transcription of target genes involved in cell cycle arrest, DNA repair, and apoptosis (Freed-Pastor & Prives, 2012). This mutation directly impairs the protein's capacity to function as a tumor suppressor by disrupting its interaction with DNA.

Studies have shown that the R273H mutation can alter various cellular pathways, including those involved in metabolism, apoptosis inhibition, and genomic instability. Mutant p53-R273H, for instance, plays a role in the metabolic reprogramming of cancer cells, enhancing glutaminolysis and glycolysis to meet the heightened needs of rapidly expanding tumor cells (C. Zhang et al., 2013). This metabolic shift is crucial for tumor growth and survival, particularly in hypoxic environments common within solid tumors.

Moreover, the mutant p53-R273H protein may be able to engage in novel protein-protein interactions that wild-type p53 cannot, affecting multiple cellular processes and signaling pathways. It might modify the activity of other transcription factors and co-regulators through interactions in order to promote oncogenesis. For example, mutant p53 may affect NF- κ B signaling, promoting inflammation and further contributing to tumor progression (Weisz et al., 2007).

1.3.3 Clinical significance of *TP53* mutations

Studies have shown that both R175H and R273H mutations of the *TP53* gene may not only diminish the tumor suppressive ability of p53 but also may lead to new oncogenic properties (Chiang et al., 2021; Freed-Pastor & Prives, 2012). R175H and R273H have both been identified across multiple cancers, including breast cancer. Studies have shown that breast cancers harboring the R175H mutation tend to show more aggressive characteristics, including higher grade, increased proliferation rates, and poorer differentiation (Chiang et al., 2021). Similarly, the R273H mutation not only disrupts p53's ability to induce apoptosis in response to DNA damage, but also its GOF phenotype may lead it to activate genes that facilitate invasion, metastasis and chemoresistance (Freed-Pastor & Prives, 2012).

The clinical significance of these mutations goes further than their impact on the biological behavior of cancer, especially in the context of breast cancer. This is because these mutations also have implications for the management of the disease. For example, *TP53* mutations are associated with resistance to certain chemotherapeutic agents, particularly those that induce DNA damage as a mechanism of action (Hientz et al., 2017). This resistance can limit the effectiveness of standard chemotherapy drugs, making it necessary to develop alternative therapeutic strategies. Recent research has focused on the potential of targeted therapies that can restore p53 function or exploit the vulnerabilities created by p53 mutations, offering hope for improved treatment outcomes for patients with *TP53* mutated breast cancer (Bykov et al., 2018; Hientz et al., 2017).

1.4 *CHRNA5* and its Role in Cancer

CHRNA5 is an ion-gated, nicotinic acetylcholine receptor, and is implicated in lung cancer (Spitz et al., 2008; Truong et al., 2010). However, recently few studies have showed the importance of this receptor in breast cancer. Lee et al. highlighted the importance of other

cholinergic receptor subunits in estrogen signaling and breast cancer (Lee et al., 2010). Others showed that this receptor could act as an oncogene, potentially by its involvement in DNA damage response (DDR) and dysregulation of certain cell signalling pathways, such as the PI3k/AKT, JAK/STAT and the ERK/MAPK pathways (Schaal & Chellappan, 2014). The variable expression of *CHRNA5* in breast cancer cell lines was shown by Koker et al., particularly in relation to genomic changes and the presence or absence of *TP53* mutations (Koker et al., 2018). Functional analyses of *CHRNA5* RNA interference (RNAi) model in MCF7 cells revealed that *CHRNA5* co-expression in CCLE breast cancer cell lines correlated positively with cell cycle regulation and DNA repair pathways. Moreover, a notable correlation was identified between *CHRNA5* RNAi expression profiles in MCF7 and A549 lung cell lines, suggesting a conserved role of *CHRNA5* in these cancers (Koker et al., 2018). Silencing of *CHRNA5* was shown to significantly reduce cell viability, DNA synthesis, and levels of phosphorylated RB, alongside decreased *CCNE2* mRNA levels—indicative of cell cycle arrest. This was further supported by an increased sub-G1 population and significantly higher BAX/BCL2 ratios, highlighting the induction of apoptosis (Koker et al., 2018).

Another possible regulation of *CHRNA5* and cell cycle, could be through the estrogen receptor (ER) pathway as shown by Shewana et al. This research showed that breast tumors lacking estrogen receptors (ER⁻) demonstrated greater *CHRNA5* levels compared to the ER⁺ tumors. Additionally, *CHRNA5* and its associated gene network were identified as secondary targets activated by estradiol, E2, predominantly suppressed by *CHRNA5* siRNA treatment in MCF7 cells (Shehwana et al., 2021).

CHRNA5 is an emerging factor in drug responsiveness. Drug resistance in cancer therapy is critical in determining patient prognosis and treatment outcomes. Consequently, using siRNA molecules to enhance the sensitivity of cancer cells to established anti-cancer treatments represents a promising strategy in cancer drug development (Grohar et al., 2014).

CHRNA5's influence on nicotine-induced changes in intracellular calcium levels may affect drug resistance (Ray et al., 2017). Cholinergic receptors especially downregulation of *CHRNA5* provide a good target for synergistically tackling the problem of drug resistivity due to the ability of targeting these receptors with antagonists or specific ligands (Ray et al., 2017).

1.5 Differential Effects of Doxorubicin, Nutlin, RITA and PRIMA-1

1.5.1 Mechanisms of Action

Restoration or inducing *TP53* activity in mutant cells via various different mechanisms has been an important research topic in cancer therapy. Some of these strategies that restore the function of p53 include the use of drugs such as Doxorubicin, Nutlin, RITA or PRIMA-1 (S. Zhang et al., 2022).

Doxorubicin, which is widely used as a chemotherapeutic agent, functions primarily by intercalation of the DNA strands that disrupts the replication and transcription processes. By embedding itself between base pairs in the DNA helix, doxorubicin inhibits the progress of the enzyme topoisomerase II, which is necessary for DNA replication and transcription. This inhibition leads to the accumulation of DNA double-strand breaks, triggering cell death pathways in rapidly dividing cells (Minotti et al., 2004). Research has shown that the interaction between doxorubicin and *TP53*, particularly in mutant forms, significantly influences treatment outcomes. Doxorubicin's effectiveness can vary greatly depending on the *TP53* status of cancer cells. It efficiently promotes apoptosis in cells expressing wild-type *TP53*, primarily via p53-mediated pathways. However, in cells with mutant *TP53*, the response to doxorubicin is often diminished, which leads to increased drug resistance and proliferation (Sun et al., 2016). Recently the challenges of using doxorubicin in treating cancers with *TP53* mutations was demonstrated by Zhang et al. The study showed that mutations in *TP53*, such as R273C and R282Q, significantly reduce the sensitivity of DLBCL (Diffuse large B-cell

lymphoma) cells to doxorubicin, with an increase in IC₅₀ values, indicating higher resistance (Zhang et al., 2022). This emphasizes the significance of *TP53* status in determining the efficacy of doxorubicin-based treatments.

Nutlin represents a class of small molecules designed to inhibit the interaction between the tumor suppressor protein p53 and its negative regulator MDM2 (Vassilev et al., 2004). Under normal conditions, MDM2 binds to p53, targeting it for ubiquitination and proteasomal degradation, thus regulating its activity tightly. Nutlin binds to MDM2 at the p53 binding site, preventing the p53-MDM2 interaction. This disruption causes stabilization and accumulation of p53 in the nucleus, where it can activate transcription of genes involved in cell cycle arrest and apoptosis (Vassilev et al., 2004).

RITA (Reactivation of p53 and induction of tumor cell apoptosis) is a small molecule that, unlike Nutlin, binds directly to p53 (Issaeva et al., 2004). By binding to p53, RITA blocks the interaction between p53 and MDM2, which causes the stabilization and activation of p53's transcriptional activity. This activation causes the activation of p53 targets that lead to a halt in cell cycle progression and apoptosis (Issaeva et al., 2004).

PRIMA-1 (p53 reactivation and induction of massive apoptosis) is a small molecule compound that targets mutant forms of the p53 protein (Bykov et al., 2002). PRIMA-1 is part of a class of drugs known as mutant p53 reactivators, which aim to restore the wild-type, tumor suppressive functions of mutated p53 proteins (Lambert et al., 2010). The mode of action of PRIMA-1 involves the conversion to its active metabolite, PRIMA-1^{MET} (APR-246), inside cancer cells. This metabolite directly attaches to the core domain of mutant p53, refolding it to restore the wild-type conformation and function. This reactivation allows p53 to regain its tumor suppressor functions, which were otherwise impaired (Bykov et al., 2018).

1.5.2 Comparative Analysis and Therapeutic Efficacy

Doxorubicin acts as a broad-spectrum anticancer agent, showing activity across a wide spectrum of cancers, which has made it an extremely versatile agent in frontline chemotherapy. However, like many cytotoxic drugs, dose-dependent cardiotoxicity is often a limiting feature of its broad clinical utility (Minotti et al., 2004; Vejpongsa & Yeh, 2013).

Another problem faced in using Doxorubicin, is drug resistance observed in cancers. Doxorubicin often encounters resistance mechanisms in cancer cells, such as enhanced drug efflux by ATP-binding cassette (ABC) transporters, or through reactive oxygen species (ROS) which can activate MAPK pathways (Christowitz et al., 2019; Sun et al., 2012). These resistance mechanisms reduce Doxorubicin's effectiveness over time, and require combination therapies or alternative treatments to overcome resistance. Despite this, Doxorubicin continues to be a crucial drug in the treatment of many cancers, particularly those that are aggressive and highly proliferative, and in combination strategies designed to limit resistance and enhance therapeutic efficacy with other agents.

Nutlin, unlike Doxorubicin, shows the importance of targeted therapies by activating p53 selectively in tumors that retain wild-type *TP53* genes (Vassilev et al., 2004). Its design to target the MDM2-p53 interaction, provides a therapeutic advantage that spares non-cancerous cells, which reduces systemic toxicity. This selectivity is especially advantageous in treating hematologic malignancies and solid tumors that harbor functional p53, providing a particularly promising approach to patients with these tumors (Tovar et al., 2006; Vassilev et al., 2004). However, Nutlin also faces challenges in drug resistance, particularly in cancers with mutant p53 or abnormalities in the downstream p53 pathway. In such cases, the restoration of p53 activity by Nutlin is less effective, leading to reduced therapeutic outcomes (Shen et al., 2008). However, research into combination therapies that can sensitize cancer cells to Nutlin's effects

or target parallel pathways shows promise in overcoming these resistance mechanisms (Bykov et al., 2018).

The unique mechanism of binding and direct activation of p53 by RITA represents an innovative strategy in cancer therapy. Even though, preliminary studies indicate RITA's potential to induce apoptosis in cancer cells with minimal off-target effects, suggesting a favorable therapeutic index much more work remains in order to establish its full clinical efficacy and safety profile (Issaeva et al., 2004). RITA's ability to induce cell death in cancer cells, even in the presence of mutant p53, suggests a potential pathway to circumvent traditional resistance mechanisms. However, like Nutlin and Doxorubicin, RITA's efficacy can be limited by alterations in cellular pathways that affect apoptosis and cell cycle arrest, highlighting the need for ongoing research into overcoming drug resistance (M. N. Saha et al., 2013).

The ability of PRIMA-1^{MET} to bind directly to p53 and make it regain its original WT conformation holds extreme potential compared to other therapies, which indirectly target p53. Clinical trials have shown that PRIMA-1^{MET} can be effective in various cancer types characterized by *TP53* mutations. Its ability to synergize with other chemotherapeutic agents and target multiple pathways which are crucial for cancer cell survival, makes it a promising candidate for combination therapies. Moreover, PRIMA-1^{MET} has demonstrated a relatively favorable safety profile, which is crucial for the development of effective cancer treatments (Mohell et al., 2015).

1.6 Combinatorial Effects of Drugs and Synergism

Because conventional chemotherapy treatments have developed resistance mechanisms, there is a need for the development combinatorial therapies to enhance therapeutic efficacy. Various studies have shown that combinatorial therapies may have a synergistic effect or at least an additive effect on the outcome compared to monotherapies. Hence, there is a need to find how

the transcriptome responds to different drug combinations. Combining multiple drugs together may also lead to less dosage for a particular drug, which reduces the chances of chemotoxicity (Diaz et al., 2020).

Combinatorial therapies involve the simultaneous or sequential use of more than one therapeutic agent, which can target different pathways involved in cancer progression. The reason for development of combinatorial therapies, is due to the heterogeneous nature of cancer and its ability to adapt and become resistance to single-agent therapies (Kwong et al., 2013). The primary mechanisms driving the efficacy of combinatorial therapies can generally be classified as synergistic or additive. Synergistic effects can be defined when the combinatorial action of drugs leads to a therapeutic outcome, higher than the actual sum of their individual effect, which is often through modulation of complementary downstream pathways (Chou, 2010). For example, for melanoma treatment, BRAF inhibitor, when used together with a MEK inhibitor, increases response rates and survival by simultaneously shutting down the MAPK pathway, which helps in overcoming resistance (Long et al., 2014). In contrast to this, additive effects occur when the sum of the actions of drugs is similar as their direct combined effect, which provides a rationale for combining drugs that individually target different mechanisms and pathways.

There is emerging research which explores the use of high-throughput methods and computational methods to find effective drug combinations and predict their outcomes. For example, Diaz et al. employed a matched time course RNAseq experiment of cells treated with single drugs or their combination and used a simple prediction method to predict synergy between drug treatments. Their gene regulatory network showed sequential activation of transcription factors over time, suggesting that drug synergy may be driven by the correlation of transcriptional responses based on unrelated treatments (Diaz et al., 2020). Similarly, Li et al. also developed a computational prediction algorithm using the largest drug combination

screening dataset and they also suggested that synergism between drugs continues to exist across different cell lines (H. Li et al., 2018). These computational methods help scientists decipher mechanisms across big datasets where experimental studies would be very difficult to carry out.

Furthermore, combining genomic and molecular profiling of tumors with different drug treatments and their combinations, could pave the way for personalized combinatorial therapies, where the treatment process can be tailored to the specific genetic makeup and resistance mechanisms of the particular tumor type.

1.7 Use of Web Applications in Transcriptomics

1.7.1 Web applications for *TP53* datasets and/or RNA-seq data analysis

With the prevalence of Next Generation Sequencing (NGS), large amounts of data need to be processed and analyzed. Analyzing these huge datasets may require some coding experience with languages such as R or python. The R shiny framework provides a solution to researchers who want to analyze data themselves, although requires knowing basics of coding languages. Users can generate dynamic web applications directly from R, using the R Shiny package. (Jia et al., 2022). These shiny based applications allow for input of raw count files, as well as differential expression analysis and visualization with modern responsive integrative interfaces with many examples in the literature (Table 1-1). Users can analyze and visualize their RNA-seq data from start to finish from differential expression (DE) analysis to pathway analysis to Gene Ontology categorizations.

The current web tools for RNA-seq data analysis provide a convenient, yet diverse range of applications to the users. Each tool has some unique features, and all capture some aspect of what is required to transform large amounts of data into interpretable results. Table 1-1 shows some features summarized between RNAseq DE related shiny based applications.

From many RNA-seq analysis tools present online, I chose the ones which could help the users most to analyze their *TP53* specific datasets.

CAP-RNAseq (Vural-Ozdeniz et al., 2024) is a novel, comprehensive webserver which allows for the analysis and exploration of bulk RNA-seq data through clustering and prioritization techniques. This webserver integrates the ability to upload RNA-seq data, perform cluster analysis, and annotate and visualize gene expression patterns within these clusters. Its distinctive feature lies in its capability to prioritize genes or clusters based on gene essentiality scores from DepMap (Tsherniak et al., 2017) and the congruency between mRNA and protein expression levels. While CAP-RNAseq is not specifically *TP53* specific, researchers can use *TP53* modulated datasets to analyze patterns across clusters while at the same time deciphering the role of *TP53* in particular pathways and their overlap with the “essential genes” described in the application.

GENAVi (Reyes et al., 2019) is another shiny application, which is designed to facilitate gene expression normalization, analysis, and visualization of RNA-seq data. The application supports normalization and differential expression analysis (DEA) to compare gene expression profiles across samples. Clustering can be done using distance metrics between genes or correlation of the genes. The application also allows the user to conduct gene set enrichment analyses (GSEA) on the uploaded dataset. Although this application is not *TP53* specific and lacks some basic visualizations and analyses for datasets, it is one of the few applications mentioned in this thesis, which allows the user to download reports of the analyses performed in the application.

DRPPM-EASY (Obermayer et al., 2022) is another Shiny web application tailored for RNA-seq analysis, using R, as the programming language. It provides users with the capability to perform clustering, visual comparison of gene expression, and uses Limma (Ritchie et al., 2015) for differential expression analysis and GSVA (Hänzelmann et al., 2013) for gene set

enrichment. It provides visualizations of the data as volcano plots, MA plots, and heatmaps. However, some limitations of the application include it not being specific to *TP53*, not having an online server, which makes it difficult for user to run the application on R locally, and not being able to analyze uploaded files.

On the other hand, the *TP53* specific web tools were either based on Shiny (Table 1-2) or on other programming languages and platforms, which were not included here.

In comparison to the other tools present, Targetgenereg.org (Fischer et al., 2022) appears to be the most relevant due to its P53 specificity. It is a shiny based database of multiple *TP53* RNA-seq and chip-seq datasets. Fischer et al. performed meta-analysis on multiple datasets to determine the regulation of genes in a p53 dependent manner. The user can look for their gene and look at statistics based on multiple datasets and how that gene is regulated among different cell lines based on different treatments. The web-tool allows the user to visualize the data as tables, pie charts, volcano plots and boxplots. However, the user cannot perform differential gene expression analysis, enrichment/over representation analysis on the provided datasets or explore the data freely.

The P53 cistrome shiny application (Nguyen et al., 2018) is a web-based tool designed for the interactive analysis and comparison of p53 ChIP-seq and cistrome data. This application was created as part of a broader effort to unveil the complex universe of human p53 interactions and functions, particularly in the context of tumor suppression and cellular stress responses. The primary function of the P53 cistrome shiny application is to provide researchers with a user-friendly interface to explore and analyze the binding patterns of p53 across different genomic contexts and investigate the potential regulatory impact of p53 on gene expression in various cell types or conditions.

Table 1-1. Summary and primary features of some of the present Shiny applications.

Application		TP53-META (This thesis)	Targetgenereg.org (Fischer et al., 2022)	CapRNA-seq (Vural-Ozdeniz et al., 2024)	P53-Cistrome (Nguyen et al., 2018)	Genavi (Reyes et al., 2019)	DRPPM-easy (Obermayer et al., 2022)
Feature	Online Server	Yes	Yes	Yes	Yes	Yes	No
	Repository (Github etc.)	Yes	No	Yes	No	Yes	Yes
Primary Features	P53 Specific	Yes	Yes	No	Yes	No	No
	User Upload	No	No	Yes	No	Yes	No
	Demo data / Help	NA / Yes	NA / Yes	Yes / Yes	NA / No		Yes / No
	Reactivity/Interactive	Yes	No	Yes	No	Yes	Yes

Table 1-2. Data types and Analysis methods/tools of the available shiny applications compared to the application presented in this thesis.

Application		TP53-META (This thesis)	Targetgenereg.org (Fischer et al., 2022)	CapRNA-seq (Vural-Ozdeniz et al., 2024)	P53-Cistrome (Nguyen et al., 2018)	Genavi (Reyes et al., 2019)	DRPPM-easy (Obermayer et al., 2022)
Feature	Data Type	NGS data type	RNA-seq	RNA-seq Chip-seq	RNA-seq	Chip-seq	RNA-seq
	Upload type	NA	NA	RNA-seq counts Condition file	NA	RNA-seq counts	NA
Data Analysis	Meta-analysis	Yes	Yes	No	Yes	No	No
	DGE methods	Deseq2 Limma	NA	Yes	NA	Deseq2	Limma
	PCA	Yes	No	Yes	No	Yes	No
	Volcano plot	Yes	Yes	Yes	No	Yes	Yes
	Heatmap	Yes	No	Yes	Yes	Yes	Yes
	Enrichment Analysis	Yes	No	Yes	No	Yes	Yes
	TF enrichment	Yes	No	No	No	No	No
	Networks	Yes	No	Yes	No	No	No

1.7.2 Meta-analysis tools for transcriptomic or biological data

Meta-analysis has emerged as a powerful tool in the analysis of RNA-seq and other biological data and allows researchers to synthesize findings across multiple studies for comparative analysis. The landscape of meta-analysis tools for gene expression data is vast, comprising both web-based applications and robust R packages designed to handle complex datasets.

NetworkAnalyst is one of these web-based tools, offering extensive functionalities including quality control, normalization, and meta-analysis methods, along with various data visualization capabilities utilizing KEGG and Gene Ontology databases (Xia et al., 2014). However, it requires users to manually format and upload their datasets, which could be challenging for non-specialists.

Another significant tool, programmed in R shiny, is ImaGEO. Using data directly from the GEO database, this app specializes in the meta-analysis of public gene expression data. It facilitates data handling by allowing users to input GEO identifiers and includes features for outlier detection and data filtering (Toro-Domínguez et al., 2021). Gemma and ExAtlas are additional web-based tools that enhance the accessibility of meta-analysis. Gemma offers a curated database of over 3000 datasets, whereas ExAtlas includes both public and user-defined data with comprehensive analysis methods (Toro-Domínguez et al., 2021).

In the case of R packages, MetaOmics and MetaIntegrator provide interactive platforms with a focus on effect size, P-value, and rank combination methods. Specifically, MetaOmics supports a variety of meta-analysis techniques and includes additional modules for deeper data exploration and analysis (Toro-Domínguez et al., 2021).

1.8 Statistical methods in RNA-seq

1.8.1 DEseq2 vs Limma

DESeq2 (Love et al., 2014) and Limma (Ritchie et al., 2015) represent two of the most commonly utilized tools for analyzing gene expression data, including the identification of differentially expressed genes derived from high-throughput sequencing experiments like RNA sequencing (RNA-seq). Both tools are designed to provide comprehensive statistical frameworks to test for differential expression in data involving complex experimental designs. Their approaches and underlying statistical models are, however, very different.

DESeq2 operates on raw count data derived from RNA-seq experiments. It uses the negative binomial distribution model for the count data and includes methods for variance estimation and data normalization. DESeq2's strength lies in its ability to accurately model the variability observed in count data, providing reliable differential expression results even in experiments with small sample sizes. It comprises a shrinkage estimator for dispersion, enabling more stable and interpretable estimates, especially when dealing with lowly expressed genes (Love et al., 2014).

Limma, first developed for data analysis from the microarray, has later been extended to handle RNA-sequencing data through the Voom (Law et al., 2014) method, enabling transformation into log₂-counts per million (CPM) for linear modeling. Developed by Ritchie et al. (2015), the limma workflow combined with Voom gives the chance to exploit Limma's linear models and empirical Bayes moderation in the context of RNA-seq data. This approach can help in experiments with complex experimental designs, because of Limma's ability to borrow information across genes to obtain more precise estimates of gene expression changes. The limma-voom pipeline thus offers flexibility and efficiency, especially in handling experiments with multiple factors and covariates (Ritchie et al., 2015).

A key difference between DESeq2 and limma-voom is in their treatment of count data. While limma-voom first transforms the counts to log2-CPM (counts per million) and then analyzes using the linear models defined in limma, DESeq2 works directly on raw counts and then fits the data from each gene into a negative binomial model. This difference in methodology influences their performance, particularly in terms of sensitivity and specificity under various experimental conditions and sample sizes (Liu et al., 2021). Although DESeq2 is often praised for its robustness in experiments with low sample sizes, limma-voom is preferred for its flexibility and its ability to analyze complex experimental designs (Law et al., 2014).

Moreover, DESeq2 uses Benjamini-Hochberg correction for p-value adjustment, while limma incorporates several techniques of p-value adjustment, which allows researchers to choose the method which best aligns with their experiment design and aims.

1.8.2 Batch Correction

In RNA-seq analysis and experiments, batch effects may show up due to the differences between samples that are not part of the experimental design. These can include different batches of reagents, different sequencing runs, or any other non-biological factors that can introduce variability in the data. Batch effects can critically affect downstream statistical analysis and if not done properly, it can lead to loss of biological data (Sprang et al., 2022). Batch correction is therefore needed to remove such technical variations.

Among the methods and tools available for Batch correction, ComBat-seq (Y. Zhang et al., 2020) and Limma have emerged as the most commonly used ones. ComBat-seq which is derived from the ComBat method specifically for RNA-seq data, uses an empirical Bayes model to adjust for batch effects to remove and technical variability while at the same time, keeping the biological variance intact (Y. Zhang et al., 2020).

Limma applies its linear modeling techniques to RNA-seq count data, fit with the direct model of batch correction. In addition, the limma option can allow for including batch as one of the covariates in the analysis to account for batch effects, therefore developing a very powerful framework for identifying differentially expressed genes (Ritchie et al., 2015).

There are other important batch correction tools such as SVA (Leek et al., 2012) and RUVSeq (Risso et al., 2014). SVA (Surrogate Variable Analysis) can identify and estimate directly all the hidden batch effects or other unwanted variations. RUVSeq (Remove Unwanted Variation from RNA-seq Data) corrects both known and unknown sources of variation through factor analysis methods that will eventually have contributed to the increase in sensitivity and a decrease in false-positive rates during tests for differential expression analysis.

1.8.3 Meta-analysis of Datasets

Meta-analysis of RNA-seq datasets is a statistical method that can be used to integrate findings from multiple studies to detect signals that are highly correlated and consistent in terms of gene expression, pathways, or markers within one or more experimental conditions. This approach can increase the scalability of results, which expands the findings to multiple studies rather than just based on standalone studies. Minor but consistent effects that may show up to be false negatives in individual studies can be detected by improved statistical power of meta-analysis (Choi et al., 2003).

One of the most basic approaches for meta-analysis is the Fisher's Inverse Chi-square test. This method combines individual P-values from different datasets into a single statistic,

$$S = -2\log(\prod P_i)$$

where S corresponds to a Chi-Square distribution, with $2I$ degrees of freedom when assuming the null hypothesis across all studies (Fisher, 1932). This method does not require additional analyses and is therefore widely used. However, this method cannot estimate the average effect

size or its direction or the magnitude of differential expression across datasets (Hong & Breitling, 2008).

Stouffer's method, an alternative to the Fisher's method, uses the inverse normal distribution (Stouffer et al., 1949). The method evolved further into converting P-values to Z-scores and then averaging them, which provides a more robust aggregated statistic in studies which have heterogeneous results (Yoon et al., 2021). This is particularly useful in studies of unequal sizes, since it gives equal weights to the contribution of each study, unlike the Fisher's method which is influenced by the number of p-values combined (Whitlock, 2005).

The Binomial test for meta-analysis assesses the number of studies that show a statistically significant result against the number expected by chance (Wilkinson, 1951). It is a direct method that does not rely on the combination of p-values but rather on the significance outcomes of individual studies. Assuming that the joint null hypothesis is *true*, the binomial approach is a way to determine whether there is a "excess of significant results" within the set of K tests (Cinar & Viechtbauer, 2022).

The Bonferroni (Neyman and Pearson, 1928) and Tippett's (Tippet, 1931) methods offer corrections for multiple testing, crucial in high-throughput data analysis studies, where the volume of simultaneous tests can inflate the rate of type I errors - null hypothesis is rejected, even though it is true (Cinar & Viechtbauer, 2022). The Bonferroni method adjusts for multiple testing by dividing the desired overall alpha level (hypothesis-wise Type I error rate) by the number of tests performed (k) and uses this as a significance threshold (Cinar & Viechtbauer, 2022). A less conservative approach for multiple test correction compared to the Bonferroni method is the Tippett's method. The Tippett's method sets an adjusted alpha level such that the overall risk of Type I errors (false positives) through all the tests remains at the desired significance level. For this method, the joint null hypothesis is rejected if the smallest p-value is less than or equal to $1 - (1 - \alpha)^{1/k}$. The combined p-value is calculated as

$$pc = 1 - (1 - \min(p_1, \dots, p_k))^k,$$

which is then compared with the critical alpha level, to assess the significance of the joint null hypothesis (Tippett, 1931; Dunn, 1958; Šidák, 1967). To apply Tippett's method, one would reject the null hypothesis formed by the combination if the minimum of the individual study p-values is smaller than the adjusted threshold. This makes it especially useful when at least one study shows a strong effect, but it relies on the assumption that all tests are independent of each other.

One of the packages recently developed for meta-analysis of P values, is the poolr package (Cinar & Viechtbauer, 2022). This package increases the capacity of doing meta-analyses with RNA-seq data and other types of genomic datasets by providing a framework for combining P values from both independent and dependent tests. This tool is particularly important in addressing the limitations of traditional methods, such as Fisher's Inverse Chi-square test, which falls short in accounting for the dependencies between tests or in estimating effect sizes. Poolr attempts to overcome these problems with several methods, including adjustments for the effective number of tests to handle dependencies among p-values and the use of empirically derived null distributions to improve accuracy. The package allows the use of 6 base methods; Fisher's method, Stouffer's method, Inverse chi-square method, Binomial test, Bonferroni method, and Tippett's method (Cinar & Viechtbauer, 2022).

2. Aims and Rationale

TP53 is the most researched protein in cancer. However, no comprehensive or modular database exists that can compile multiple RNA-seq datasets for analysis at the same time. Checking each dataset one by one to address the degree of *TP53* target expression modulation, may be very extensive and time consuming. Although there are several Shiny R based applications in the literature, a comprehensive tool that can analyze multiple expression datasets from start to end simultaneously, while allowing for multivariate statistics and network visualizations, does not exist.

Moreover, the case of drug resistance in breast cancer is very alarming and new combination therapies may be needed to tackle this problem. Previously in our lab, it was established that *CHRNA5* when silenced, can upregulated *TP53* or pathways related to it (Koker et al., 2018); and this was confirmed by microarray and RNA-seq experiments. This thesis builds further on these experiments by also checking the effect of silencing *CHRNA5* in *TP53* mutant cell lines, developed by Burcin Aras in our lab (Arıcı, 2022).

Continuing these projects, the main aims of this thesis are the following;

- Develop a *TP53* database called TP53 META to analyze and compare multiple datasets simultaneously. To do this the application should be able to provide a user-friendly interface, where the application can perform DGE analysis, meta-analysis between datasets, gene enrichment analysis, TF enrichment, and visualization of the data across datasets for better understanding of *TP53* modulated datasets.
- Perform preprocessing and multi-omics analyses of the heterozygous mutant MCF7 cell lines; integrate them into and further elaborate on the possible pathways regulated by silencing *CHRNA5* to induce *TP53* targets based on heterozygous mutant and homozygous WT overexpressing cell lines.

- Analyze the role of doxorubicin and its drug resistance in these mutant cell lines alone or in combination with si*CHRNA5* to tackle this problem of drug resistance.
- Develop a synergy assessment methodology based on the Limma framework to better annotate genes as synergistic/antagonistic in RNA-seq datasets and compare this with the previously developed method in our lab (Keşküş, 2021).



3. Materials and Methods

3.1 RNA-seq analyses

There are three different experiments of RNA-seq datasets from our lab, which are included in the TP53-META application. Two of these datasets were processed from scratch for the purpose of this thesis. Table 3-1 shows these particular datasets, and associated summary statistics for them. For Data 2 and Data 3, they were analyzed prior to batch correction and also after batch correction while Data 1 were first analyzed by Ayse G. Keskus (Keşküş, 2021). Since the TP53-META application does not apply batch correction, we provide files for the batches as separate files and as batch corrected files from Combatseq (Y. Zhang et al., 2020).

3.1.1 Data acquisition

The selected online datasets present in the application (Table 3-1), were acquired from NCBI GEO (Barrett et al., 2013). The initial treatment using si*CHRNA5*/si*TP53* in MCF7 cell line, RNA extraction and analysis was done by Dr. Ayse gokce Keskus (Keşküş, 2021). Cell culture studies and preparation of stable isogenic mutant cell lines, using MCF7 cell line were done by Burcin Aras (Arıcı, 2022). Doxorubicin treatment along with si*CHRNA5* and RNA extraction in these cell lines was performed by Cagdas Yanyatan (MSc) in our lab. Sequencing of the samples was outsourced as mentioned in Table 3-1.

Table 3-1. Summary of the datasets present in the TP53-META application.

Data Features	siData (Data 1) Konu lab	Mutant data (Data 2) Konu lab	Doxorubicin Data (Data 3) Konu lab	GSE47042 (Janky et al., 2014)	GSE173364 (Hassin et al., 2022)	GSE68248 (Walerych et al., 2016)
Samples	4	5	15	2	8	2
Platform	RNA-seq	RNA-seq	RNA-seq	RNA-seq	RNA-seq	RNA-seq
Cell line	MCF7	MCF7 (stably expressing <i>TP53</i> mutations)	MCF7 (stably expressing <i>TP53</i> mutations)	MCF7	SW480	MDA-MB-231
Treatments	si <i>CHRNA5</i> , si <i>TP53</i> , si <i>CHRNA5</i> +si <i>TP53</i> (combination), siScrambled	EV, WT, R175H, R273H, R248W	EV_Doxo_A5, EV_Doxo, EV_A5, EV, R175H_Doxo_A5, R175H_Doxo, R175H_A5, R175H, R273_Doxo_A5, R273H_A5, R273H, WT_Doxo_A5, WT_Doxo, WT_A5, WT	Control, Nutlin	<i>TP53_KO</i> , <i>TP53_KO_R175H_OE</i> , <i>TP53_KO_R273H_OE</i> , shP53, shP53_R175H_OE, shP53_R273H_OE	si <i>TP53</i> , siControl
Replicates	2 each	3 or 2	2 or 1 (for some samples)	2 each	3 or 2	3
Sequencing company	BMLabosis	GENera	Szaomics	-	-	-
Paired-end	Yes	Yes	Yes	No	No	Yes

Table 3-2. Sequence data analysis results for the first batch of the doxorubicin (Data 3) treated mutant samples.

#	Sample ID	Mreads (Fragment)	Mreads(R1+R2)	>=Q30(%)
1	EV_Doxo_A5_2	58.1	116.2	0.867
2	EV_Doxo_sc_2	17.1	34.2	0.835
3	EV_Serum_A5_1	60.6	121.2	0.866
4	EV_Serum_A5_2	56.5	113	0.866
5	EV_Serum_sc_1	60.4	120.8	0.864
6	EV_Serum_sc_2	55.6	111.2	0.867
7	R175H_Doxo_A5_2	48.3	96.6	0.868
8	R175H_Doxo_sc1	61.7	123.4	0.865
9	R175H_Serum_A5_2	48.2	96.4	0.863
10	R175H_Serum_sc_1	68.3	136.6	0.87
11	R175H_Serum_sc_2	74.7	149.4	0.866
12	R273H_Doxo_A5_1	63.1	126.2	0.856
13	R273H_Doxo_A5_2	56.1	112.2	0.866
14	R273H_Serum_A5_1	62.3	124.6	0.862
15	R273H_Serum_A5_2	56.9	113.8	0.867
16	R273H_Serum_sc_1	48.5	97	0.868
17	R273H_Serum_sc_2	63	126	0.868

Mreads : Total million reads (or pairs) for a sample

Q30 : Percentage of sequences with 1/1000 or less error

Table 3-3. Sequence data analysis results for the second batch of the doxorubicin (Data 3) treated mutant samples.

#	SAMPLE ID	Mreads (Fragment)	Mreads(R1+R2)	>=Q30(%)
1	R175H_Serum_A5_1	55.4	110.8	0.859
2	WT_Doxo_A5	76.6	153.2	0.86
3	WT_Doxo_SC	75	150	0.862
4	WT_Serum_A5_1	40.4	80.8	0.864
5	WT_Serum_A5_2	96.7	193.4	0.863
6	WT_Serum_SC_1	55.5	111	0.86
7	WT_Serum_SC_2	58.3	116.6	0.862

Table 3-4. Sequence data analysis results for the mutant samples (Data 2).

Sample Name	Total Reading	Aligned Reading	Alignment Rate
EV1	40,348,734	40,001,135	99.14%
WT1	40,814,550	40,607,325	99.49%
WT2	38,146,130	37,970,151	99.54%
R273H1	40,229,522	40,018,822	99.48%
R273H2	42,353,462	42,126,198	99.46%
R175H1	43,658,816	43,371,030	99.34%
R175H2	39,206,554	38,921,310	99.27%

3.1.2 RNA-seq data processing - QC, Alignment and Quantification

Quality control and alignment for Data 1 was done previously in our lab, by Dr. Ayse Gokce Keskus (Keşküş, 2021).

3.1.2.1 Mutant samples (Data 2)

To confirm the development of the isogenic mutant stable cell lines by Burcin Arıcı (Arici, 2022), cell culture experiments were performed and RNA was extracted (Appendix A2) and the samples were sent for RNA-sequencing. mRNA extracted from stable MCF-7 cells with empty vector, *TP53* WT and *TP53* mutation (R175H, R273H, R248W) vectors were sent to GEN ERA Diagnostics in duplicate (Arici, 2022). The sequencing study was carried out using the Illumina NovaSeq 6000 next-generation sequencing platform, with an average of 30 million 150 bp long paired-end reads per sample.

Raw fastq files were uploaded on Galaxy (useGalaxy.org). Galaxy is a comprehensive online browser which helps the users in storing and the computation and analysis of high-throughput data (Afgan et al., 2022). It contains a variety of bioinformatics tools which run on the Galaxy server. These tools if run on local machines may need high computation and hardware requirements.

Quality control of the fastq files was performed using the FastQC tool (version 0.12.1) that allows the results to be viewed as an html file on galaxy (Andrews, 2010). One of the most often utilized tools for quality control analysis in high-throughput sequencing is FastQC. It provides basic statistics such as sequence length, mapped reads and GC percentage. It also shows metrics such as per base sequence quality or per sequence quality scores which are important measures showing the quality scores based on their position in the reads. Further metrics include, duplication levels, over-represented sequences, per base N content, adapter and Kmer content. MultiQC tool (version 1.11) (Ewels et al., 2016) was further used to aggregate all the QC reports together for easier visualization (Figure 3.1). Based on the QC

results, no trimming was required and adapter sequences were removed by GENERA before they provided the fastq files.

After QC checks, the fastq files were aligned using HISAT2 (version 2.2.1), which is a fast and sensitive alignment tool (Kim et al., 2015). The paired-end files were mapped to the hg38 human genome assembly (Genome Reference Consortium, Human Build 38 (GRCh38)). BAM files obtained from HISAT2 were then used as input for the featureCounts tool (Liao et al., 2014) to create a read count matrix with samples as columns and rows as genes/features.

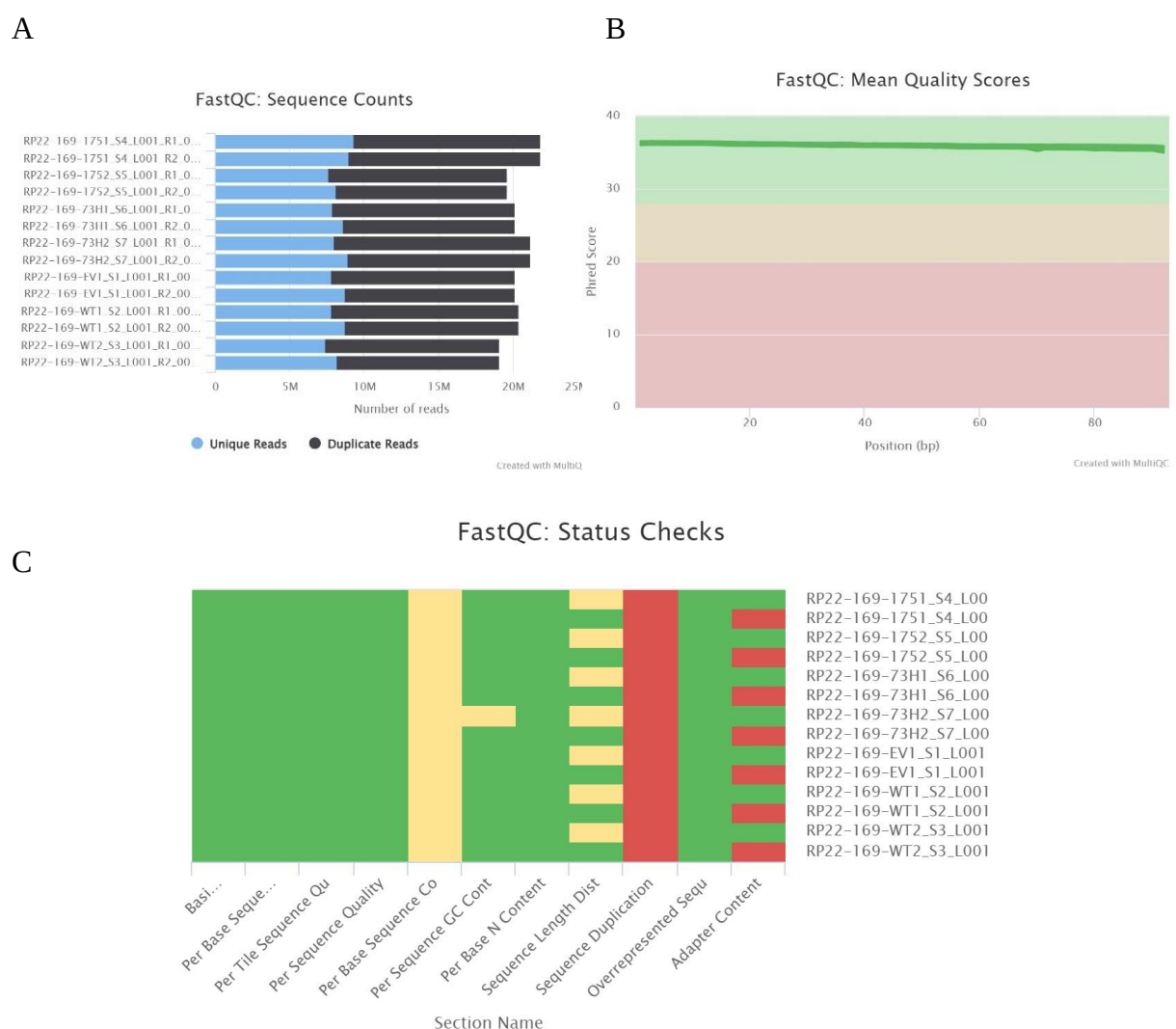


Figure 3.1 A, B Sequence counts and mean quality scores for the samples in our mRNA-seq mutant dataset (dataset 2), as analyzed by fastQC and aggregated by multiQ. **C** Each FastQC section is marked to indicate if results are completely normal (green), somewhat abnormal (orange), or highly unusual (red).

3.1.2.2 Doxorubicin and si*CHRNA5* treated samples (Data 3)

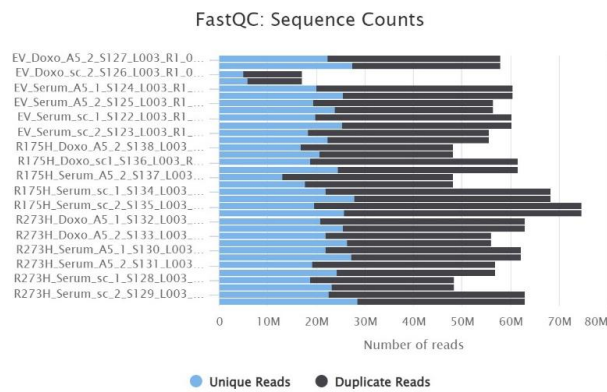
To analyze and investigate the transcriptomic effects of doxorubicin along with si*CHRNA5* in the stable isogenic mutant MCF7 cell lines developed by Burcin Arıcı (Arici, 2022); RNA was isolated and sent for RNA sequencing. Total RNA sequencing was done with Novaseq-6000 platform at 100bp paired-end and 25M read depth parameters. Raw fastq files were provided with the purchased service.

The quality control was carried out with the same tools and workflow as described for the mutant samples.

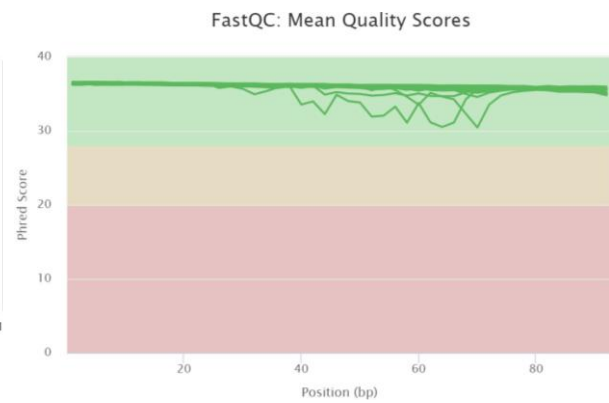
One of the replicates of the EV_Doxo file showed some variation, in the mean quality scores and showed less mapped reads as can be observed in Figure 3.2. The variability in the mean quality scores for this one sample, even though it passed the quality thresholds, could be due to various technical artifacts, or variation in the sequencing cycles. Figure 3.3 shows the summary of the total status checks for our RNA-seq data sets.

After QC checks, the fastq files were aligned on Galaxy using the same workflow and tools as mentioned for the mutant cell lines.

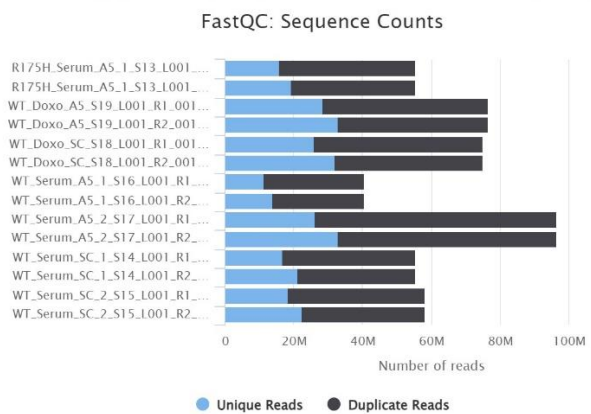
A



B



C



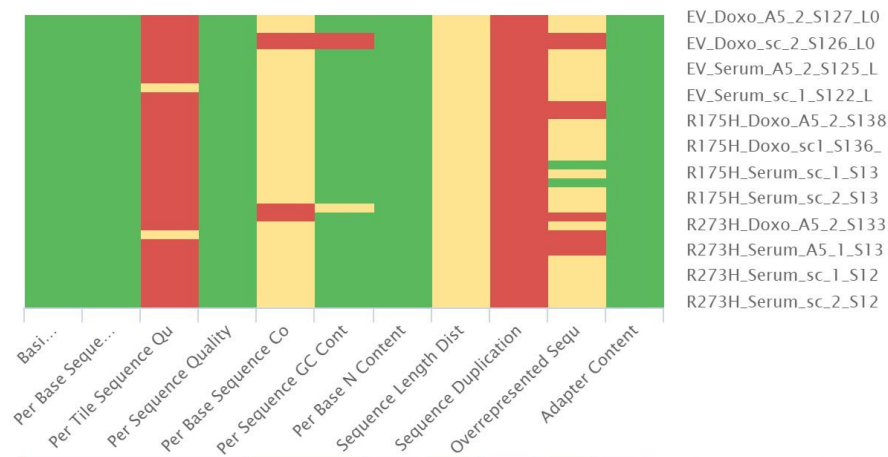
D



Figure 3.2 Sequence counts and mean quality scores for all the samples in our RNA-seq doxorubicin treated dataset (dataset 3), as analyzed by fastQC and aggregated by multiQC. **A, B** Sequence counts and mean quality scores of the first batch of the dataset. **C, D** Sequence counts and mean quality scores for the second batch of the dataset.

A

FastQC: Status Checks



B

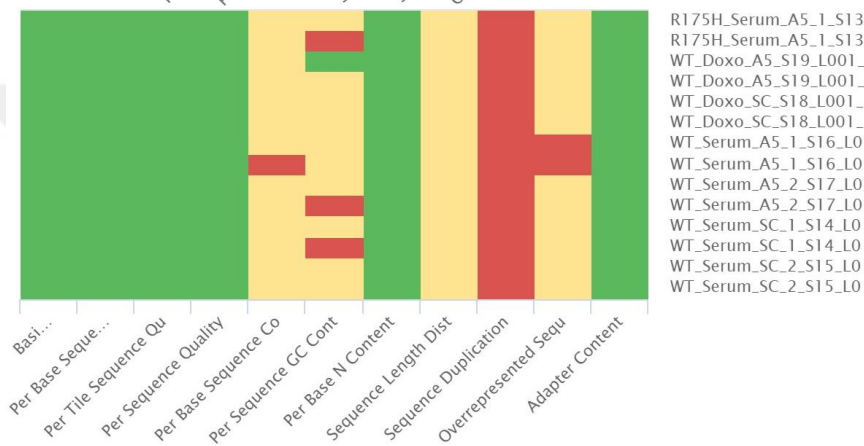


Figure 3.3 A, B All the fastQC checks presented here for all the dataset 3 (doxorubicin treated mutant cell lines). Green shows that that the filter is passed, while yellow indicates moderate quality and red shows low quality.

3.2 Transcriptomic data exploration:

3.2.1 DGE analysis

The data analysis included library size bar plots, PCA and boxplots using base R functions, which was followed by differential gene expression analysis done by either Deseq2 (Love et al., 2014) or Limma (Ritchie et al., 2015). BiomaRt (Durinck et al., 2009) along with org.Hs.eg.db package (Carlson, 2022) was used to get the human annotation for labeling the genes in the count matrix.

DGE analysis from Deseq2/Limma provided log-transformed fold change values, p values, adjusted p values and other statistics. These results were visualized as volcano plots, scatter plots, and heatmaps using various packages in R mentioned in table 3-5. Batch correction on the datasets was performed by ComBat-seq or Limma by introducing a batch equation in the design.

3.2.2 Downstream functional analysis

The functional analysis of the datasets included the Gene Set Enrichment Analysis (GSEA), Over Representation Analysis (ORA), Transcription Factor Enrichment (TFE) and profiling of particular gene sets. The GSEA and the ORA results were reported for specific genes which fulfilled particular statistical filters. For these analyses, the annotation data were used from the Kyoto Encyclopedia of Genes and Genomes (KEGG) Database (Kanehisa et al., 2012) and the Gene Ontology (GO) Resource (Aleksander et al., 2023).

The TFE (transcription factor enrichment) was done based on the Dorothea database using the DoRothEA package in R (Garcia-Alonso et al., 2019). The transcription factors with the most overlapped genes with the user defined filter were plotted as networks using visnetwork (Almende B.V., et al., 2022). The particular gene lists used for filtering the genes were based on prior research present in the literature about *TP53* pathways or related pathways (Fischer et al., 2016; Uxa et al. 2019).

Table 3-5. Tools and packages reference list for data analysis and the development of the shiny webserver.

Tool or Package	Reference	Identifier
Online tools		
Galaxy platform	Afgan et al., 2022	https://usegalaxy.org/
FastQC	Andrews, 2010	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
MultiQC	Ewels et al., 2016	https://multiqc.info/
Hisat2	Kim et al., 2015	https://daehwankimlab.github.io/hisat2/
Ncbi geo	Barrett et al., 2013	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi
Human Genome assembly	Genome Reference Consortium	https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.26/
cBioportal	Cerami et al., 2012	https://www.cbioportal.org/
stringDB	Szklarczyk et al., 2023	https://string-db.org/
R packages		
limma	Ritchie et al., 2015	https://bioconductor.org/packages/release/bioc/html/limma.html
edgeR	Robinson et al., 2010	https://bioconductor.org/packages/release/bioc/html/edgeR.html
DESeq2	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
viridis	Garnier et al., 2023	https://cran.r-project.org/package=viridis
tidyverse	Wickham et al., 2019	https://www.tidyverse.org/
ggplot2	Wickham, 2016	https://ggplot2.tidyverse.org/
dplyr	Wickham et al., 2023	https://CRAN.R-project.org/package=dplyr
poolr	Cinar & Viechtbauer, 2022	https://cran.r-project.org/web/packages/poolr/index.html
shiny	Chang et al., 2024	https://shiny.rstudio.com/
shinydashboard	Chang et al., 2021	https://rstudio.github.io/shinydashboard/
rintrojs	Ganz, 2016	https://cran.r-project.org/package=rintrojs
fresh	Perrier et al., 2020	https://dreamrs.github.io/fresh/
shinyWidgets	Perrier et al., 2024	https://dreamrs.github.io/shinyWidgets/
DT	Xie et al., 2023	https://rstudio.github.io/DT/
heatmaply	Galili et al., 2018	https://talgalili.github.io/heatmaply/
plotly	Sievert, 2020	https://plotly-r.com/
visNetwork	Almende B.V., et al., 2022	https://cran.r-project.org/web/packages/visNetwork/index.html
shinyBS	Bailey, 2022	https://cran.r-project.org/web/packages/shinyBS/index.html
shinycssloaders	Sali and Attali, 2020	https://cran.r-project.org/package=shinycssloaders
EnrichmentBrowser	Geistlinger et al., 2016	https://bioconductor.org/packages/release/bioc/html/EnrichmentBrowser.html
stringr	Wickham, 2023	https://cran.r-project.org/package=stringr
markdown	Xie et al., 2023	https://cran.r-project.org/package=markdown
msigdb	Dolgalev, 2022	https://cran.r-project.org/package=msigdb
hyper	Federico & Monti, 2020	https://bioconductor.org/packages/release/bioc/html/hyper.html
enrichplot	Yu et al., 2022	https://bioconductor.org/packages/release/bioc/html/enrichplot.html
clusterProfiler	Wu et al., 2021	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
DOSE	Yu et al., 2015	https://bioconductor.org/packages/release/bioc/html/DOSE.html
clusteprofiler2	Kolberg et al., 2020	https://cran.r-project.org/package=gprofi2
EnhancedVolcano	Blighe et al., 2022	https://bioconductor.org/packages/release/bioc/html/EnhancedVolcano.html
biomaRt	Durinck et al., 2009	https://bioconductor.org/packages/release/bioc/html/biomaRt.html
ComBat-seq	Y. Zhang et al., 2020	https://github.com/zhanguyqing/ComBat-seq
DoRothEA	Garcia-Alonso et al., 2019	https://bioconductor.org/packages/release/data/experiment/html/dorothea.html

3.3 Development of TP53-META application

The TP53-META application was developed using R and R shiny framework which allows to build and host interactive web applications (Chang et al., 2024). The R shiny framework includes the integration of a UI and a Server file. The application's user interface is based on CSS and HTML components which enhance the layout and design. Custom HTML elements were incorporated to facilitate advanced features and allow for precise placement of UI elements. The UI was carefully designed to allow the user to navigate freely between the tabs and download or save any visualizations. The user interface has a landing page from which users can select their datasets and set parameters for analysis. This application makes use of shinyWidgets (Perrier et al., 2024) as a part of the UI, which permits the easy implementation of dropdown menus and action buttons providing a streamlined workflow with easy navigation. The server file complements the features of the UI and allows for all the back-end processing and real-time analysis. The server analysis is reactive and allows for dynamic processing of the selected data.

Various R packages were utilized in the application's development (Table 3-5). The dplyr (Wickham et al., 2023) and tidyverse (Wickham et al., 2019) packages were used to filter and structure the data. For the differential gene expression analysis, the Deseq2 (Love et al., 2014) and Limma (Ritchie et al., 2015) packages were integrated into the application and the meta-analysis of their p-values was performed based on the poolr (Cinar and Viechtbauer, 2022) package. The various visualizations such as bar plots and volcano plots, heatmaps were based on ggplot (Wickham, 2016) and plotly (Sievert, 2020) packages. Plotly and heatmaply (Galili et al., 2017) extend the ggplots' ability to make the plots interactive. The clustering methods for the heatmaps are defined by various dropdown options, giving the user the freedom for their analysis.

The networks were all based on the visnetwork package (Almende B.V., et al., 2022) which allows for interactive visualizations for exploring the datasets. Enrichment analysis was based on the clusterprofiler package (Wu et al., 2021). Other enrichment analysis included the integration of various databases such as Dorothea (Garcia-Alonso et al., 2019), which is a Transcription Factor target database; DisGeNET (Piñero et al., 2019), which is a disease-gene database; and MSigDB (Dolgalev, 2022), which is a molecular signatures database. The datasets incorporated in the application and the gene lists were acquired as mentioned before in this section.

Our RNA-seq datasets on the application are uploaded with and without batch effect correction through Combatseq (Y. Zhang et al., 2020). Combatseq normalizes the read counts and gives integer values which can also be used as an input for Deseq2. Public datasets were obtained from NCBI GEO (Barret et al., 2013) and their raw counts were extracted directly from the database.

3.4 Synergy Calculations

To devise a method for better annotation of synergistic and antagonistic genes, limma's statistical framework was used. The synergy annotation for each gene, based on the p-values obtained for selected pair-wise comparisons as well as the combinatorial and synergy logFCs, are mentioned in the results section in more detail.

To calculate the scores for performance of different methodologies, the randomforest package (Liaw and Wiener, 2002) was used to calculate the accuracy of the annotation. The data was split into 80:20 for train and test respectively. The features to train and predict were the p-values and the logFC values of the combinatorial and the respective synergy method score. The clustering and correlation were done with the apcluster (Bodenhofer et al., 2011) and umap (Konopka, 2023) packages in R.

4. Results

4.1 TP53 META shiny web server

The TP53-META is a shiny web server which compiles multiple public and in-house datasets in which *TP53* was modulated and allows for effective analysis and visualization of data all together. The application also allows for a meta-analysis of the p-values of the chosen comparisons in the datasets using the poolr (Cinar and Viechtbauer, 2022) package. In this section, I have used the public datasets, GSE47042, GSE173364 and GSE58248 and its subsets to demonstrate how to use TP53-META.

The application consists of a “Home” page, which includes the summary of the datasets and a tutorial. The “Analysis” page performs all selected comparisons and displays their analyses. The two primary tabs in the “Analysis” module are: “Data Summary” and “Results”. The “Data Summary” page makes possible to examine the dataset selected using raw data visualization and volcano, bar/box and PCA plots.

The “Results” tab is the main module in which there are seven sub-tabs under this tab: “Analysis”, “Scatter Plot”, “Heatmap”, “Functional Profile”, “DoRothEA”, “Treatment-Gene Networks”, “MsigDB” and “Compare Cluster”. All of these analyses are performed based on certain user-defined thresholds, such as p-value thresholds or log fold change thresholds. The application also consists of a set of cell regulatory gene lists, from which the user can select, or the user can upload their own gene lists to limit the number of genes for their downstream analyses. The packages used for the development of this web server are listed in Table 3-5. A simplified workflow of the web application is illustrated in Figure 4.1.

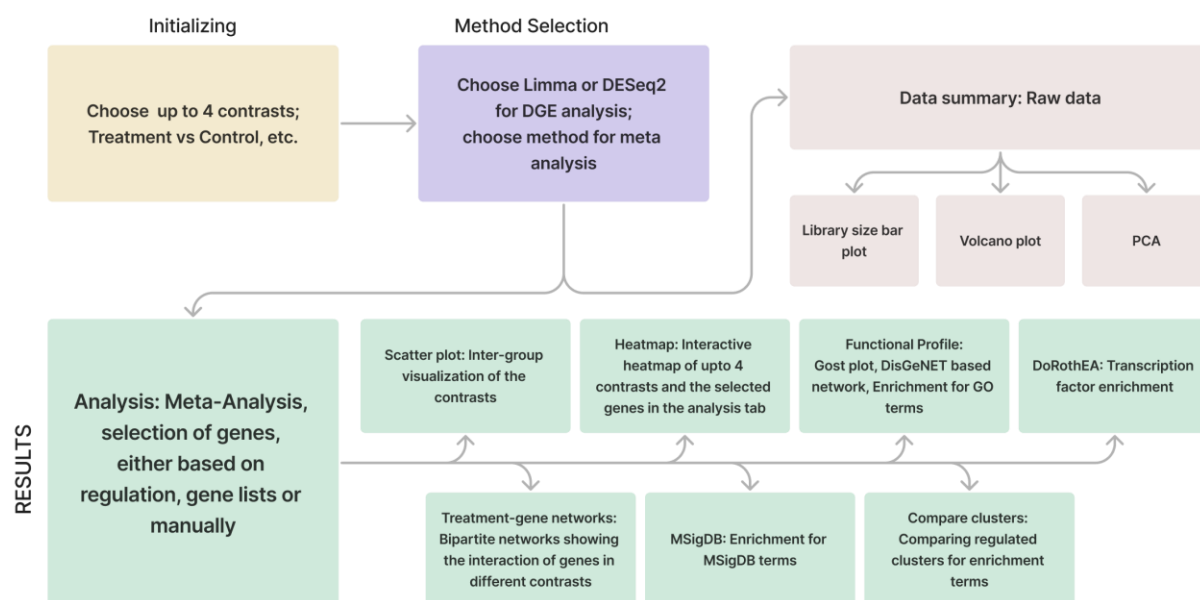


Figure 4.1 Simplified workflow of the TP53-META application.

4.1.1 Initialization and choosing contrasts

The TP53-META application starts by giving information about the datasets when selected and includes a tutorial (Figure 4.2). The application allows the user to choose a minimum of one (up to four) contrast/comparison(s), from the datasets provided (Table 3-1). The user has the ability to allocate each group as Control or Treatment from the dropdown menu, which allows comparisons between treatments from a particular dataset selected (Figure 4.3).

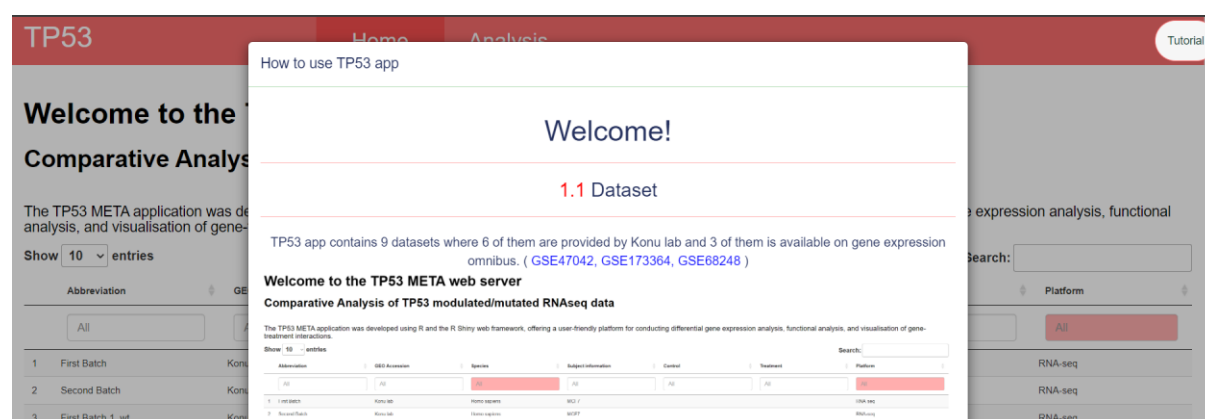


Figure 4.2 Initialization, summary of the datasets and the tutorial page.

After selecting the datasets and the contrasts, the user can choose a method for the meta-analysis of p-values, if more than one contrast has been chosen. This is integrated from the poolr package and the methods available to use are, “Fisher’s”, “Stouffer’s”, “Inverse chi-square”, “Binomial test”, “Bonferroni” and “Tippett’s”. The differential gene expression analysis can be done either with Limma-Voom or DESeq2 (Figure 4.3).

The interface displays four data selection panels (Data 1 to Data 4) with dropdown menus for dataset names and contrast pairs (Control vs Treatment). Below these panels are options for the meta-analysis method (Bonferroni method selected) and the DGE method (Limma-Voom selected). A red 'Analyze' button is present. At the bottom, there are checkboxes for columns to display: LogFC, Average Exp, t value, p value, p adj value, and B value.

Figure 4.3 Data selection and contrast comparison (Treatment vs Control). The user also selects the method for DGE and the meta-analysis of P-values. Here, the three different GSE datasets were chosen and the meta-analysis was selected as Bonferroni.

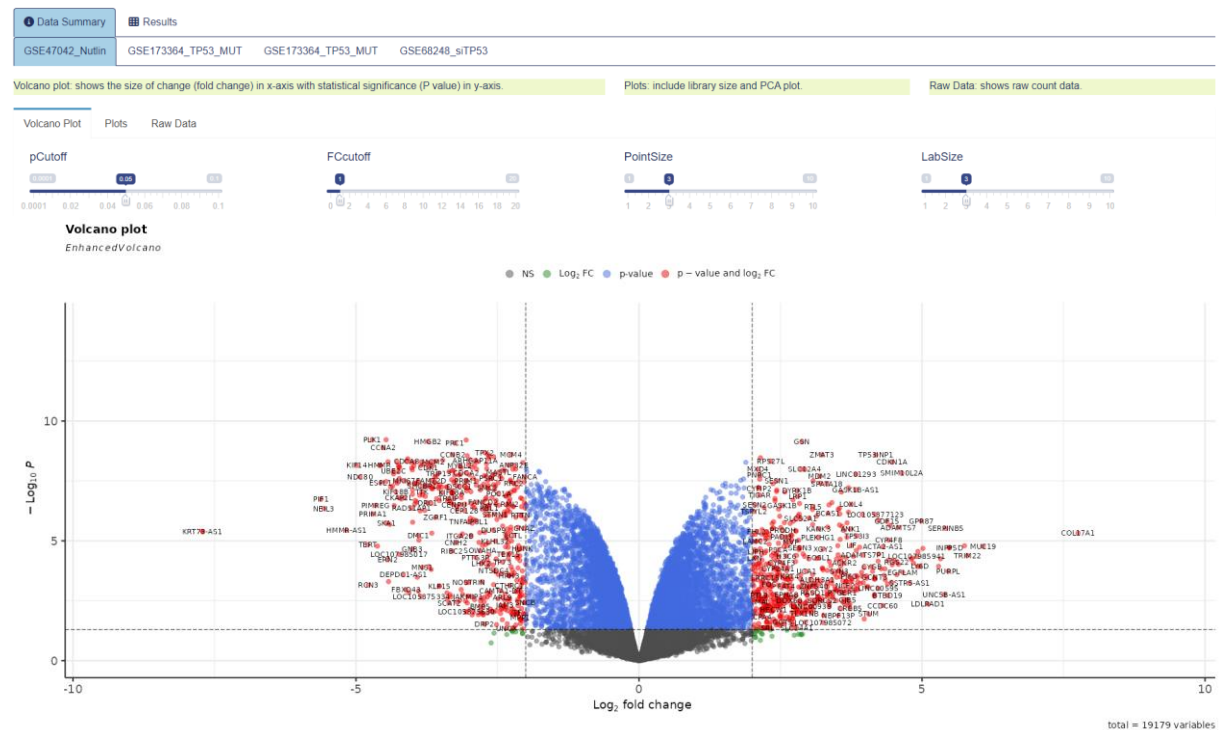
4.1.2 Data Summary

After selection of the contrasts/comparisons and the methods, the user can proceed to view the “Data Summary” tab that shows the raw statistics for each of the datasets separately. Here the user can visualize a volcano plot for the selected contrast to assess the overall differential expression of the contrast. The user via the slider inputs can change the colors, fonts, sizes and cutoffs manually (Figure 4.4A).

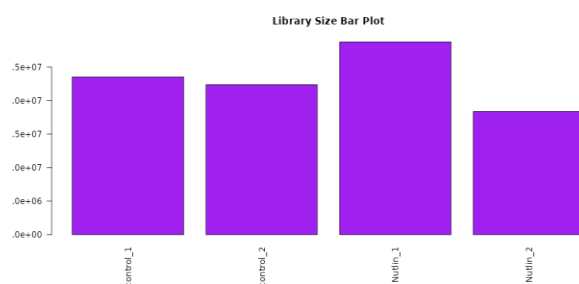
In the “Plots” sub-tab of the Data summary, the library size for the data can be visualized (Figure 4.4B), and the clustering of the samples for the data can also be viewed in a PCA (Principal Component Analysis) plot (Figure 4.4C). The library size plot gives information on the depth of sequencing and quality of the dataset. PCA allows the user to visualize the similarity between samples and/or replicates. These plots can help the user understand the quality of the data and distributions of the samples in a reduced dimension. Good quality datasets usually have replicates clustered together and clear separation in the different samples. They should also have large library sizes and equally distributed counts.

The user can view the count data under the “Raw data” sub-tab and can also download it if needed for analysis outside the web application. Except for the Volcano plot, all other visualizations would be the same if the contrasts selected are from the dataset.

A



B



C

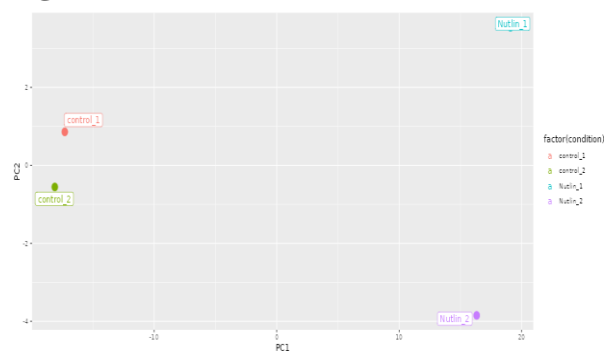


Figure 4.4 **A** Volcano plot for the chosen the contrast with log2FC plotted against the $-\log(P\text{-values})$. The red colored genes fulfil the criteria of having a p-value less than 0.05 and logFC of over 1. These cutoffs can be changed along with the label size. **B** Library size bar plot for all of the samples present in the dataset (GSE47042). **C** The respective PCA for the samples of the chosen dataset.

4.1.3 Gene selection

Under the *Results* tab, the first section is the “Analysis”. The table here shows the results of the DGE analysis and the meta-analysis. By default, the table shows the p-value and the logFC columns for all the contrasts. In addition to this if more than one contrast is selected, then three more columns are shown; “Meta.analysis.of.Pvalues”, “Sig.upregulated” and “Sig.downregulated”. The user can also select to show more columns such as the adjusted P-value etc. (Figure 4.3).

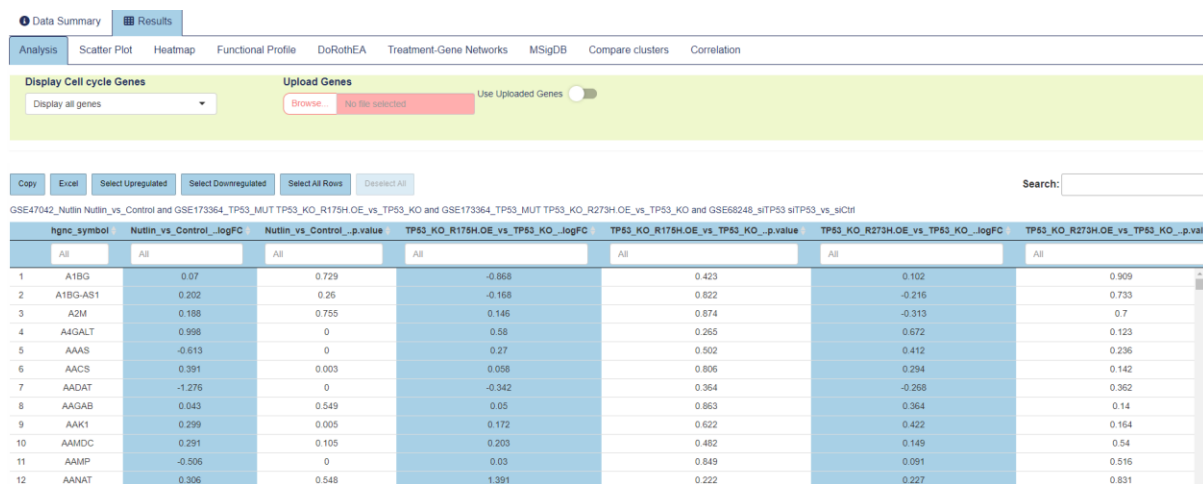
The meta-analysis of P values shows the combined (pooled) p-values for all the chosen contrasts. If only one contrast is chosen, this column does not show up. The sig.upregulated and the sig.downregulated columns show the direction of the logFC values, upregulated if positive and downregulated if negative. These columns show the number of contrasts with significant p-values (less than 0.05) and their directions.

For the analysis throughout the application, the user has to define or **select** rows from the table shown in figure 4.5 A. These genes can be selected by three different ways. The user can:

- manually select each gene by clicking on the row.
- choose from the predefined list uploaded in the application such as the DREAM targets etc., or upload their own gene list either as a CSV file or TXT file, with gene names in rows.
- choose based on upregulated or downregulated genes given the user defined cutoffs (Figure 4.5). These filter options show up once the *select upregulated* or *select downregulated* buttons are clicked. These filters can be applied to the chosen contrasts as can be seen in Figure 4.5B.

The selection of genes, based on the chosen gene lists or filters, is shown as another table just underneath the first one. Both of these tables can be downloaded as CSV files.

A



B

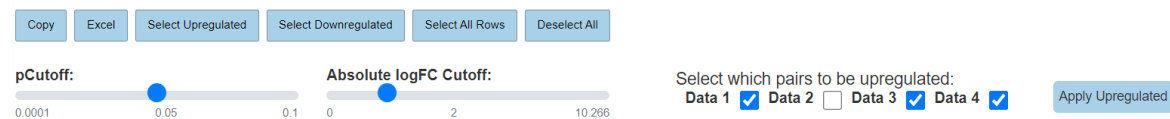


Figure 4.5 A Shows the options to select genes based on the predefined gene lists or an upload option. The table for the contrasts can be seen right underneath. The user has to select genes from this table to be used further in the analysis. **B** Shows the p-value and the logFC filters for upregulated selection and downregulated selection of genes across datasets. Using the check boxes, the user can choose which datasets to apply these particular filters to; for example, if a p-value cutoff of 0.05 is chosen and applied to only Data 1, the genes will be filtered only based on Data 1. However, if someone wants to choose significant genes across all contrasts, all the check boxes can be chosen.

The filters shown in Figure 4.5 B, when applied across all contrasts can be very strict. This is because the selection is then done by using genes with logFCs and P-values over certain threshold. For choosing both upregulated and downregulated genes together, the user can filter from the sliders shown at the top of the table. The p-values can be filtered for each contrast either separately, or according to the meta-analysis column. These selections prove to be helpful for the enrichment analysis, where the user may want to see certain pathways based on just upregulated genes or downregulated genes.

4.1.4 Scatter plot and heatmap

The “scatter plot” tab of the *Results* section, plots either the logFC or the p-value of the defined contrasts. These plots can be between the contrasts as well. This can help the user to see the trend and pattern of either the entire set of genes or only the chosen gene set.

As an example, I have obtained the following contrasts for the demonstration for the application (also shown in Figure 4.3);

- 1) *Nutlin vs control* (GSE47042)
- 2) *R175H Overexpression vs TP53 Knock Out* (GSE173364)
- 3) *R273H Overexpression vs TP53 Knock Out* (GSE173364)
- 4) *siTP53 vs siControl* (GSE68248)

This provides the user with a data table with P values and logFC values for each of the contrasts (Figure 4.5). Next I have chosen DREAM targets as the list from the pull-down menu, *Display cell cycle genes* (Figure 4.5 A) to limit the genes of interest. The scatterplot between the log FC values from the two contrasts for GSE173364 dataset, for the chosen gene set was plotted.

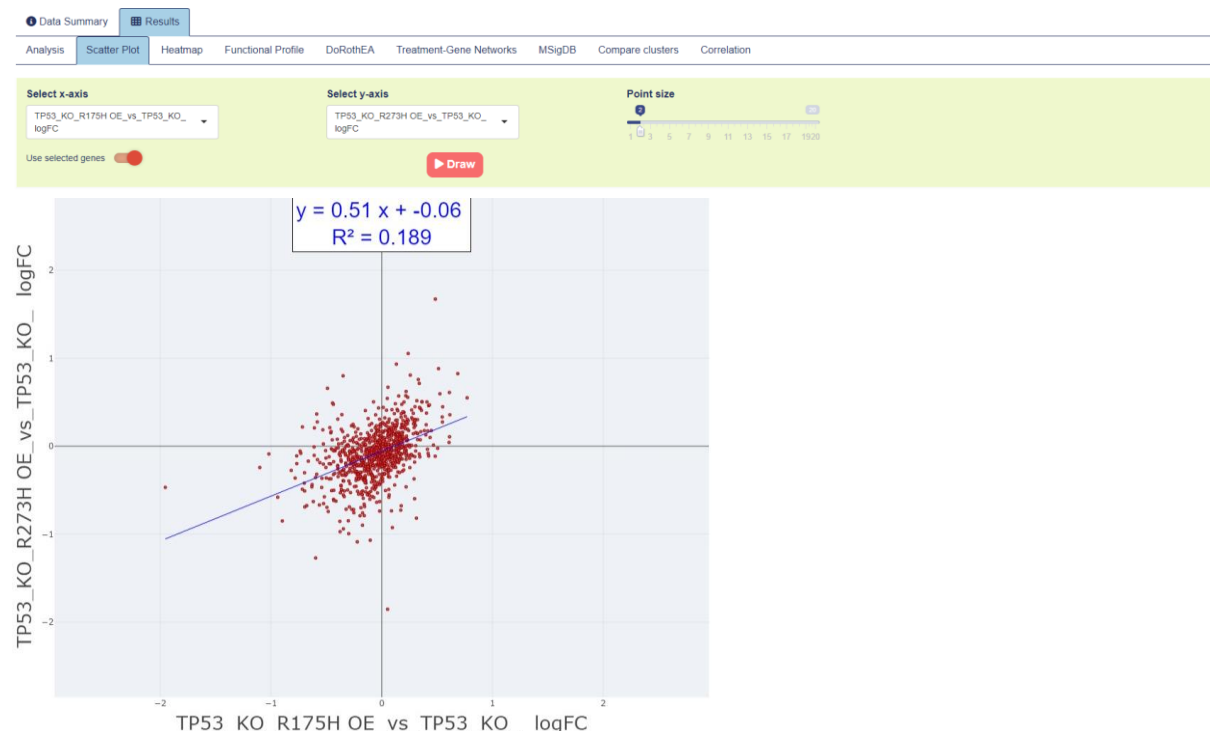


Figure 4.6 Scatter plot for the logFC values of the two chosen contrasts from GSE173364. Here the selected genes (DREAM targets) were plotted only instead of all the genes.

This overexpression of R175H and R273H vectors, even though had a linear relationship for the DREAM targets, they did not show an extensive variation in logFC values and were rather scattered around the origin. This goes in line with the hypothesis that *TP53* affects DREAM targets and cell regulatory genes and overexpression of these particular hotspot mutations (R175H and R273H) would dysregulate the cell cycle genes. This “scatter plot” tab also allows inter-contrast plotting of the chosen genes. For example in Figure 4.7, I plotted *Nutlin vs Control* contrast from the GSE47042 dataset on the y-axis and *R175H Overexpression vs TP53 Knock Out* from the GSE173364 dataset on the x-axis. Nutlin is known to be a p53 inducer in *TP53* WT expressing cells, and it shows that most of the DREAM targets appear to be downregulated when Nutlin is used, in contrast to the R175H overexpression.

This allows the user to plot different treatments against each other and visualize their transcriptomic effects, in contrast to each other. The scatter plots use the plotly framework and are interactive so the user can view the outliers and identify them.

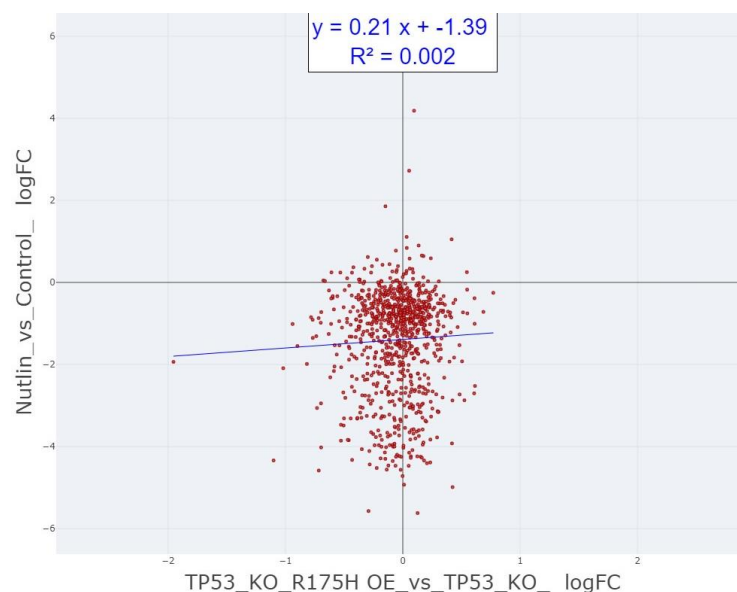


Figure 4.7 DREAM targets for contrasts from different datasets (GSE173364 vs GSE47042) plotted against each other. This allows for transcriptomic comparison of different datasets with respect to each other.

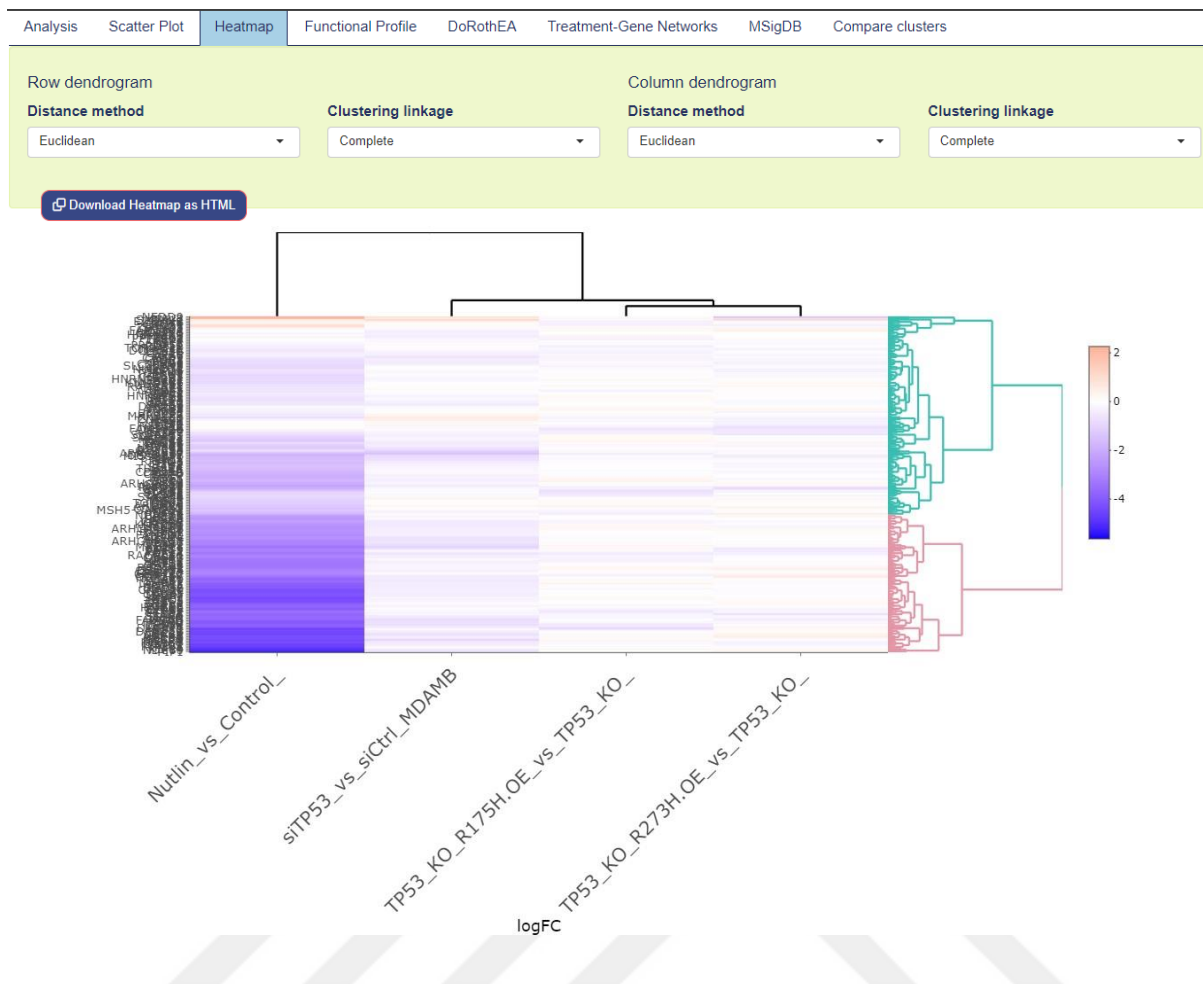


Figure 4.8 Heatmap of the chosen contrasts and the MMB-FOXM1 geneset with “Euclidean” defined as the distance method for both rows and columns; and “complete” defined as the linkage method for both rows and columns. The colors are based on the logFC values with red representing values over 0 and blue representing negative values. The genes can be visualized at the left and the samples are clustered at the top. This is an interactive plot so the user can zoom in/out to clearly see the gene names and their effect.

The “Heatmap” tab of the application plots the log fold change values of the chosen contrasts and the selected genes. The heatmap has been made using plotly and heatmaply packages, so that it is interactive and in case of too many genes, the user can zoom-in or hover over the row and observe the gene. The distance and the linkage of the clustering can be defined by various methods which are shown as drop-down options over the heatmap (Figure 4.8).

The most commonly used distance method is the “Euclidean”, which is calculated based on the Cartesian coordinates of the two points, using the Pythagoras theorem. The “manhattan” distance method uses the sum of the absolute differences of the coordinates. Other methods

include, "maximum", "canberra", "binary" and "minkowski". Similarly, the distance between sets of observations is calculated using the linkage criterion as a function of the pairwise distances between observations (Nielsen, 2016). As a result, the linkage specifies how the distances are aggregated. The methods available for linkage include "ward.D", "ward.D2", "single", "complete", "average" (= UPGMA), "mcquitty" (= WPGMA), "median" (= WPGMC) and "centroid" (= UPGMC). A comparison between the heatmap in Figure 4.8 can be made by using different distance and linkage methods by the user.

The heatmap of the four contrasts for the MMB-FOXN1 geneset showed that nutlin highly downregulated these genes (Figure 4.8). This is in fact what we expect since nutlin induces P53 and MMB-FOXN1 targets are downregulated by *TP53* in cell cycle, along with DREAM and B-MYB targets (Fischer et al., 2016). However, the *TP53* mutants and silencing *TP53* showed slightly upregulated genes and mostly unaffected genes. This is because *TP53* is non-functional in these cases, which causes the cell cycle to be dysregulated.

4.1.5 Functional Profile

The “functional profile” tab shows the enrichment of terms based on the selected genes in the analysis tab. The GSEA results are shown as a GOST plot (Kolberg et al., 2020), a table of the enrichment terms, their visualization the form of a dotplot, cnetplot and an emapplot (Wu et al., 2021). In the “functional profile” tab the user can also view a gene-disease association network or a heatmap for the selected genes based on the DisGeNET database (Piñero et al., 2019).

The GOST plot, using information and data from the Gene Ontology (GO) terms, biological pathways (BP), KEGG, Reactome, regulatory DNA elements, human disease gene annotations, and protein-protein interaction networks, performs functional profiling of the gene lists. The GOST function uses Fisher’s one-tailed test, to calculate a p-value, which is shown

on the y-axis. This P-value is a measure of the randomness of the intersection of the input gene and the ontology term. The x-axis shows the functional terms which are color-coded based on the database and the terms are grouped together with the size showing the number of genes in that particular ontology term (Figure 4.9A).

To demonstrate this, I selected a P-value cutoff of 0.01 and genes with a logFC of less than -2, for the first contrast (*Nutlin vs Control*, GSE47042). In the interactive Gost plot, the GO:BP (Gene ontology, Biological process) enriched terms were mostly related to cell cycle. These enriched terms can be further explored in the “GO terms Enrichment Analysis”. GSEA results can be viewed as multiple plots. In Figure 4.9B the Gene ontology table is shown. The size for minimal overlap can be changed which can allow the user to be more stringent when visualizing the ontology terms. The overlapping genes between the chosen gene set and the term are shown under the *geneID* column of the table. The P-value here also shows the probability of getting a number of genes overlapping with the total genes in the particular term.

These enrichment results are shown as either a dotplot, cnetplot or an emapplot (Figure 4.10 A-C). A dotplot shows the enriched gene sets or terms, where each dot represents a specific term and its position on the plot indicates the term’s significance. A cnetplot represents the relationship between the enriched terms and their linkage to the genes in those terms. The emapplot shows the enriched terms as a network. The terms which have overlapping genes are more likely to cluster together, making it easier for interpretation. These visualizations are crucial for the interpretation of the enrichment analysis and the functional significance of the chosen gene set.

The most significant enriched terms for this contrast included, nuclear division, chromosome segregation, mitotic nuclear division. I set the number of displayed terms at eight for Figure 4.10 A, B and four for Figure 4.10 C. The emap plot shows the overlap of the genes between the enriched terms as well and here DNA replication has no genes in common with

the other enriched terms (Figure 4.10 E). The cnet plot here shows cell cycle genes such as *CDC6*, *CDC20*, *AURKA*, *AURKB* among others, which are color coded based on the logFC of the contrast. These were all significantly downregulated, in line with the hypothesis that Nutlin induced p53, which in turn regulated these cell cycle genes.



Figure 4.9 **A** GOST plot of the selected genes. This is an interactive plot and the term can be viewed by clicking on the individual circle. **B** Table for the GO enrichment analysis results. It shows the overlap of the selected geneset with the terms present and the FDR values for the terms based on the overlap.

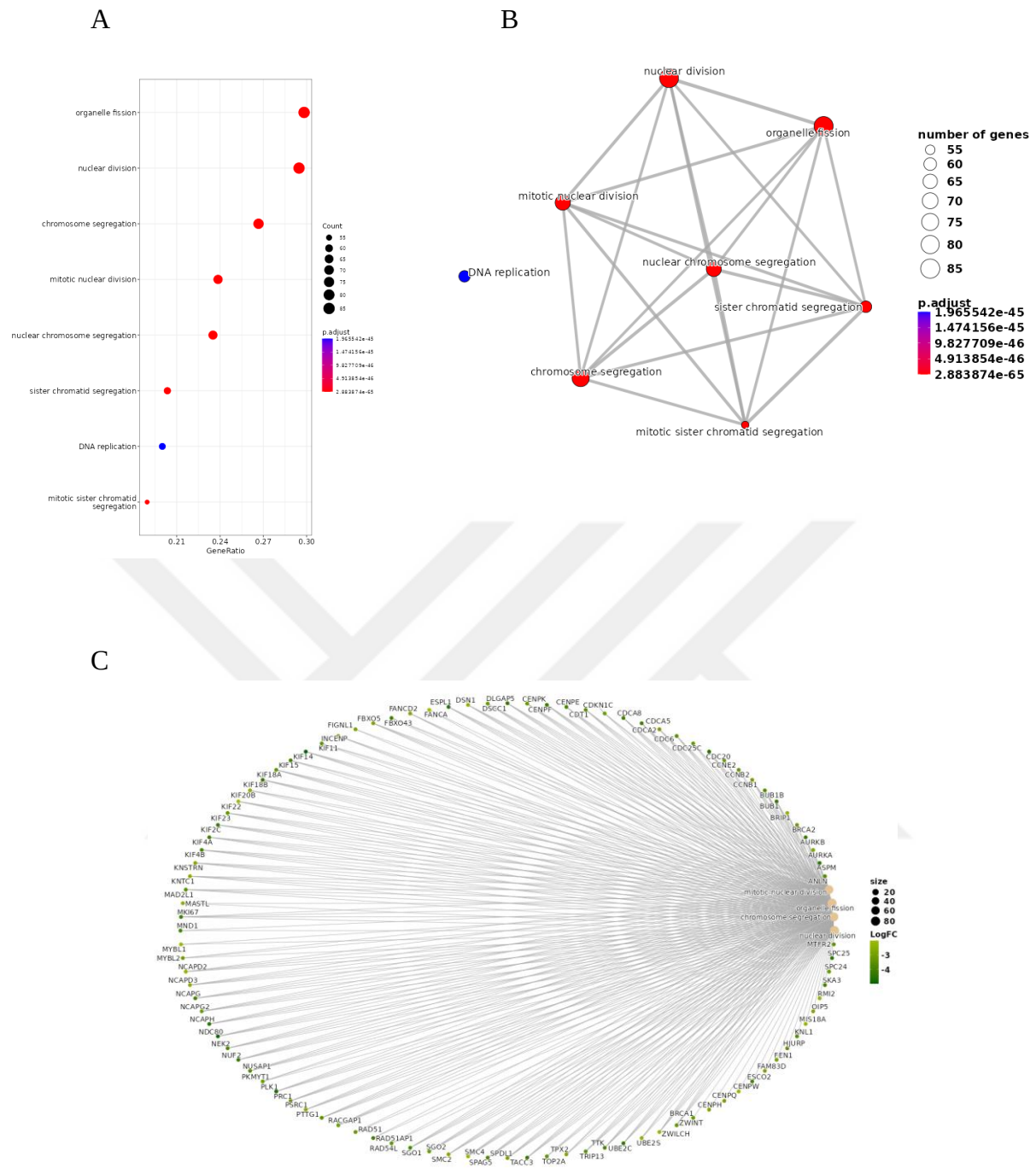


Figure 4.10 **A** Dot plot for the enriched terms. **B** Emapplot, shows the enriched terms as a network with terms with overlapping genes, clustered together. **C** Cnetplot, shows the enriched terms and their associated genes, with gene nodes colored based on the user's selection. The figures above indicated that, the most significant enriched terms appeared to be cell cycle pathway related. This is expected since we had chosen the downregulated genes for the *Nutlin* vs *Control* contrast (GSE47042).

Under the same “Functional Profile” tab, the DisGeNET analysis shows a network or heatmap of the enrichment analysis of the gene list over diseases present in the database (Piñero et al., 2019). The Gene-Disease associations in the database are curated from various databases such as ClinGen, UniProt, Orphanet etc. The enrichDGN function (Yu et al., 2015) based on the p-value, FDR cutoffs, plots the gene-disease associations. All these parameters can be changed in the application as shown in Figure 4.11B. The user can also select between the contrasts to show the logFC’s of the genes on the plots (Figure 4.11 A-C). For the chosen contrast and its downregulated genes, “abnormality of chromosome stability” showed up as one of the enriched terms from this database. For better visualization, the number of enriched terms was set to three.

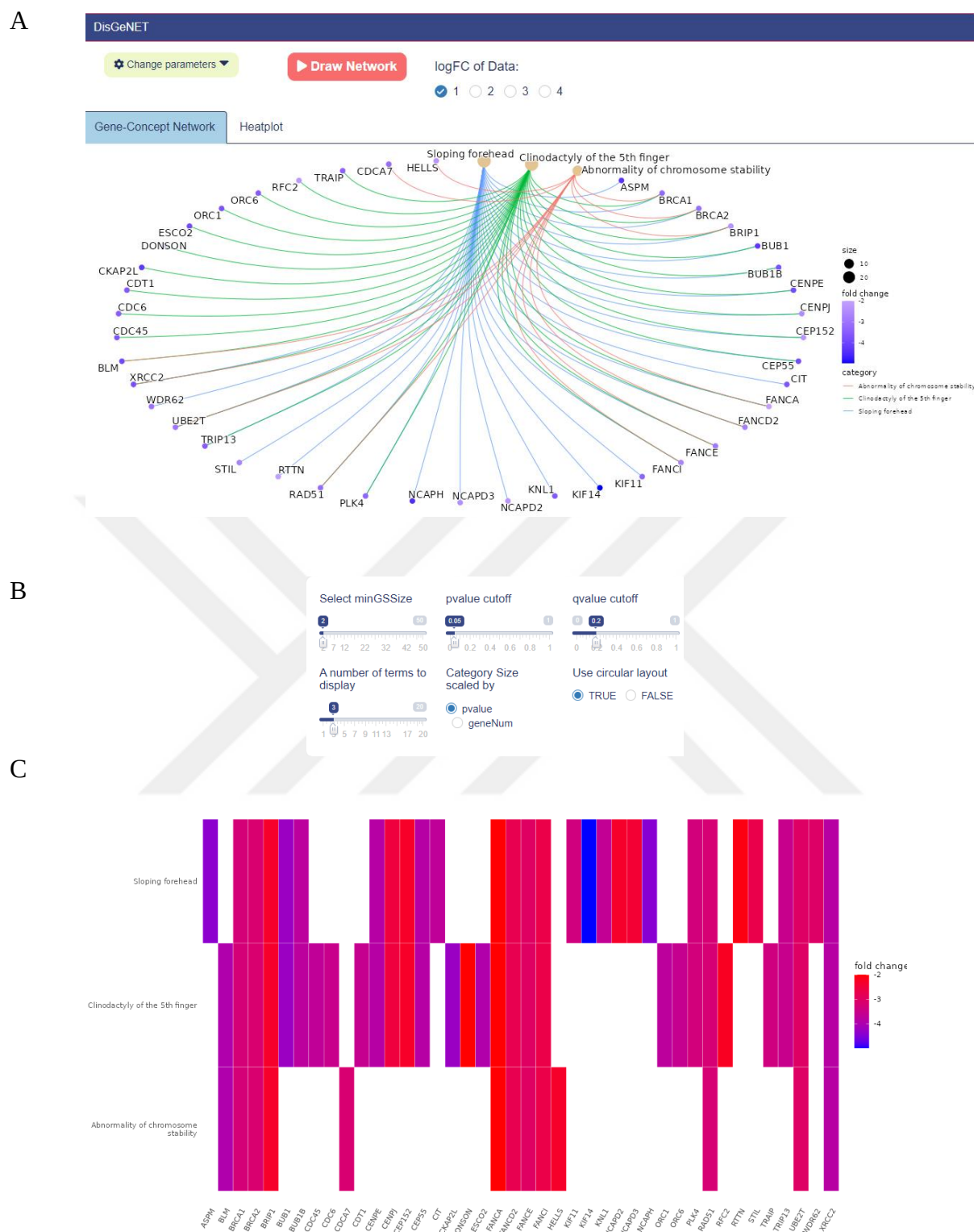


Figure 4.11 **A** The disease gene association based on the chosen genes and logFC of one contrast (contrast 1 chosen here) are shown as a network. **B** This shows the parameters that can be changed by the user when plotting disease-gene associations in the DisGeNET plots. **C** The heatmap shows the enriched disease terms on the y-axis and the genes on the x-axis.

4.1.6 DoRothEA – Transcription Factor Enrichment

DoRothEA is a database for transcription factor – target gene interactions (Garcia-Alonso et al., 2019). This database includes manually curated transcription factor- target gene pairs from multiple sources, and has evidence scores ranging from A to E, with A being the highest. Here, the overlap between the chosen gene set and the TF target genes is calculated using the hypeR package in R (Federico and Monti., 2020). The hypeR function takes the Dorothea gene sets and calculates the hypergeometric overlap between the TF-targets and the selected genes.

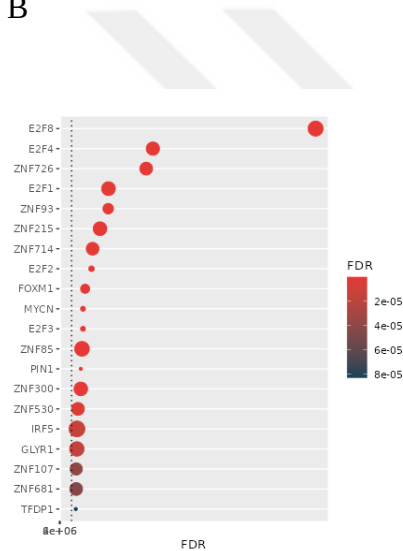
This outputs a table with p-values and the overlap between the gene set and the transcription factor targets which can help the user to understand if a transcription factor is dysregulated in the study. The table is ordered based on the p-values of the TF with respect to the significance of the overlap. By clicking on the TF in the table, the user can visualize a heatmap of the gene targets of the TF and their log fold changes values across the contrasts. The confidence value from A to E (or NA) is also color coded. A network of the transcription factors can also be viewed. This network is based on the interaction and overlap of genes between themselves. The nodes representing the TFs can be selected either based on FDR values of the overlap or the number of TFs desired (which will show the topmost TF in the table, ranked on the basis of p-value).

Accordingly, I have found that the downregulated genes for the nutlin contrast showed E2F family of transcription factors along with FOXM1, which are all involved in cell cycle progression. Along with these, zing finger domain factors also showed up. The E2F and FOXM1 targets when plotted for all the contrasts, showed no regulation for the other three contrasts.

A

Analysis	Scatter Plot	Heatmap	Functional Profile	DoRothEA	Treatment-Gene Networks	MSigDB	Compare clusters	Correlation
Table	Network							
Column visibility	Copy	Excel	PDF	Show 10	entries			
TF	pval	fdr	signature	geneset	overlap	background	hits	Search:
E2F8	1.4e-167	1.9e-164	361	740	182	20295	E2F1,ANLN,ARHGAP11A,ARHGAP11B,ASF1B,ASPM,ATAD2,AURKA,AURKB,BIRC5,BLM,BRCA1,BRCA2,BRIP1,BUB1,BUB1B,C5orf34,CCDC15,CCDC150,CCNA2,CCNB1,CCNB2,CCNF,CD	
E2F4	6e-59	3.999999999999999e-56	361	237	65	20295	ANLN,ASF1B,AURKB,BARD1,BIRC5,BRCA1,CCNA2,CCNB1,CCDC25A,CCDC25C,CCDC45,CCDC6,CCDC7,CDK1,CDKN2C,CDT1,CENPE,DHFR,DUT,E2F1,E2F2,E2F8,FBXO5,FEN1,FO	
ZNF726	2.5e-54	1.1e-51	361	179	56	20295	ARHGAP11B,ATAD5,BRCA2,BRIP1,BUB1,C5orf34,CCDC15,CCDC150,CCNB2,CCNF,CCDC20,CENPE,CENPK,CHAF1A,CHTF18,DSN1,ESCO2,FANCA,FANCD2,FBXO43,FOXO1,GN	
E2F1	4.2e-29	1.4e-26	361	324	47	20295	ASF1B,ATAD2,AURKA,BIRC5,BRCA1,CCNA2,CCNE2,CCDC25A,CCDC45,CCDC6,CCDC7,CDK1,CDKN1C,CDKN2C,CDT1,DHFR,DTL,FANCA,FANCD2,FBXO43,FBXO5,FOXA2,FOXO1,GN	
ZNF93	8e-29	2.1e-26	361	58	25	20295	ARHGAP11A,ARHGAP11B,ATAD5,BARD1,C4orf46,CCDC45,CDK1,CENPK,CLSPN,DTL,ESCO2,FANCI,FBXO5,FOXO1,GIN51,HELLS,KIF15,KNL1,KNTC1,MAD2L1,MCM10,MMS22L1	
ZNF215	2.7e-23	6e-21	361	303	40	20295	ARHGAP11B,AURKA,BLM,BRCA2,CCDC45,CCDC6,CENPF,CENPK,CIP2A,DDIAS,DTL,E2F2,EXO1,FANCA,FANCI,GIN52,GIN54,HELLS,KIF11,KNTC1,LCTL,MND1,NDC80,NUF2,PARF	
ZNF714	2.7e-18	5.2e-16	361	168	27	20295	ARHGAP11B,ATAD5,BRCA2,BRIP1,BUB1B,CENPE,CIP2A,FANCD2,GIN54,HELLS,KIF11,KIF20B,KNL1,KNTC1,MCM2,MMS22L1,NCAPG2,PLK4,POLQ,PRR11,RAD51AP1,RFC3,SKA	
E2F2	1.7e-17	2.8e-15	361	14	11	20295	ASF1B,CCDC6,CCDC7,CDT1,E2F1,MCM10,MCM5,MCM6,MCM7,PLK1,PLK4	
FOXO1	3.2e-13	4.8e-11	361	34	12	20295	AURKB,BIRC5,BRCA2,CCNB1,CCDC25A,CCDC25C,CCDC6,CDK1,CENPA,CENPF,CKS1B,PLK1	

B



C

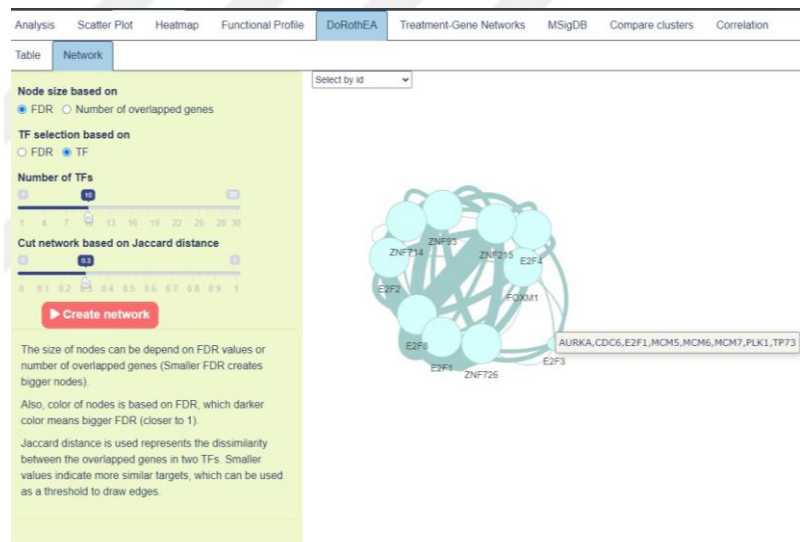


Figure 4.12 **A** Enriched Transcription factor terms based on the Dorothea database. **B** Shows the FDR value for each of the TF enriched with the geneset. The size shows the relative size of genes and the overlap with that TF. **C** Network of the TF can be viewed based on either TF number or the FDR. The edges show the number of overlapping genes between the TF and the node size shows the number of gene targets for that particular TF.

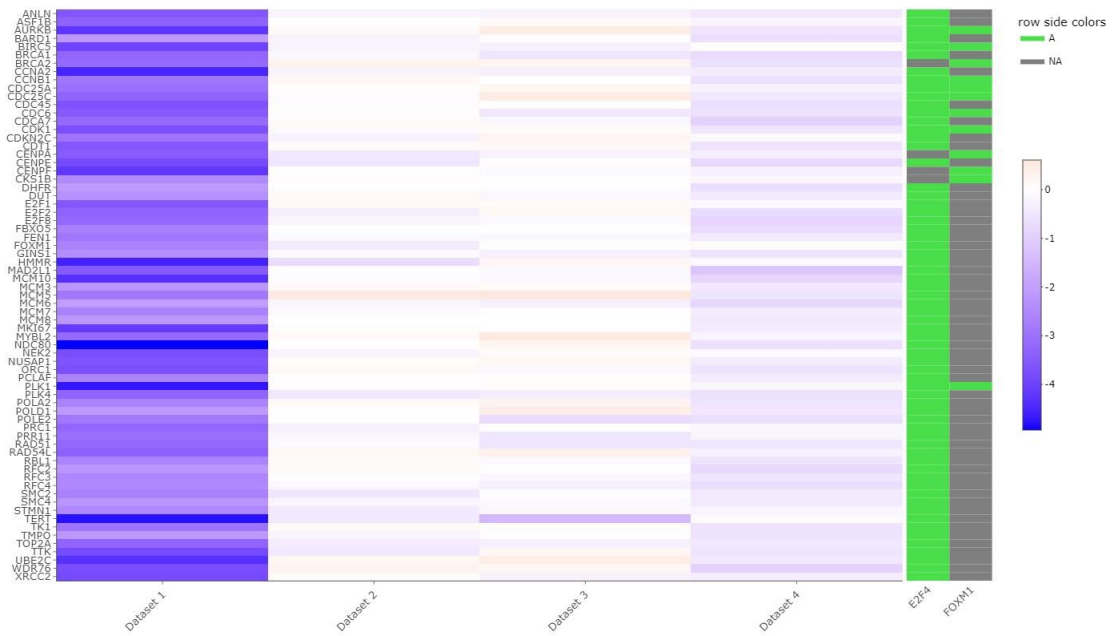


Figure 4.13 Selecting a particular TF, results in a heatmap of the gene targets of the particular TF. In the annotation part of the Heatmap, values show the confidence scores classifying regulons based on their quality (from A to E or NA). To better estimate TF activities, it is recommended to select regulons from the confidence levels A, B and C. The datasets refer to: Dataset 1, GSE47042; Dataset 2, GSE173364 (R175H contrast); Dataset 3, GSE173364 (R273H contrast) and Dataset 4, GSE68248.

Transcription factor networks based on filters with FDR values showed an interconnected set of TFs that included E2F4, E2F8, E2F1 with FOXM1 along with zinc finger domain factors (Figure 4.12 C). The E2F4 targets particularly showed a slight downregulation in the dataset 4 (*siTP53* vs *siControl*) compared to the mutant datasets 3 and 4 (Figure 4.13). This was interesting, but could be due the differences in cell lines, or due to the reduced expression of *TP53*, not fully diminished.

4.1.7 Treatment-Gene Networks

The TP53-META application also allows visualization of the genes and contrasts as a bi-partite network. The edges in the network represent the connection between the contrast/treatment and the gene. The size of the edge represents the logFC value of that particular gene and color represents if it is either upregulated (red) or downregulated (green). If the P-value is not significant (greater than 0.05), then edge does appear between the contrast and the gene. The

genes and the contrasts are represented as nodes and the color of the node for each gene represents the significance in the contrasts. The colors of the network nodes are relative; darker grey color means the gene is significant in a higher number of contrasts while white color would suggest it is significant in the least number of contrasts. The networks use the visNetwork framework in R and are interactive for the user to observe the trends and patterns and to identify key genes which could have a probable synergistic effect. Due to the server capacity and processing limits, there is a cutoff of 500 genes that can be plotted as a network.

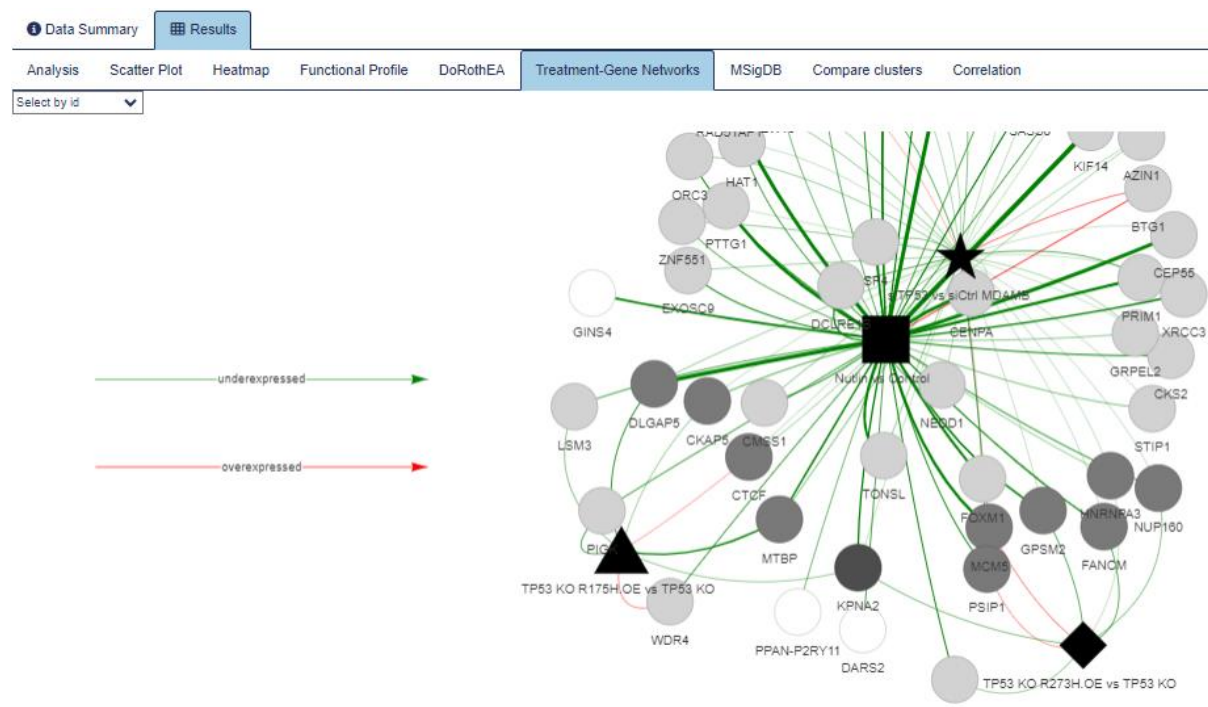


Figure 4.14 Bi-partite network of the four dataset contrasts. Here only 49 genes were used for analysis instead of the whole geneset of the MMB-FOXM1 targets as used previously in Figure 4.8. The edges show the logFC and the contrasts are represented with different shapes, i.e., triangle, square diamond, or star while the circles represent the genes.

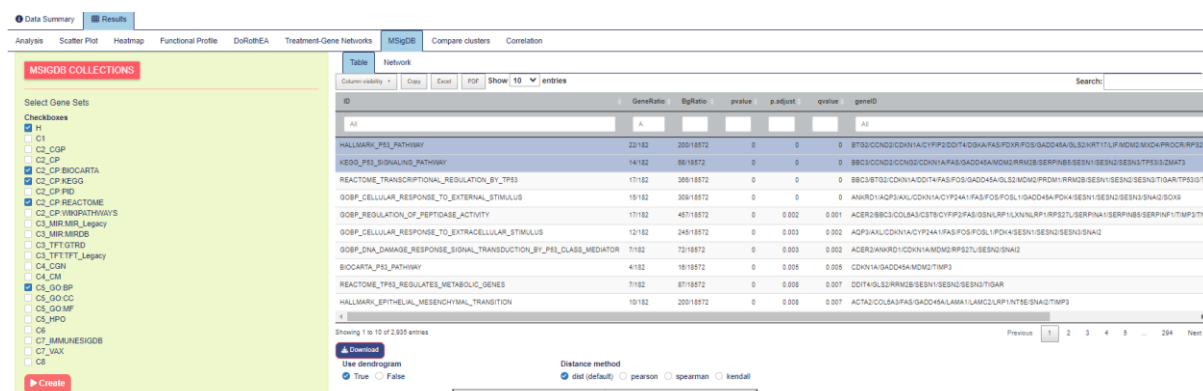
In this particular example (Figure 4.14), I have chosen manually genes with a low p-value for the first contrast. For these genes, the p-value cutoff is also satisfied in other contrasts and hence an edge is seen connecting them to more than one contrast. In this example, the square represented the *Nutlin vs Control* (GSE47042) contrast, the triangle represented the *R175H Overexpression vs TP53 KO* (GSE173364); diamond represented the *R273H*

Overexpression vs TP53 KO (GSE173364); and star represented the last contrast, *siTP53 vs siCtrl* (GSE68248). The size of the edge showed the absolute magnitude of the logFC. Here the *Nutlin vs Control* had genes which had the lowest logFC amongst all the contrasts as shown by the (large) size and (green) color of the edges.

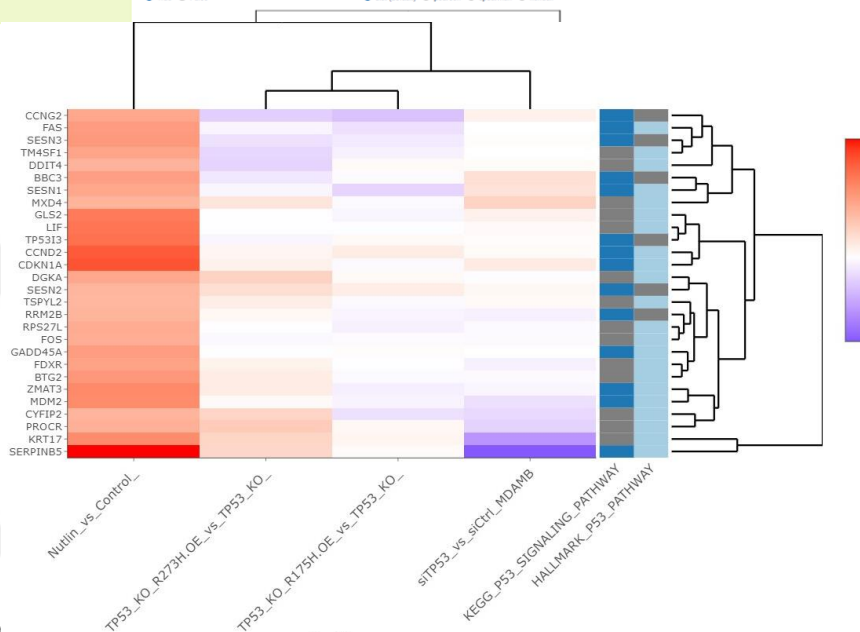
4.1.8 MSigDB and Enrichment Clusters

Another Gene Set Enrichment option in the application is through the MSigDB database (Molecular signatures database; Dolgalev, 2022). The database has nine major collections, H (hallmark gene sets) and then from C1 through C9, varying from cancer gene sets to regulatory target genes etc. Using this database and its R package, *msigdb* (Dolgalev, 2022), enrichment for the chosen genes among various collections can be done. These results can be visualized by a heatmap of the enriched pathways, manually chosen by the user or by visualizing a network of the enriched terms, based on an FDR cutoff.

A



B



C

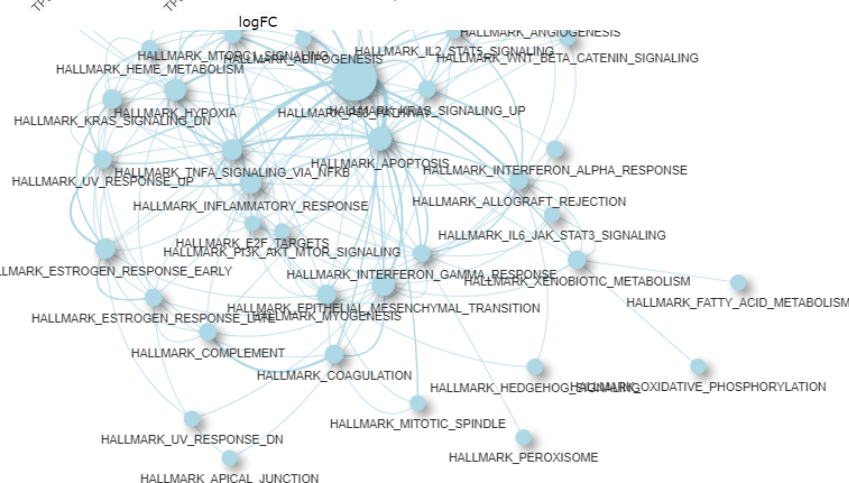


Figure 4.15 **A**, Enrichment based on the terms present in the MSigDB terms. The collection of terms can be chosen from the left. **B**, **C** Similar to the Dorothea TF enrichment heatmap and network, here the user can select the enriched terms together to show its particular targets and visualize a network for the terms. Width of edges are represented by the GeneRatio column. Hovering over the node, shows the overlapping genes between the term and the genes the user selected for the analysis.

Accordingly, when I performed the enrichment based on the significantly upregulated genes ($\log_{2}FC > 2$ and $p\text{-value} < 0.05$) in the *Nutlin vs Control* contrast which filtered 259 genes, the terms enriched were mostly p53 related (Figure 4.15 A). When genes for these terms were plotted as a heatmap for the four contrasts, the *R273H OE vs TP53 KO* contrast showed a small cluster of slightly upregulated genes (Figure 4.15B). However, for all the contrasts other than Nutlin vs Control, no significant regulation was observed for these p53 related pathways. The selected databases from MSigDB, such as KEGG, Reactome, Biocarta and the hallmark enriched terms, showed p53 terms at the top of the table.

4.2 Analyses of in house TP53 RNA-seq data

The analyses of our RNA-seq data were provided as a case study for the TP53-META application. Some analysis related to other gene lists or visualization of more than four contrasts for better understanding of data was performed outside the application. For the analysis, the ComBatseq batch effect corrected files were only used where a comparison was not possible between a batch itself. This was due to the fact that the raw counts were heavily influenced when using ComBatseq and two batches with unequal distribution may not be appropriate for batch correction (Yu et al., 2023).

4.2.1 Mutant cell lines

Using the luminal A type MCF7 cell line, isogenic mutant cell lines were created for the *TP53-R175H*, *TP53-R273H* and *TP53-R248W* by Burcin Aras in our lab (Arıcı, 2022). The summary for these samples has been outlined in Table 3-1 in the material and methods section.

To analyze the R175H and R273H mutations, the analyses were done using the original counts from the first batch of data. For R248W, the second batch counts were used. The Batch effect corrected results for this dataset removed some variability between the mutant cell lines especially R175H and R273H (Figure 4.16).

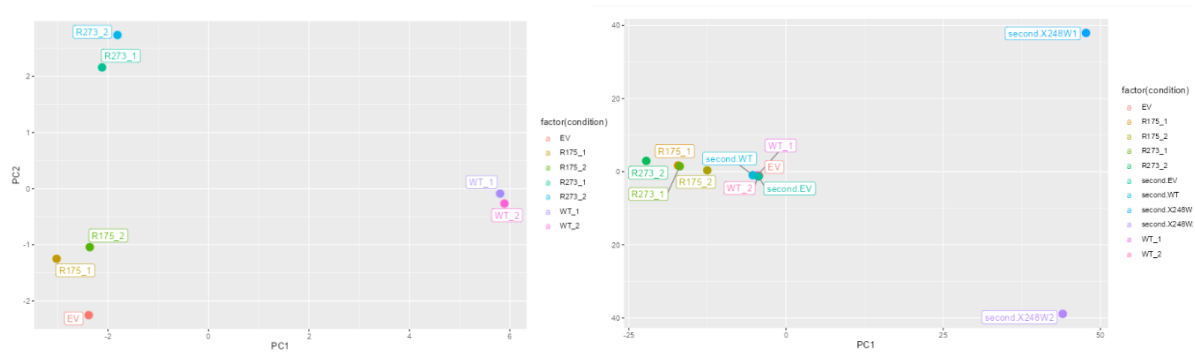


Figure 4.16 **A** PCA for the mutant cell line (only first batch) before batch correction. **B** PCA for the mutant cell line after batch correction (first and second batch combined).

First, I plotted the batch corrected results for the same mutated cell line against each other to see how Combatseq effects the counts. The scatter plots plotted for the DREAM target genes clearly show that the normalization with Combatseq reduces the log Fold Changes (Figure 4.17). This could be due the normalization of the algorithm to minimize the non-biological variations which also leads to overall less variation.

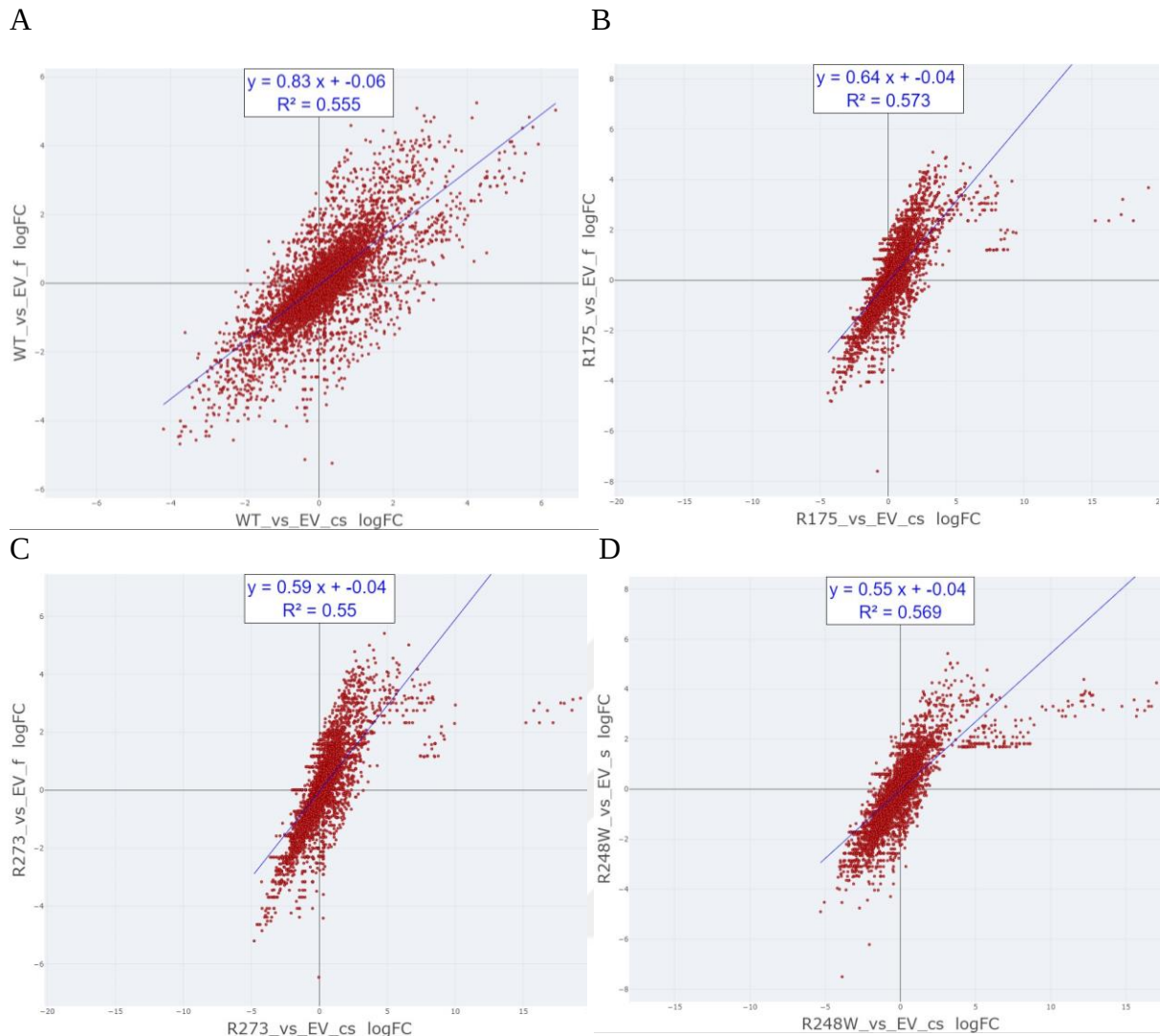


Figure 4.17 A-D, Scatter plot of all the genes for the WT and mutant cell lines (R175H, R273H, R248W) vs EV contrasts, plotted from the separate batches against the Combatseq results for a comparative analysis. The Combatseq contrast is plotted on the x-axis. The slope and the R^2 values can be seen for each scatter plot. These results are based on Limma-voom. Most logFC values are grouped near the x-axis, where the Combatseq results are plotted, while on the same x-axis, some genes have very inflated results for all the contrasts.

The slope/gradient for all of these contrasts are positive. For the WT vs EV contrast, the gradient is 0.83. The Combatseq results being plotted on the x-axis show that this method reports consistently lower values compared to when not corrected for batch effect. Similarly, for all the other contrasts as well, the batch effect corrected values report lower values based on the slope. However, there is some unexpected inflated logFC values. This was most likely because the batch effect correction was done on two unequally distributed batches; and would perform better if the samples had equal distribution in the batches. The R squared value for all

the contrasts is around 0.55, which shows that there is around 45 percent unexplained variance. A higher R squared value would represent smaller differences, which was not the case for these datasets. Hence, I continued my analysis without the batch effect corrected files.

The contrasts for the first analysis included WT, R175H, R273H and R248W as treatment groups against the EV (empty vector).

- 1) *WT vs EV* (Mutant data first batch)
- 2) *R175H vs EV* (Mutant data first batch)
- 3) *R273H vs EV* (Mutant data first batch)
- 4) *R248W vs EV* (Mutant data second batch)

Results WT, R175H and R273H were used from the first batch of mutant data, while for R248W, the second batch counts were used. Genes with significant p-values (less than 0.05) for all contrasts were selected. This filtered the results down to 233 genes. Results for R175H and R273H were quite similar compared to the WT and R248W (Figure 4.15). These results were confirmed from the scatter plots where the R175H and R273H, when plotted against each other showed a linear distribution. The functional enrichment and networks did not show any conclusive results from this filtered geneset. There I further explored the significant genes in the contrasts separately.

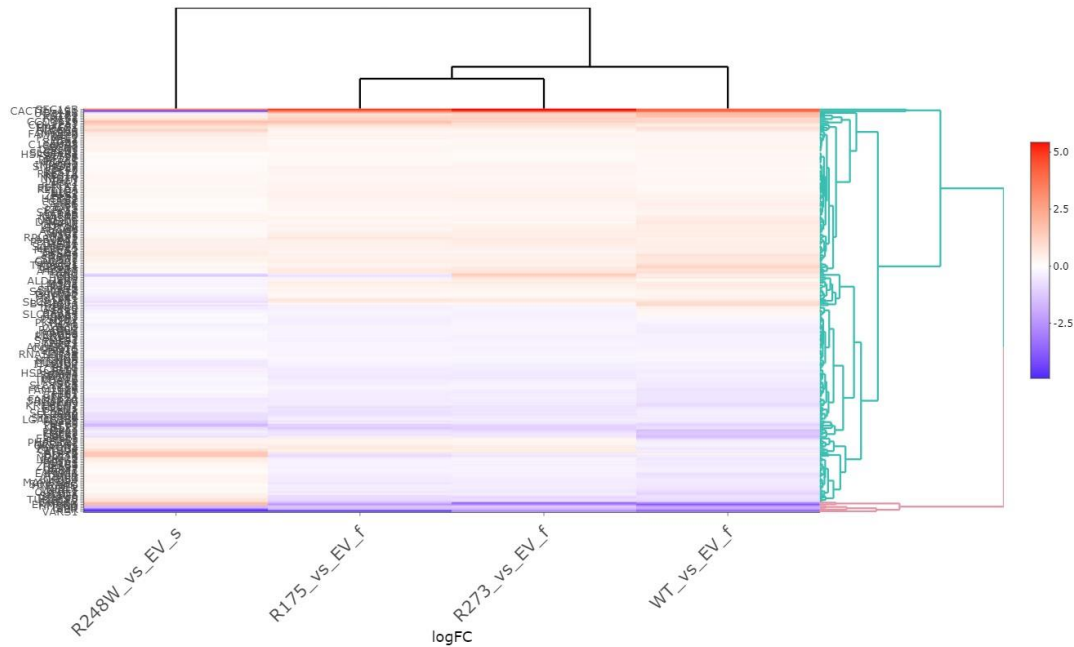


Figure 4.18 Heatmap for the four contrasts, R175H, R273H, R248W, WT vs EV. Here 125 genes were selected. These were selected by filtering only genes that fulfilled a threshold of 0.05 for the p-value across all the contrasts.

I then selected genes with an FDR cutoff of 0.2 in R175H mutant cell line. This FDR cutoff was selected to filter down further the genes, since I get around 500 genes when a p-value cutoff of 0.01 is selected. Only choosing 0.01 for p-value would result in many false positive genes, hence I filter down with the FDR cutoff. We get 111 genes with these cutoffs for the R175H mutant cell line (when compared to EV).

From these 111 genes, one notable gene was *EGR1*, a transcription factor. Downregulation of *EGR1* has been associated with various cancers and can influence tumor suppression pathways (S. K. Saha et al., 2021). Other genes included *TGFB3* and *FGFR3*, both of which are growth factors involved in cell growth, and differentiation. Enrichment analyses based on the DisGeNET database highlighted “early stage breast carcinoma” which could be indicative of progression of cancer based on the R175H mutation (Figure 4.19 A). The GSEA results indicated the regulation of growth factor pathways in addition to some neuronal pathways (Figure 4.19 B). *FOS* gene was specifically downregulated in R175H mutant MCF7

while it was relatively higher in R273H and lower in R248W; *TP53* expression was also higher in R175H and R248W (Figure 4.19C).

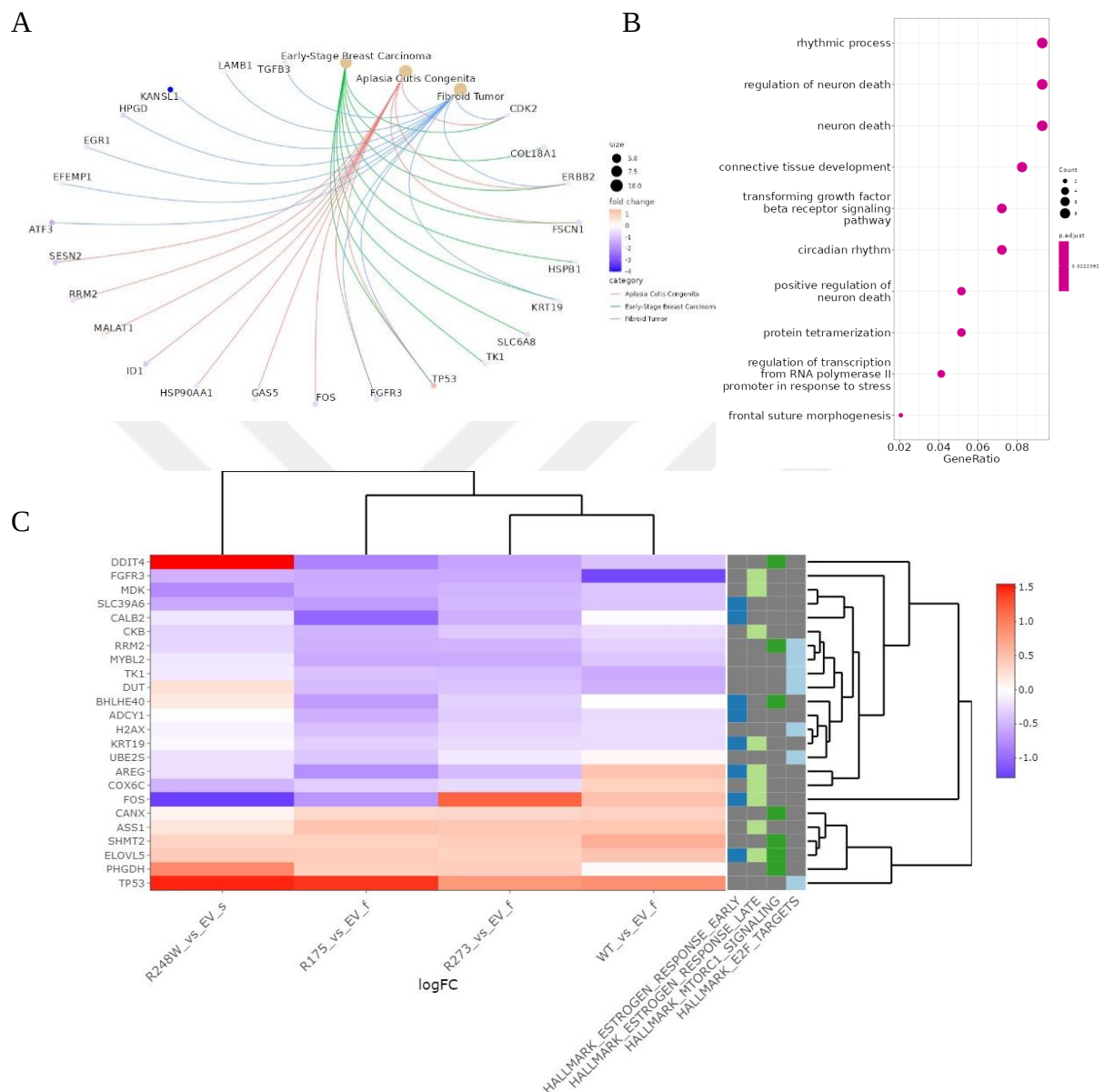


Figure 4.19 **A** Enrichment analysis based on the DisGeNET database. Fibroid tumor and Early stage carcinoma are two of the terms enriched for this R175H mutant geneset. **B** Enrichment based on Gene Ontology terms show some neuronal pathways along with the TGFβ signaling pathway. **C** Heatmap for the chosen enriched terms from the MSigDB tab. The R175H contrast shows a higher downregulation for these target genes compared to the other mutations.

Then for these 111 genes, the hallmark terms from the MSigDB database showed estrogen response early terms mostly downregulated and estrogen response late terms upregulated; and E2F and MTORC1 targets, even though the overlap between the 111 genes and the genes present in the terms was very low. E2F and MTORC1 are both central regulators of cell proliferation and are also implicated in cancer. The R175H mutation may disrupt cell cycle pathways and apoptotic processes which may lead to dysregulation of these pathways and transcription factors.

Using the same cutoff as the R175H cell line, the R273H contrast was filtered for differentially expressed genes. 126 genes came up with an FDR cutoff of 0.2. The clustering in the heatmap showed R175H and WT to be clustered together for this geneset, which could indicate a higher deleterious effect of the R273H point mutation (Figure 4.20 A). Enrichment analysis most frequently showed, response to steroid, epithelial and epidermal cell differentiation terms. Even though most of the patterns for R175H and R273H were linear and similar, one striking difference was in the regulation of *EGR1* and *FOS*, which appeared to be downregulated in *R175H-EV* contrast but were upregulated in the *R273H-EV* contrast. These confirmation of mutant cell lines allowed us to continue with doxorubicin and si*CHRNA5* treatments to see if *TP53* expression can be restored.

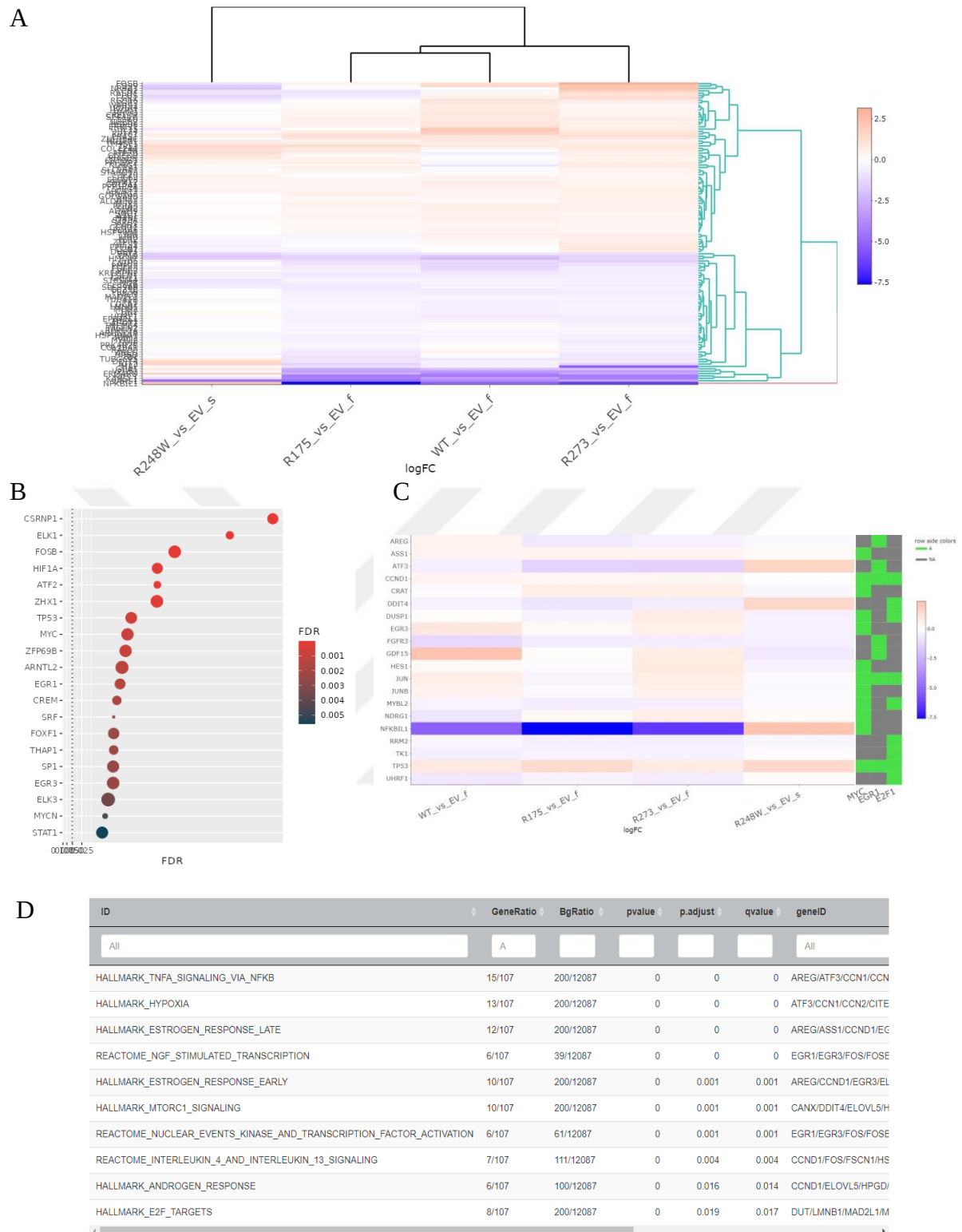
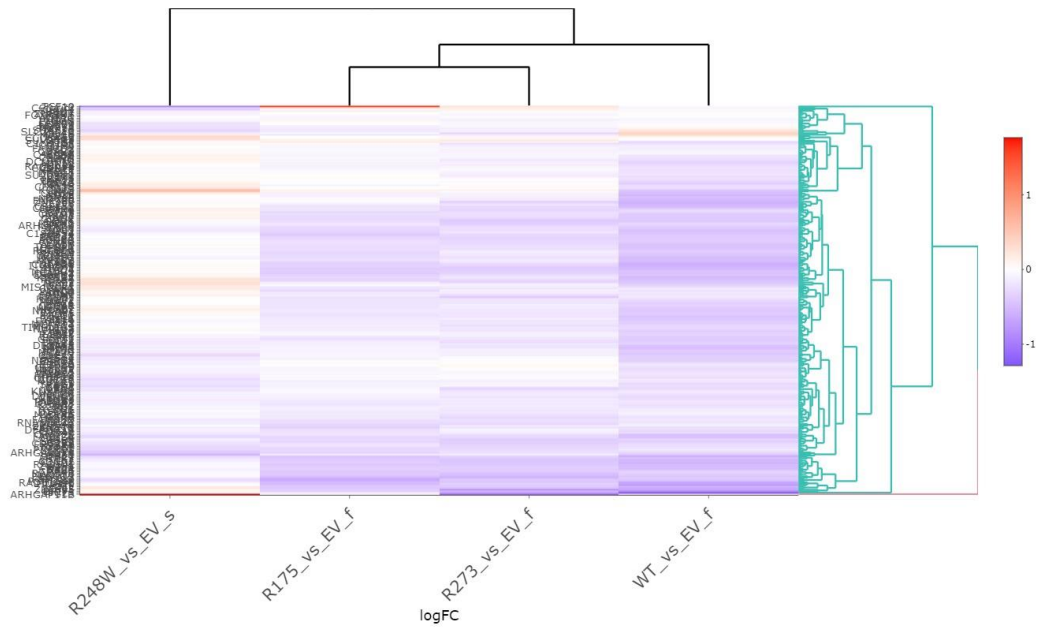


Figure 4.20 **A** Heatmap for all the 126 genes filtered for the *R273H* vs *EV* contrast. Column-wise clustering shows that *R273H* is separated from *R175H* and *WT*. **B** Transcription factors enriched for this gene set. **C** Heatmap of MYC, EGR1 and E2F1 TFs. MYC targets seem to be slightly upregulated compared to the other contrasts. **D** Enrichment of hallmark terms based on the MSigDB database. NFKB, NGF and Estrogen related pathways show up. These terms are very similar to the ones by the *R175H* contrast.

For the *R248W* vs *EV* contrast, 2443 genes passed the FDR filter of 0.2 as used for other contrasts. However, for this contrast no significant pathways related to p53 or Cancer were enriched. Due to these results we decided to disregard this cell line from our treatments with doxorubicin and si*CHRNA5*.

In particular, when filtered for DREAM targets, LIN37 dependent targets and cell regulatory genes, with a filter of 0.05 for the aggregated p-value (meta p-value), the genes in WT appeared more downregulated when compared to R175H, R273H and R248W, which could indicate the abrogation of p53 functionality or a dysregulation of cell cycle (Figure 4.22 A). The treatment-gene network for LIN37-dependent P53 targets, depicted the genes which were not significantly expressed in either R175H or R273H, in the lower right cluster (Figure 4.22 B). This confirmation, that in WT the p53 was functional causing an expected downregulation of the LIN37 targets, while in the mutants, the downregulation of these targets could be negatively affected at least for some of the targets, led us to further experiment our hypothesis of silencing *CHRNA5* by itself or in combination with doxorubicin to restore functional p53.

A



B

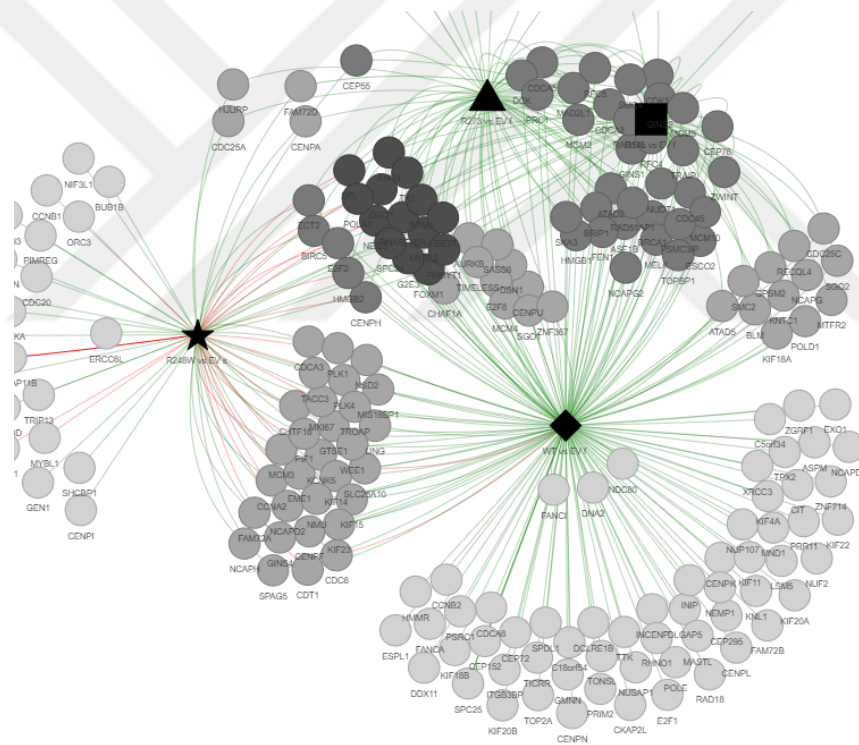


Figure 4.22 **A** Heatmap of the LIN37 dependant P53 targets, plotted across all the contrasts. WT vs EV contrast shows a downregulation of most of the targets, which is normally expected. This is due to the regulation of the LIN37 component of the DREAM complex by P53. However, the mutant cell lines clearly show a dysregulation of these targets. **B** The treatment-gene network confirms the genes with no significance with the other contrasts. The edges are only shown if the p-value is significant in that particular contrast. Here, diamond represents the WT vs EV contrast.

4.2.2 Doxorubicin and si*CHRNA5* Treated Mutant Cell Lines

Following the validation of stable mutant cell lines, the cells were subjected to treatment with either si*CHRNA5* or doxorubicin, or both. The role of silencing of *CHRNA5* as a *TP53* inducer has been discussed previously. There was a total of 15 different samples sent for RNA-sequencing (Dataset 3 in two batches). These were summarized in Table 3-1. The quality control and alignment was described in the methods section.

4.2.2.1 Exploratory data analysis

The raw data were analyzed to confirm the clustering of the samples. EV, R175H and R273H cell lines and their corresponding treatments were sent in one batch, while WT and its treatments were sent in another batch. PCA for both the first batch and the Combatseq normalized batches were plotted (Figure 4.23).

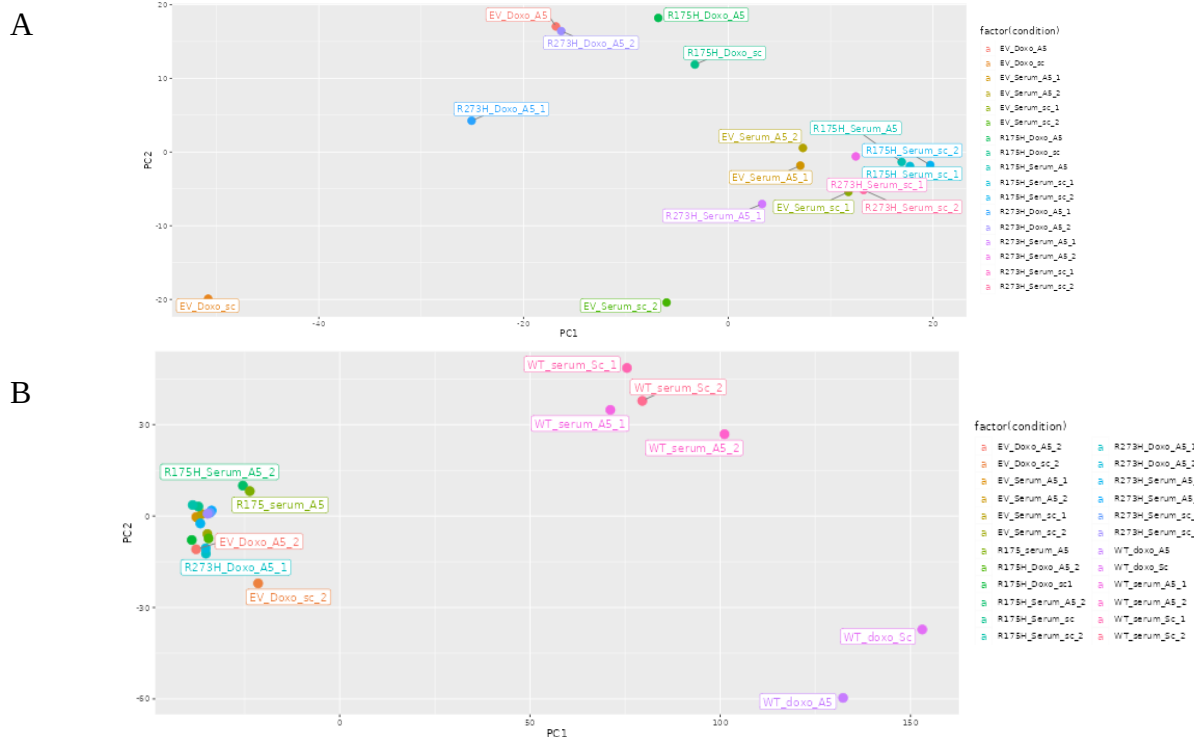


Figure 4.23 **A** PCA of the doxorubicin treated samples before Combatseq. The different samples are mostly far apart from each other with their replicates clustered somewhat together. **B** PCA of the doxorubicin treated samples after Combatseq. Clusters all of the samples from the first batch all together, thereby reducing variability.

For the data analysis of EV, R175H and R273H mutant cell lines, the first batch data was used without correcting for batch effects. To check the effect of siCHRNA5 and doxorubicin in combination and their effect on these mutant cell lines, the combinatorial treatments vs their respective controls were plotted, in addition to single treatments for each cell line. As in the dataset 2; unequal distribution of the batches did not allow for batch correction.

4.2.2.2 EV contrasts

Based on the PCA plot (Figure 4.23 A), it appears that the EV_Doxorubicin_siCHRNA5 sample could have been mislabeled. Since we also did not have replicates for this particular sample, we could not confirm this. On further inspection in IGV (Integrative Genomics Viewer), the expression of the *TP53* gene varied and was not consistent with what would have been expected in an MCF cell line with an empty vector (Appendix B). However, for consistency in the analysis of the cell lines, DEGs for the EV contrasts were explored. The following contrasts were plotted for this;

EV_Doxorubicin vs EV_Control (Doxo First Batch);

EV_siCHRNA5 vs EV_Control (Doxo First Batch);

EV_Doxorubicin_siCHRNA5 vs EV_Control (Doxo First Batch);

Nutlin vs Control (GSE47042).

*The labeling in the application is as follows and the same format for the other cell lines;

~ EV_Control = EV

~ EV_Doxorubicin = EV_D

~ EV_siCHRNA5 = EV_A5

~ EV_Doxorubicin_siCHRNA5 = EV_D_A5

The differentially expressed genes for the EV cell line, showed similarity between the doxorubicin treatment and the combinatorial treatment (Figure 4.24). Since the effect of doxorubicin has been extensively researched upon as a P53 inducer, I mainly focused on the effects of si*CHRNA5* and the combinatorial for this EV cell line and its contrasts.

I used 0.05 as the FDR cutoff for the *siCHRNA5* vs *control*, EV contrast. This filtered a total of 885 genes. These genes when plotted against the combinatorial showed a linear trend with a R squared value of 0.706, which shows that there is around seventy percent similarity in the variance and hence can be dependent on each other (Figure 4.25).



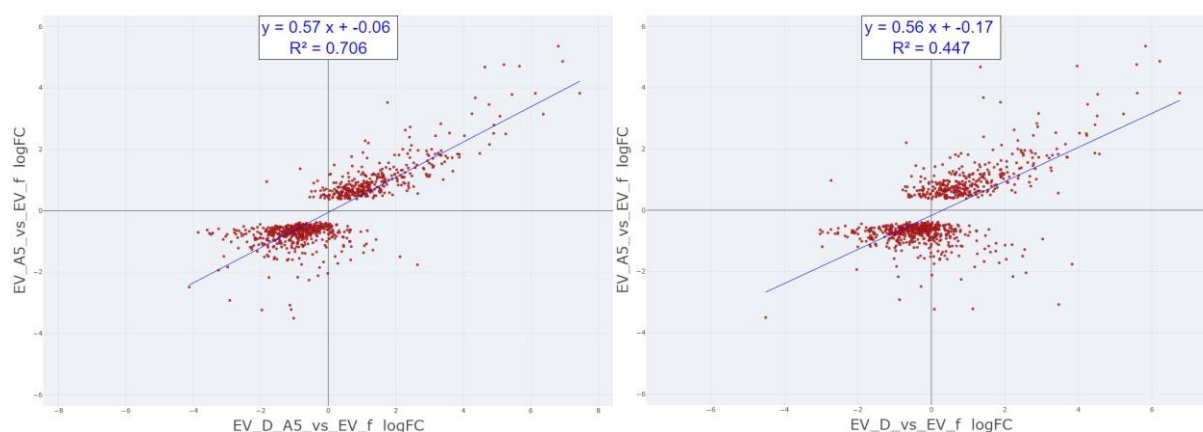
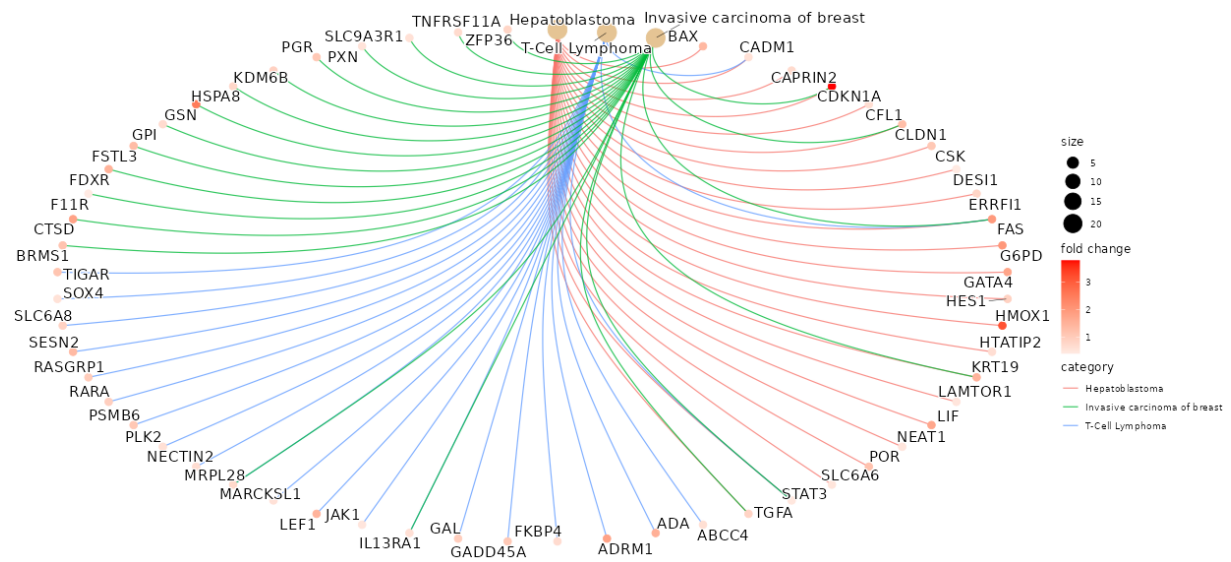


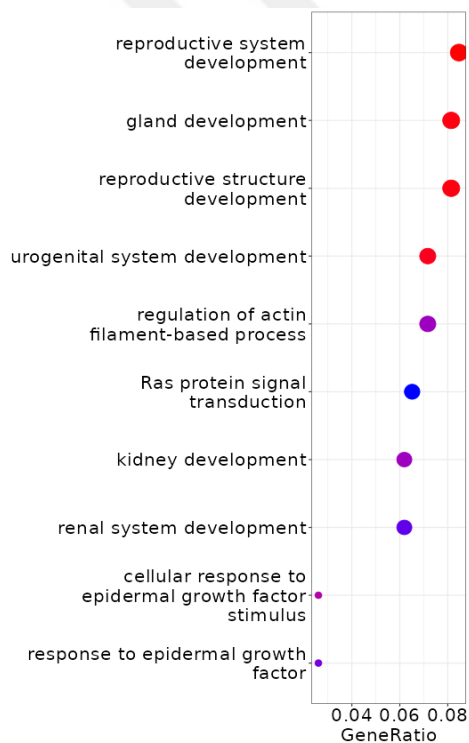
Figure 4.25 Scatter plot for the gene with a significant FDR for the si*CHRNA5* treatment. **A** si*CHRNA5* treated EV is on the y-axis and the combinatorial treatment is on the x-axis. **B** si*CHRNA5* treated EV is on the y-axis and the doxorubicin treatment is on the x-axis. R^2 value with the doxorubicin treatment is comparatively low.

These filtered genes were further filtered based on their direction of regulation. There were 358 genes, which were upregulated based on the logFC cutoff of 0 or above. The functional profile for these genes included “Invasive carcinoma of breast”, “Hepatoblastoma” and “T-cell lymphoma” from the DisGeNET database (Figure 4.26 A). *P21* was significantly upregulated with this si*CHRNA5* treatment, along with *HMOX1* which is an NRF2 target. The role of silencing *CHRNA5*, may also be explored in other carcinomas as a therapeutic strategy.

A



B



C

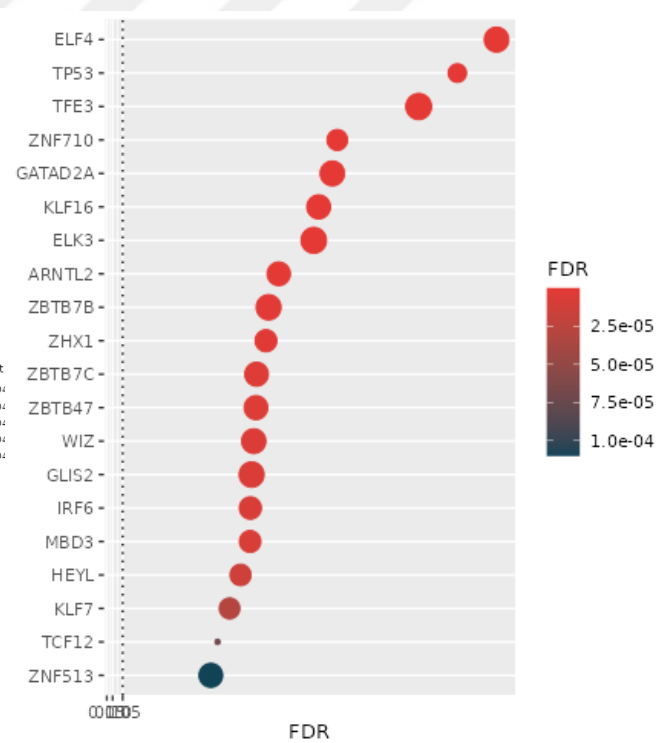


Figure 4.26 **A** Disease-gene interactions from the DisGeNET database. “Invasive carcinoma of breast”, “Hepatoblastoma” and “T-cell lymphoma” showed as the top three enriched terms. **B** GO:BP terms for the upregulated genes. **C** TFE showed *TP53* as one of the significantly enriched TFs.

In the MSigDB database, the enriched terms included, “P53 pathway”, “Estrogen response” and “Apoptosis”, among the top 10 terms. The regulation of the genes across all the contrasts was similar. The upregulation of *P21* and *TP53* TF targets showed a promising

response to the si*CHRNA5* treatment when at least copy of WT *TP53* present. I then filtered the downregulated genes, because we hypothesized that si*CHRNA5*, through regulation of p53 and p21, regulate (or downregulate) various cell regulatory genes which would otherwise be upregulated in breast cancer.

For filtering the downregulated genes, a cutoff of (less than) 0 was used. This filtered 527 genes which fulfilled the previous FDR criteria of 0.05 and this logFC cutoff. Surprisingly, the functional profile for these downregulated genes showed unrelated pathways with very low overlap with the enriched terms. Enriched terms from MSigDB, however, showed E2F targets and G2M checkpoint targets. These genes were also downregulated further in the other contrasts (Figure 4.27).

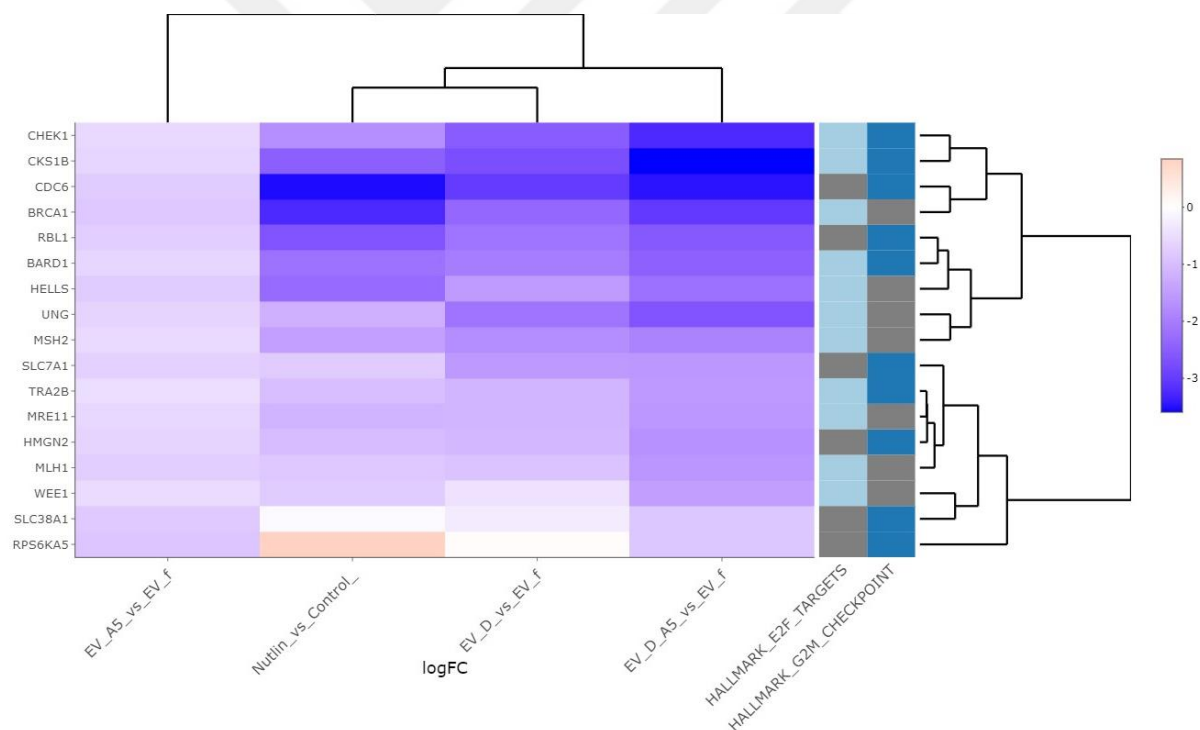


Figure 4.27 MSigDB enriched terms for the downregulated si*CHRNA5* treated downregulated genes.

The effects of doxorubicin in the EV cell line were explored next. An FDR cutoff of 0.05 filtered a total of 3757 genes. Since this geneset was too large to explore, these genes were further filtered with logFC greater than 1, to get upregulated genes. There were 1410 genes which fulfilled this criterion. The GO:BP included various migration terms and “wound healing” as some of the enriched terms. Doxorubicin’s affect on the DNA and cellular replication mechanisms may interfere with normal wound healing process or cell migration and enhance them. This goes in line with the enriched terms from MSigDB which include p53 pathway genes, apoptosis genes and TNFA signaling.

Next I filtered downregulated genes with the same cutoffs of $FDR < 0.05$ and logFC less than -1 leading to 1053 genes. Enrichment of GO terms showed various cell regulatory terms, such as “chromosome segregation”, “nuclear division” and “DNA replication”.

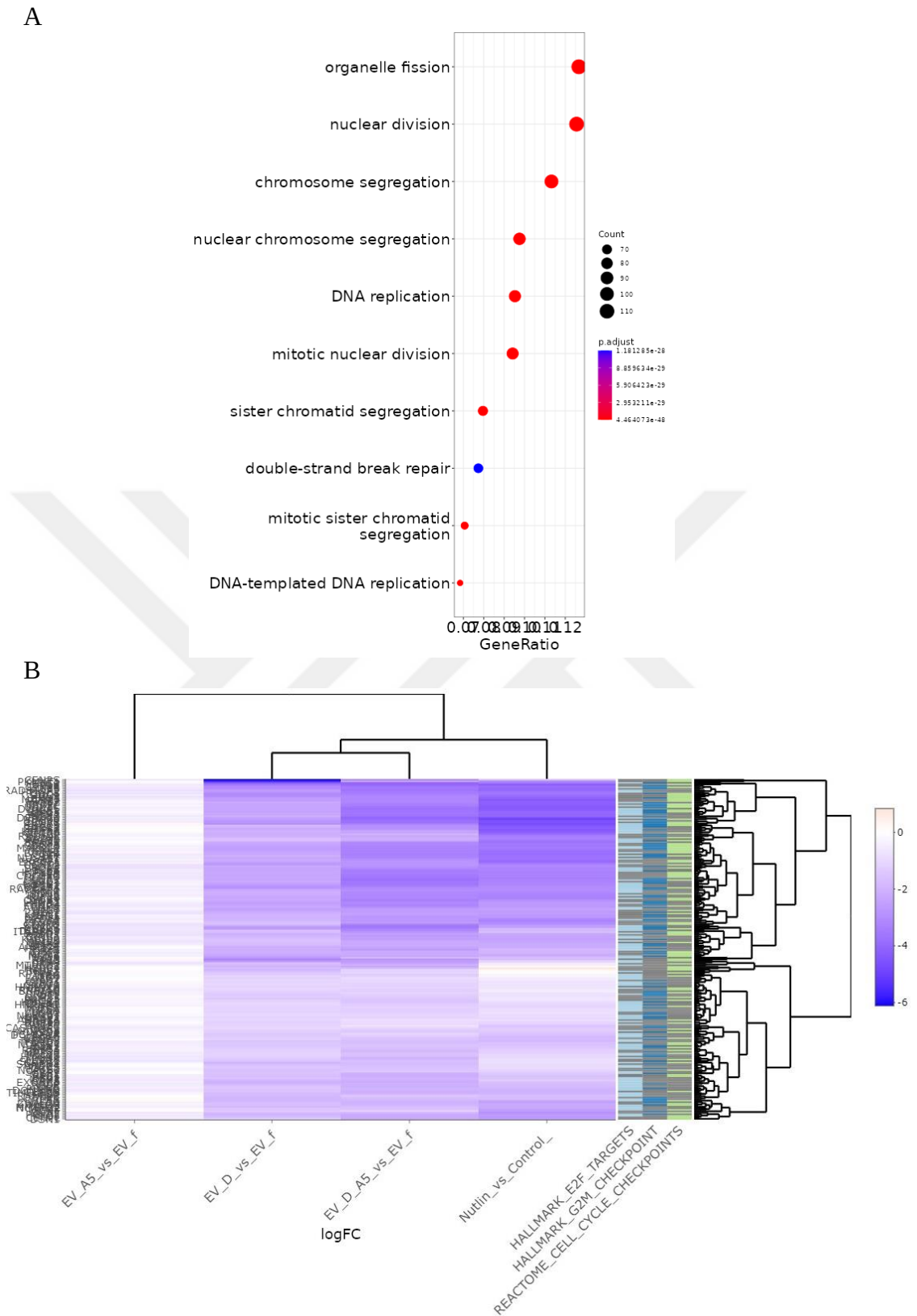


Figure 4.28 **A** GO enrichment based on the downregulated genes of the doxorubicin treated EV cell line. Most enriched terms are related to cell cycle pathways or replication. **B** Heatmap of the “Hallmark E2F targets”, “G2M targets”, and “Reactome cell cycle checkpoint” genes.

Enrichment of terms in Figure 4.28 demonstrated the effect of doxorubicin in regulating various pathways relating to cell division. The MSigDB enriched terms also included cell regulatory targets and E2F targets, all with a significant overlap between the enriched terms' targets and the selected downregulated genes.

The role of combinatorial treatment in the EV cell line was explored in sub section 4.2.2.6. To further explore the role of si*CHRNA5* and doxorubicin, I filtered the LIN37 dependent *TP53* targets across the contrasts with an aggregated p-value (meta p-value) cutoff of 0.01, based on the Bonferroni method. si*CHRNA5* treatment interestingly showed relatively lower downregulation of these targets (Figure 4.29). This may be due to the cell line's tumorigenic nature, and hence provides resistance to the depletion of *CHRNA5* alone or could be an artifact of the vector.

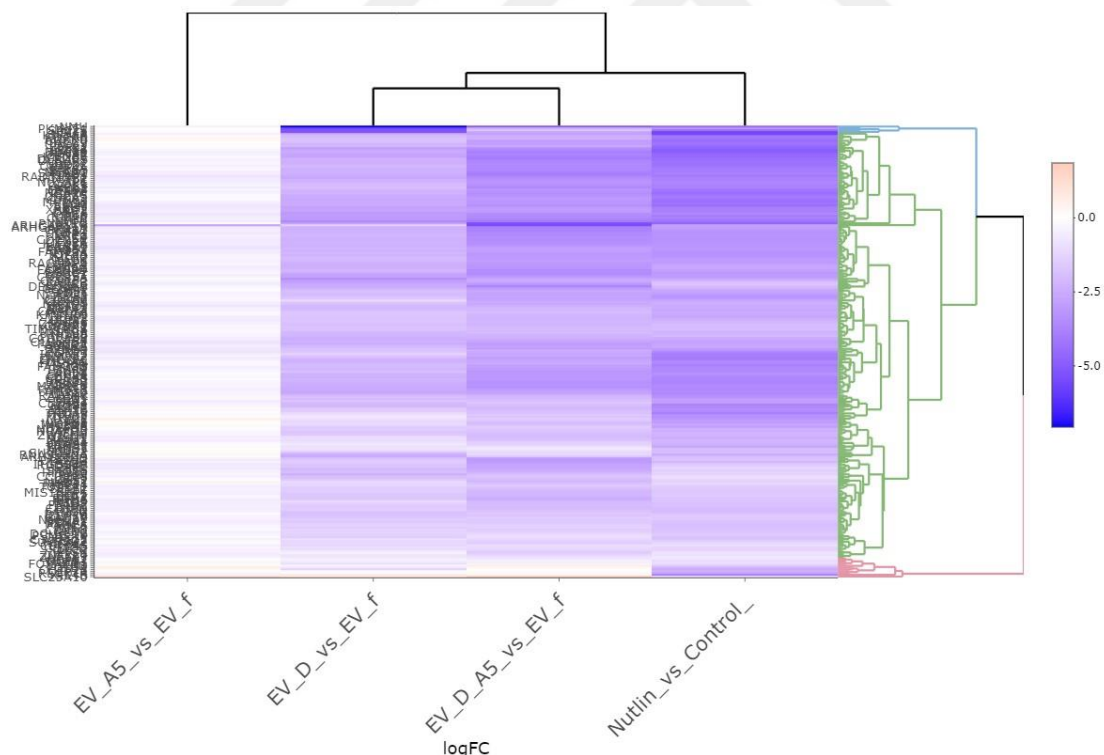


Figure 4.29 Heatmap of LIN37 dependent p53 targets.

Comparison of cell cycle genes among different contrasts showed that doxorubicin had a more similar effect, with respect to the magnitude, to nutlin than compared to si*CHRNA5* (Figure 4.30). This was surprising, since previously with our MCF7 cell line (Data 1, Table 3-1), cell cycle genes were highly downregulated in response to si*CHRNA5*. This variation in results may be due to the MCF7 cell batch that might show differences due to different passages and/or laboratories (Nugoli et al., 2003).

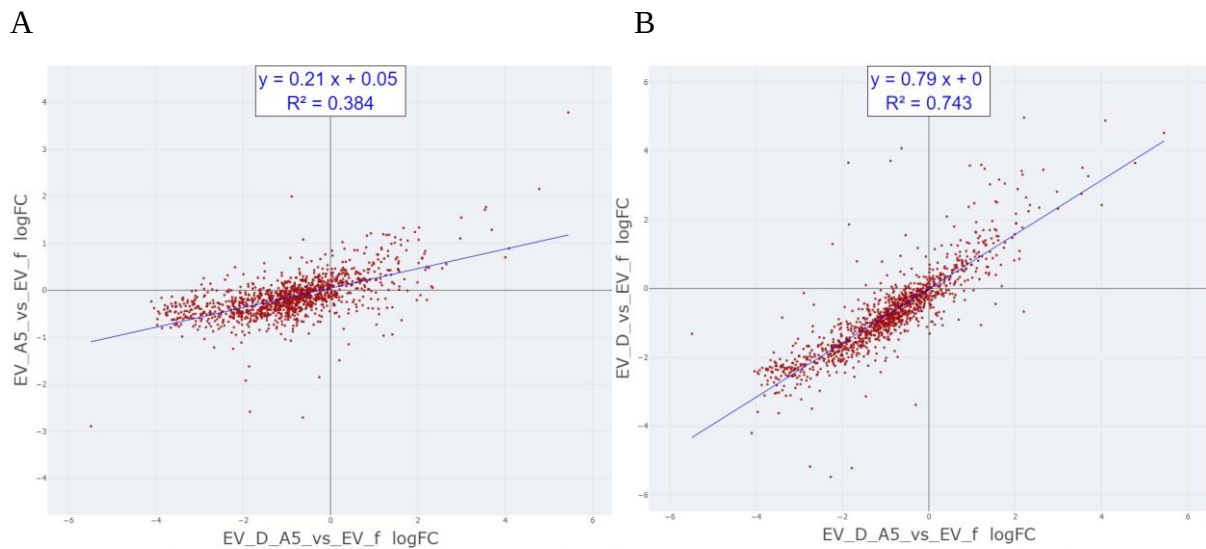


Figure 4.30 **A** Scatter plot of combinatorial against si*CHRNA5* for cell cycle genes. The points were closer to the x-axis representing the lower levels of regulation from the si*CHRNA5* treatment R^2 value here is 0.384 which may indicate less significance and dependence between the contrasts. **B** Scatter plot of combinatorial against doxorubicin for cell cycle genes. The R^2 value is 0.743, which represents the co-dependence of the contrasts.

4.2.2.3 R175H contrasts

The effect of individual treatments and combinatorial in R175H mutant cell line was explored next. The following contrasts were plotted;

R175H_Doxorubicin vs R175H_Control (Doxo First Batch);

R175H_siA5 vs R175H_Control (Doxo First Batch);

R175H_Doxorubicin_siA5 vs R175H_Control (Doxo First Batch);

WT_Doxo vs *WT_control* (Doxo Second Batch).

Using only doxorubicin as the treatment in the R175H mutant cell line, there were 1176 genes which fulfilled a FDR filter 0.01. These differentially expressed genes for the R175H doxorubicin contrast, were plotted against si*CHRNA5* treatment for the same cell line. A very low R^2 was observed which suggests that si*CHRNA5* and doxorubicin target different genes and/or pathways (Figure 4.31 A). However, the combinatorial and doxorubicin genes showed a linear trend with a R^2 value of 0.923.

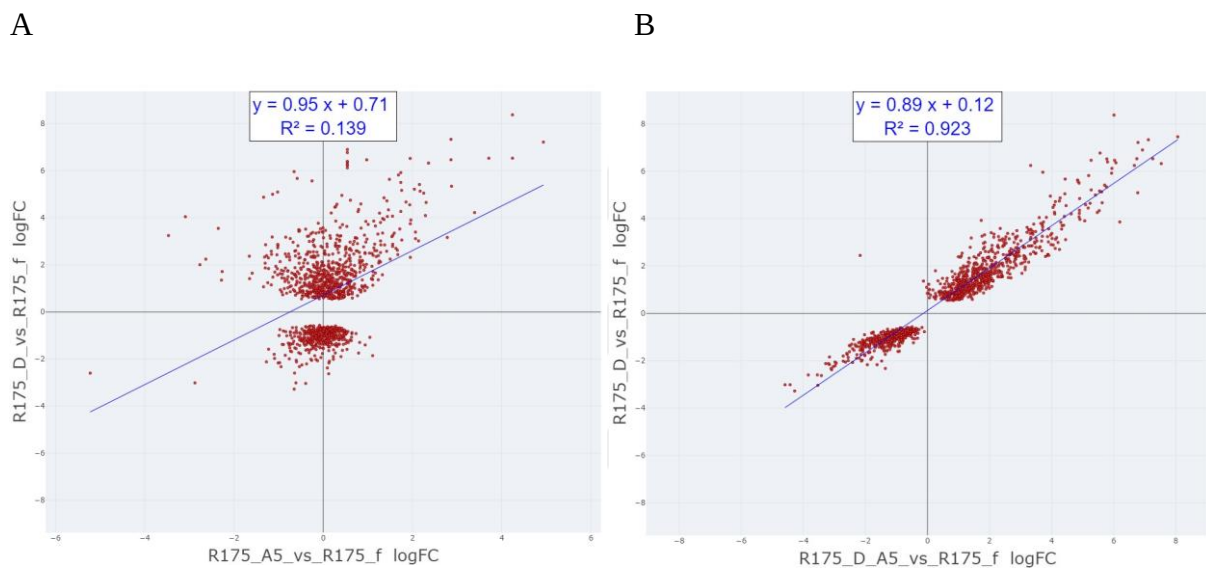


Figure 4.31 **A** Doxorubicin treated R175H contrast is plotted on the y-axis and si*CHRNA5* treatment on the x-axis. These genes are significantly expressed in the doxorubicin contrast and appear to be not mutually dependent. **B** These significantly expressed genes in the doxorubicin treatment are highly correlated with the combinatorial treatment.

There were 563 upregulated genes filtered, based on a FDR cutoff of 0.01 and logFC of 1 and above. Enrichment analysis for the upregulated gene set showed “response to hypoxia” related terms and “signal transduction by p53 class mediator”. The transcription factor enrichment based on Dorothea showed *TP53*, *ESR1* and *JUN* as the top-most enriched TFs (Figure 4.32). The particular targets of *TP53*, *ESR1* and *JUN*, show a similar effect to when the treatments are done in combination. To particularly target the genes in the R175H mutation, it is expected that si*CHRNA5* along with doxorubicin will play a vital role.

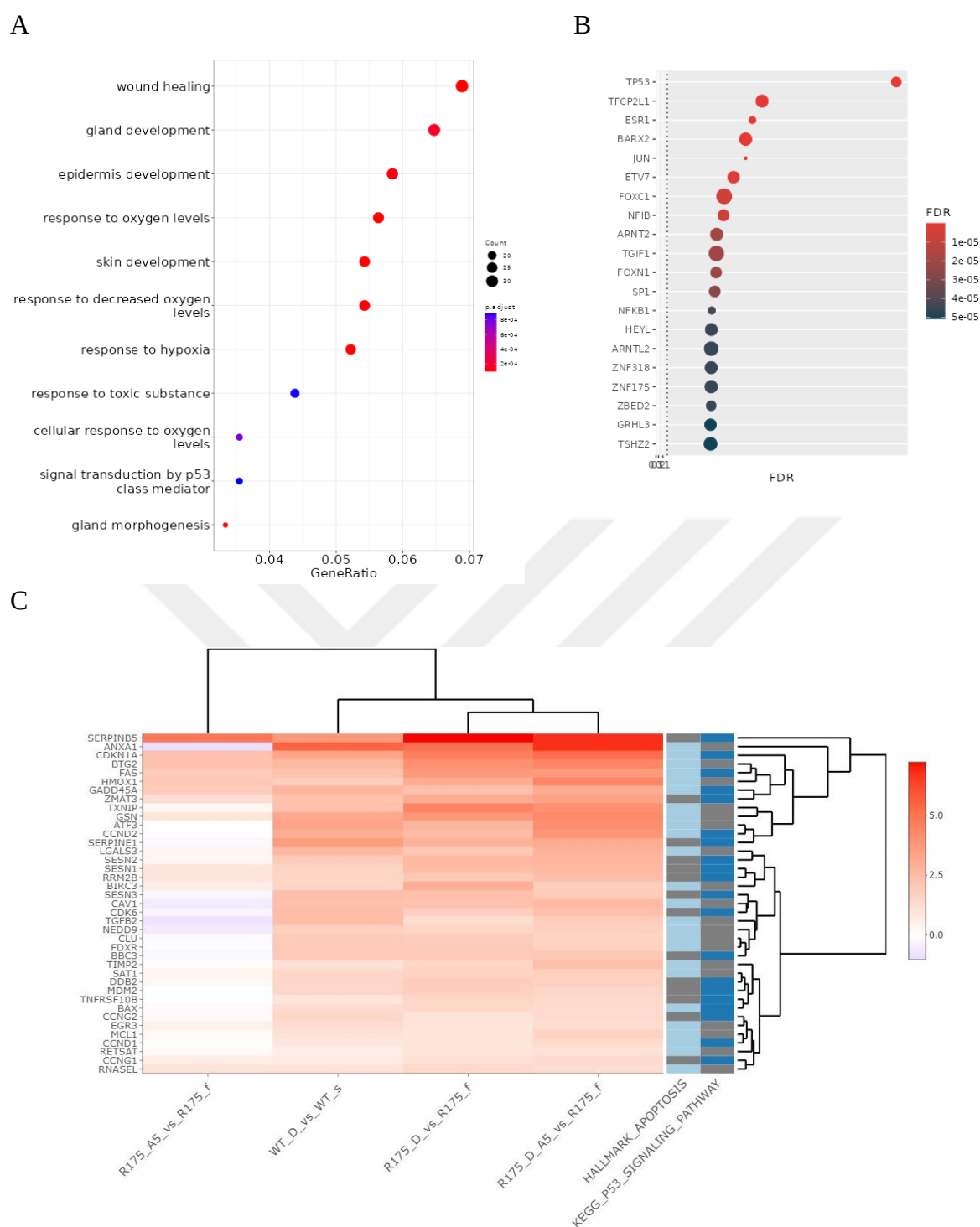


Figure 4.32 **A** shows the GO functional enrichment of the upregulated R175H doxorubicin contrast. Most enriched terms include response to oxygen. **B** Dorothea TFE showed *TP53*, *ESR1* and *JUN* as the top-most enriched TF. All of these TF play important roles in cancer progression. **C** Upregulated genes from the doxorubicin treatment also show apoptosis as an enriched term. Important apoptosis targets upregulated by doxorubicin alone and in combination include, *P21*, *HMOX1*, *BAX*, *MDM2*.

From these crucial genes that doxorubicin is able to upregulate, one important pattern that appears is the upregulation of *P21* and *HMOX1* in an additive/synergistic manner in the combinatorial treatment. Both of these genes are also upregulated with si*CHRNA5* treatment.

With the same cutoffs ($\log_{2}FC < -1$ and $FDR < 0.01$), there were 230 genes downregulated in the doxorubicin treated contrast for the R175H cell line. These genes include *MYC*, *MCM2*, *E2F8* among some of the cell regulatory genes. *MCM2* and *E2F8* genes were slightly less regulated compared to the WT contrast with the doxorubicin treatment. However, *MYC* appeared to be slightly more downregulated in the R175H mutant upon doxorubicin treatment compared to the WT with the same treatment. The GO enrichment showed terms very similar to how doxorubicin responded in the EV cell line. The enriched terms included various replication terms and also “double strand break repair” (Figure 4.33). The TFE showed E2F family of transcription factors in addition to *FOXM1* and *MYCN*, which are proliferation, associated TF and a proto-onco gene respectively.

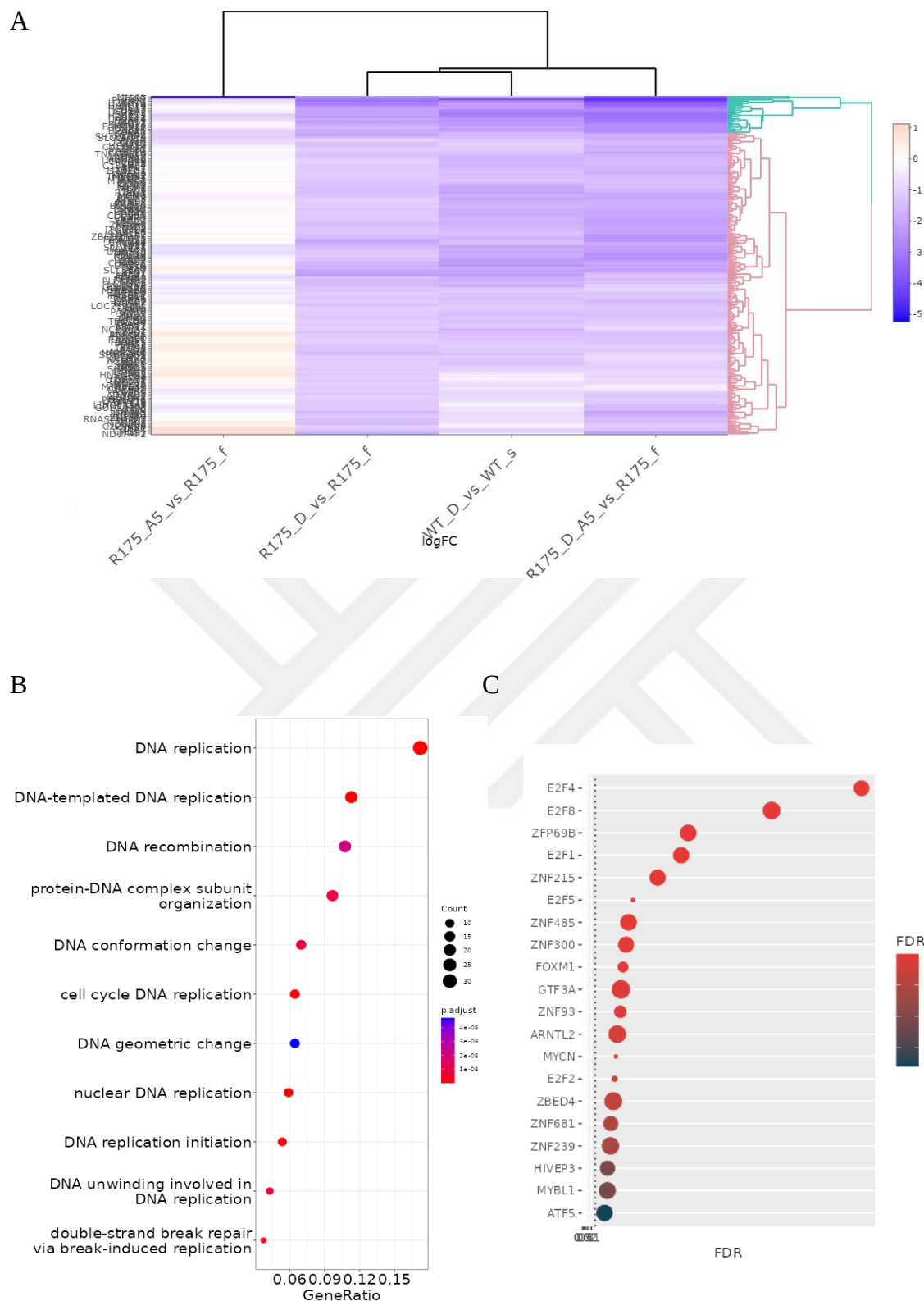


Figure 4.33 **A** Heatmap for the 230 downregulated by doxorubicin in the R175H mutant cell line. Regulation of these genes in the si*CHRNA5* contrast varies significantly compared to the doxorubicin and combinatorial. **B** functional enrichment based on the GO database. Various replication terms are enriched. **C** TFE shows the E2F family of transcription factors.

The transcriptomic effects of si*CHRNA5* on the R175H cell line was explored next. Here the DEGs were selected based on the p-value. This is because the FDR proved to be very stringent here. For the si*CHRNA5* treated R175H mutant with a p-value cutoff of 0.01, there were 929 genes differentially expressed. The scatter plot shows all the differentially regulated genes for the si*CHRNA5* treatment plotted against the combinatorial and the doxorubicin treatment.

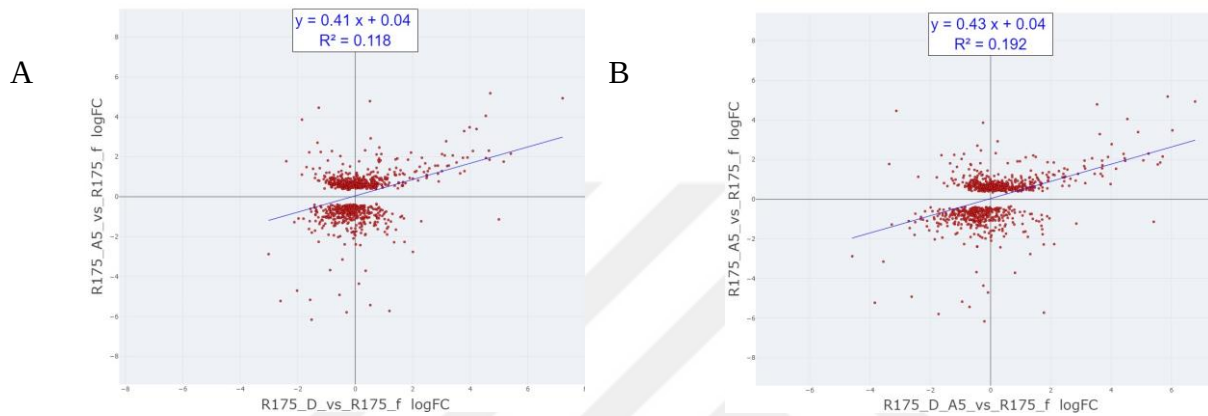


Figure 4.34 Scatter plot of the 931 DEGs for the si*CHRNA5* treatment in R175H mutant cell line. **A** si*CHRNA5* is plotted on the y-axis and doxorubicin on the x-axis. **B** si*CHRNA5* is plotted on the y-axis and the combinatorial on the x-axis.

The doxorubicin DEGs appeared to have a higher influence on the combinatorial genes compared to the si*CHRNA5* treatment, for the R175H cell line. The effect of combinatorial was further explored in section 4.2.2.6. The 929 genes were further filtered into 544 upregulated ($\logFC > 0$) and 385 downregulated ($\logFC < 0$) genes. The upregulated genes for this contrast showed cellular and aerobic respiration as enriched terms, along with various mitochondrial related terms (not shown in the dotplot). “Translational initiation” and “ncRNA processing” are both important in regulating gene expression at the transcriptional and post transcriptional levels.

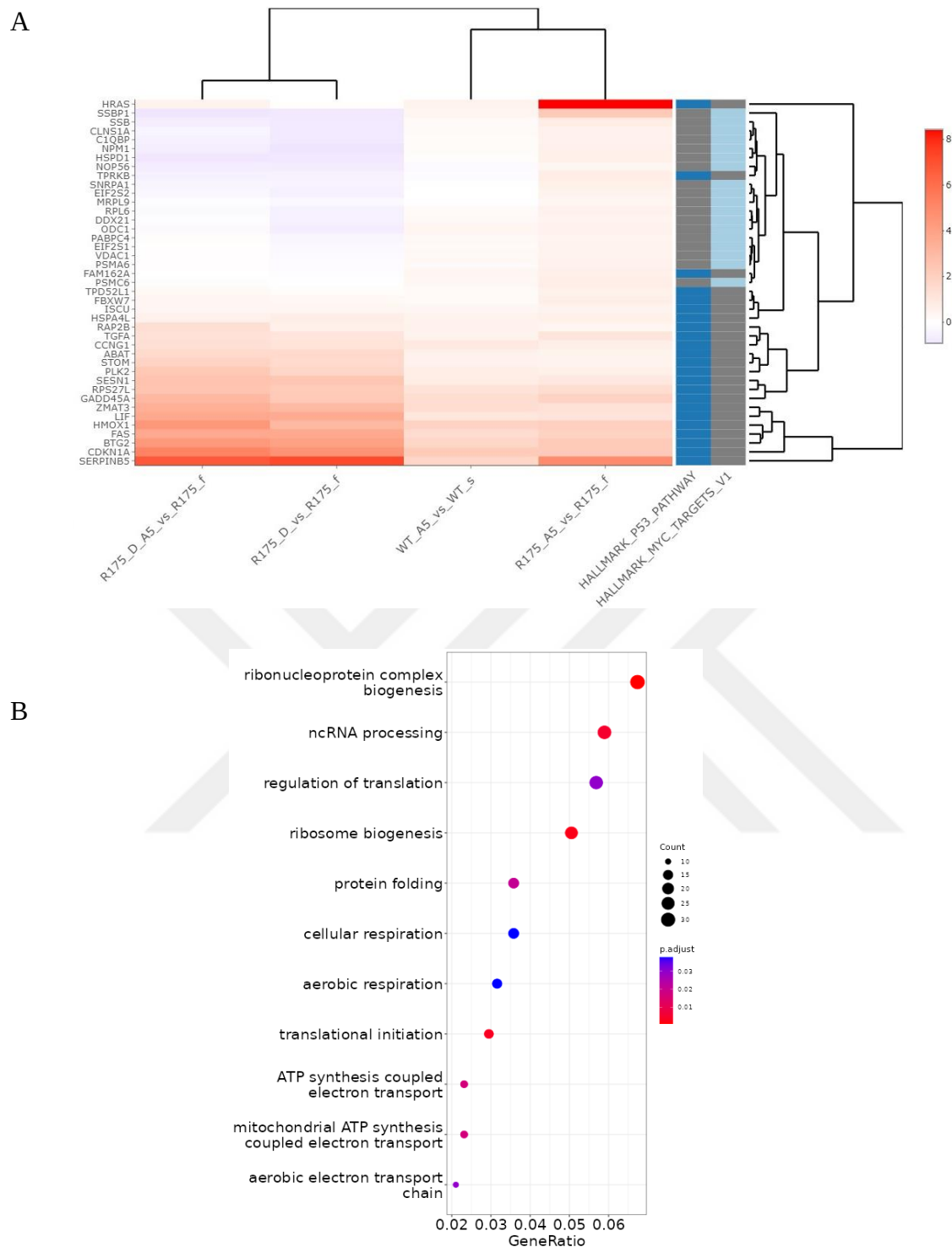


Figure 4.35 **A** Heatmap of hallmark P53 and MYC targets for the upregulated genes from the siCHRNA5 treated R175H mutant cell line. **B** GO enrichment for these genes. “Translational processing” and “ncRNA processing” appear to be important enriched terms in regulating gene expression.

From the MSigDB database, Hallmark p53 targets and MYC targets are enriched. One surprising gene which showed up as a *TP53* target was the *HRAS* gene which is only upregulated in the si*CHRNA5* treatment and not in doxorubicin treated or in the combinatorial. This was interesting since, there is association between Ras activation and breast cancer (Zhumakayeva et al., 2019).

In the R175H mutant cell line treated with only si*CHRNA5*, there were 385 genes downregulated which satisfied a p-value cutoff of 0.01 and a logFC less than 0. Enrichment for these downregulated genes showed various “miRNA transcription” terms, along with “TGF beta receptor signaling pathway” and “intrinsic apoptotic signaling pathway in response to DNA damage”. The Dorothea enriched Transcription factors mainly showed zinc finger domain factors. Enriched MSigDB pathways showed hallmark estrogen response targets, KEGG Cell cycle and DNA replication targets. For these targets, the effects of doxorubicin were more prevalent in the combinatorial in comparison to the si*CHRNA5* contrast.

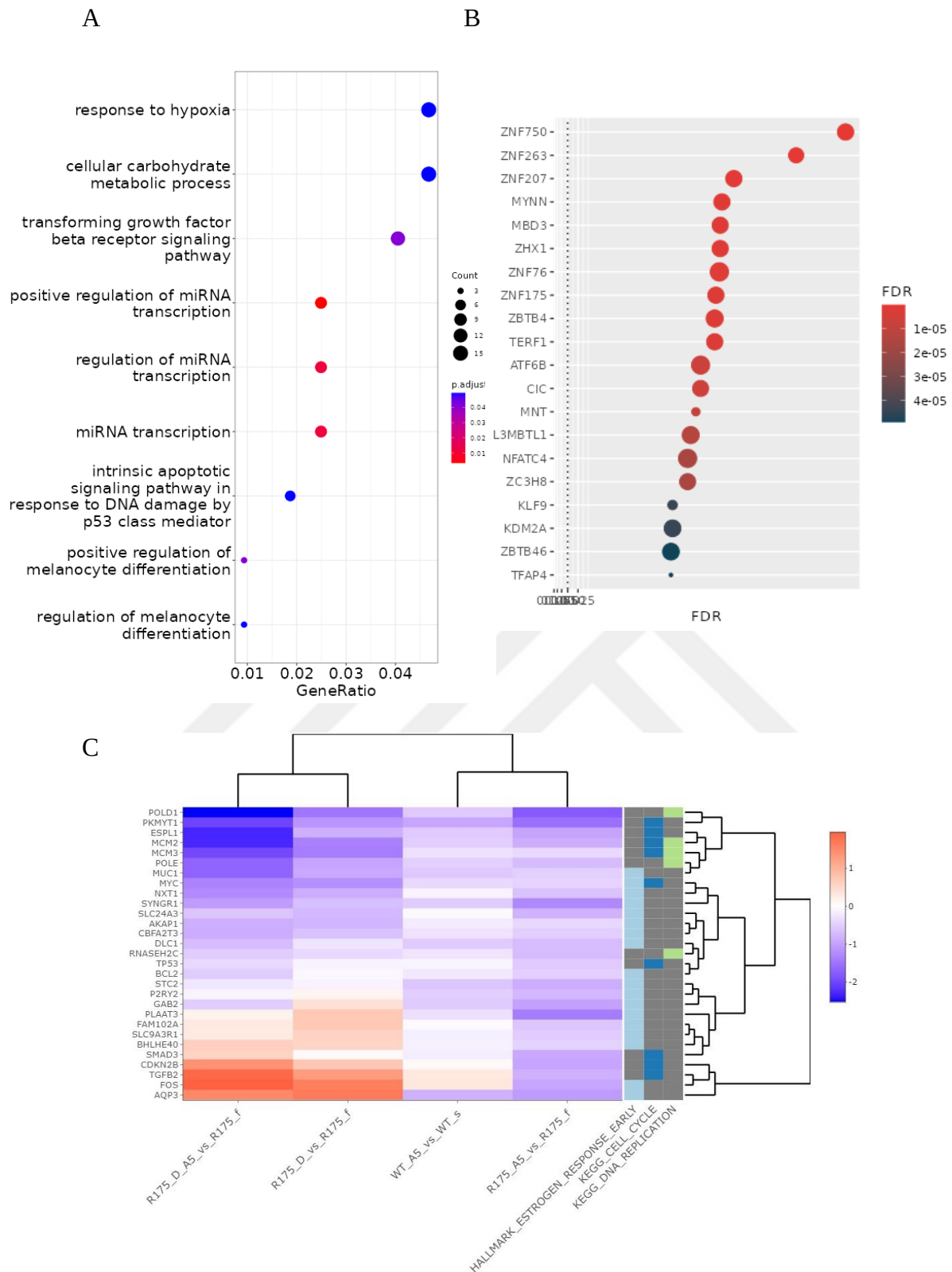


Figure 4.36 **A** GO enrichment for the downregulated genes of the R175H si*CHRNA5* treated contrast. Transcription and apoptotic enriched terms show up. **B** TFE shows mostly zinc finger domain factors as enriched. **C** Heatmap for the hallmark estrogen response, KEGG cell cycle and DNA replication terms. Column wise clustering show that these targets appear to be more correlated between the combinatorial and the doxorubicin treatments.

Then I filtered the DREAM targets and the LIN37 targets for these three contrasts. I filtered the DREAM targets further with p-value threshold of 0.05 across the three R175H contrasts. The si*CHRNA5* treatment on its own shows a cluster of upregulated genes, which is probably overcome by doxorubicin in the combinatorial. These genes could be further explored to see if doxorubicin and si*CHRNA5* are working antagonistically for a particular pathway or during a certain phase of the cell cycle.

Functional profile for these filtered targets showed nuclear division, DNA replication and other cell regulatory pathways. The TFE showed the E2F family of TFs as most significant. This is expected since these genes are a subset of the DREAM targets.

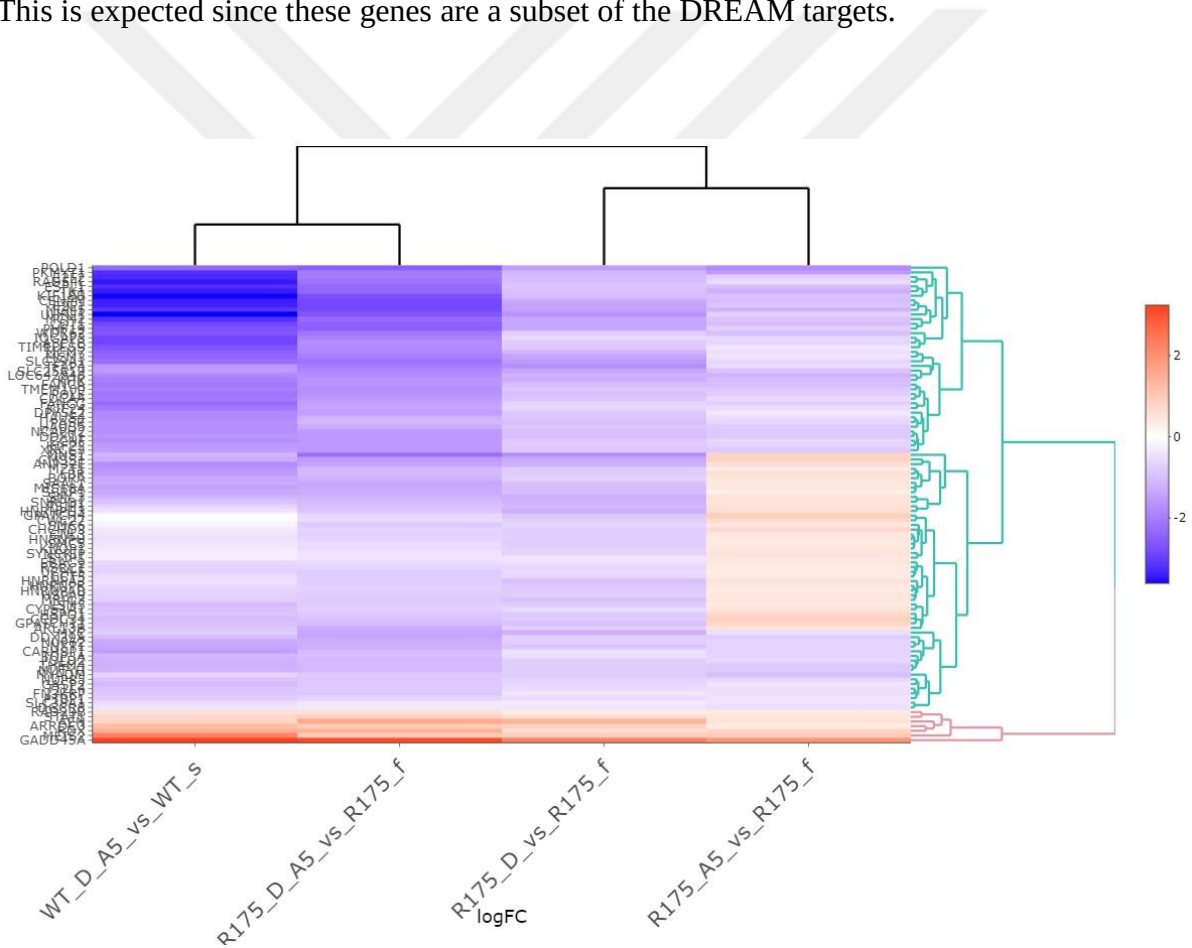


Figure 4.37 Heatmap of the DREAM targets (filtered for p-value < 0.05 for the R175H contrasts).

For the LIN37 dependent *TP53* targets, I first set a p-value threshold of 0.05 throughout all the R175H contrasts. 27 genes fulfilled this filter. All of these genes appeared to have an additive or a synergistic effect in the combinatorial treatment. Enriched pathways for these genes showed up as mitotic nuclear division, double strand break repair, and replication. All of the genes for the LIN37 gene list showed a similar pattern, with higher expression in the combinatorial compared to both the doxorubicin treatment and the si*CHRNA5* treatment.

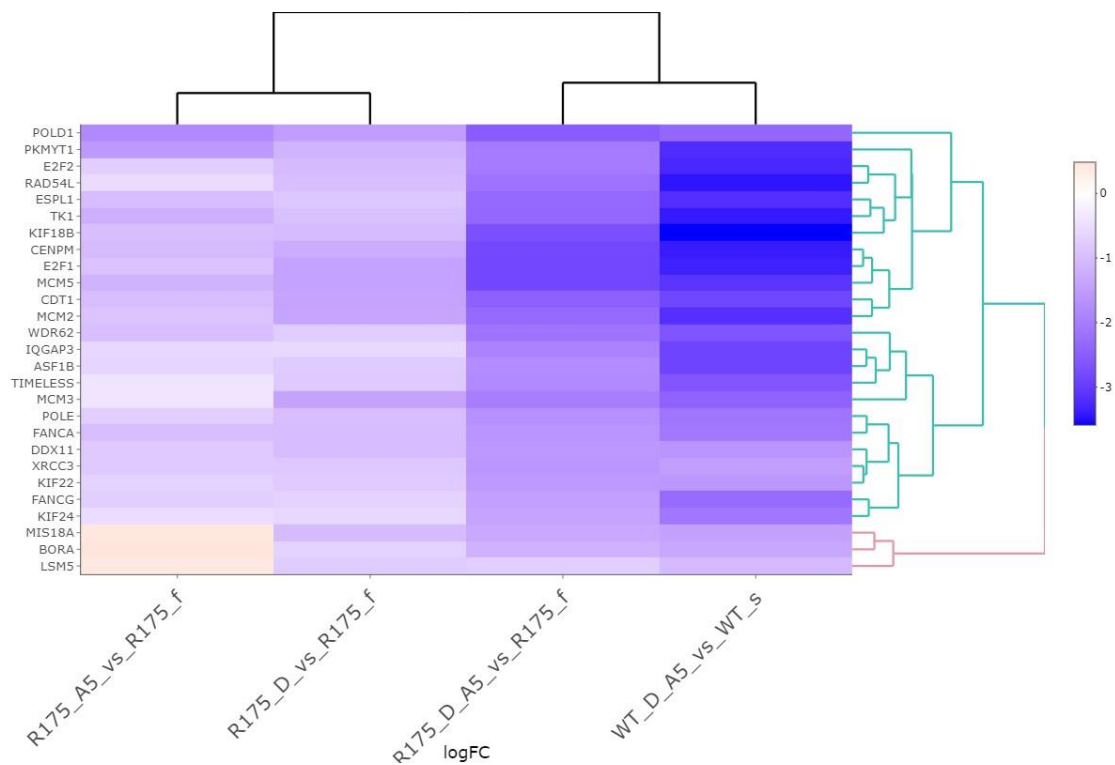


Figure 4.38 LIN37 targets significant in all three contrasts.

Then I filtered the LIN37 dependent *TP53* targets, with a 0.01 FDR cutoff, in the combinatorial treatment alone. This filtered 191 genes out of the total 260 LIN37 gene set. The treatment gene network for these genes suggested there was certain group of genes that were synergistically affected. In the network, the genes colored in white, solely connected to the combinatorial treatment (diamond), showed that these genes were significant only in this treatment. These genes included *AURKA*, *AURKB*, *CENPL*, *SHCBP1*, *EME1*, *FAM72D*,

DLGAP5, ANLN, RNASEH2A, MASTL, TOP2A, CDCA8, FAM83D, SPAG5 (Figure 4.39).

This could suggest that these genes were synergistically affected in this cell line even though most trends pointed towards a higher influence of doxorubicin in the combinatorial. The dark grey colored genes were the 27 genes shown in Figure 4.35. *AURKA, AURKB* and *CENPL* are all very important genes for cell division and the synergistic effect for these genes should be explored further.

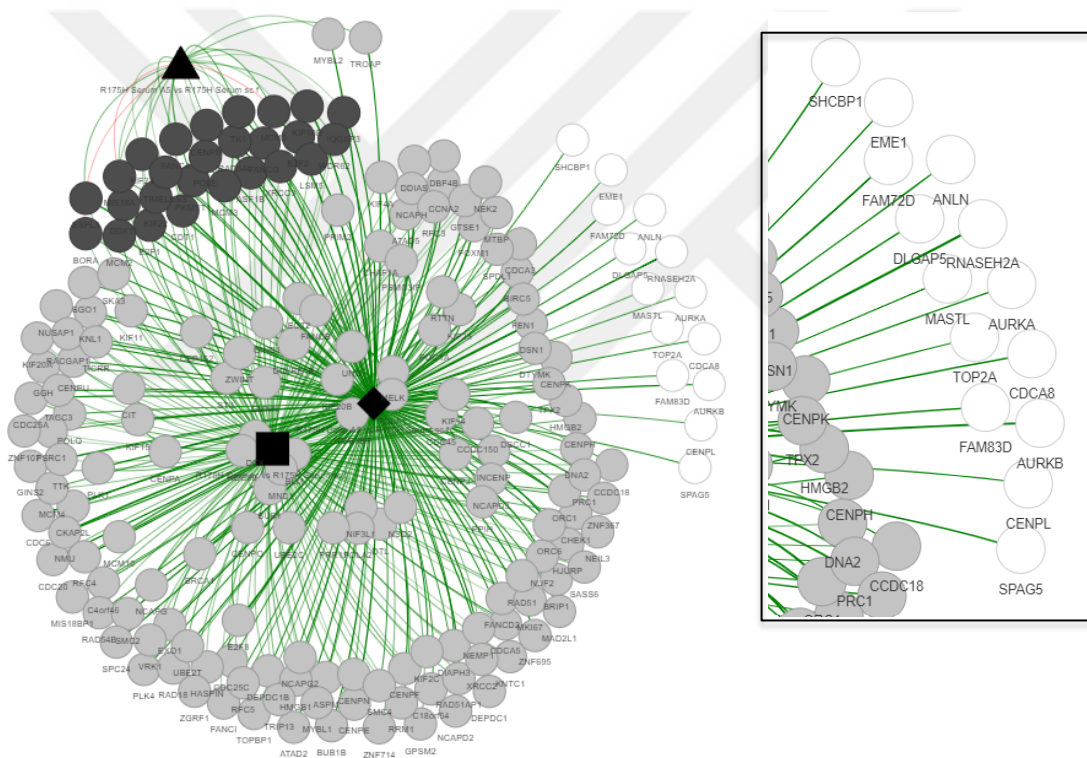


Figure 4.39 Treatment-gene network for the LIN37 targets for the R175H contrasts. The dark grey nodes are genes which are significant in all three contrasts (previous heatmap). The white colored nodes are connected to the diamond node (combinatorial treatment) and hence could be potentially synergistic genes since they not significant in the other treatments.

4.2.2.4 R273H contrasts

The effect of combinatorial in R273H cell line was explored next. The following contrasts were plotted;

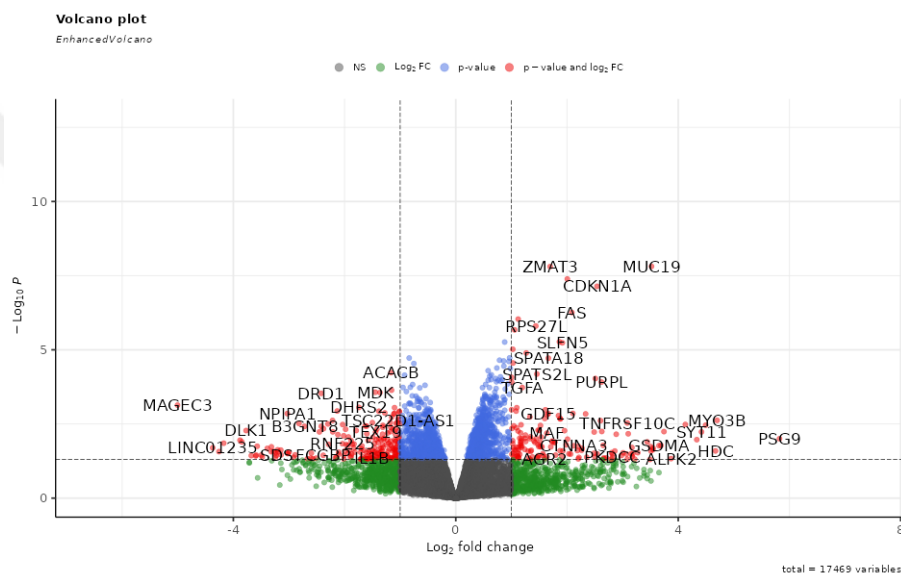
R273H_siCHRNA5 vs R273H_Control;

R273H_Doxorubicin_siCHRNA5 vs R273H_Control;

WT_Doxorubicin vs WT_control;

WT_Doxorubicin_siCHRNA5 vs WT_control.

A



B

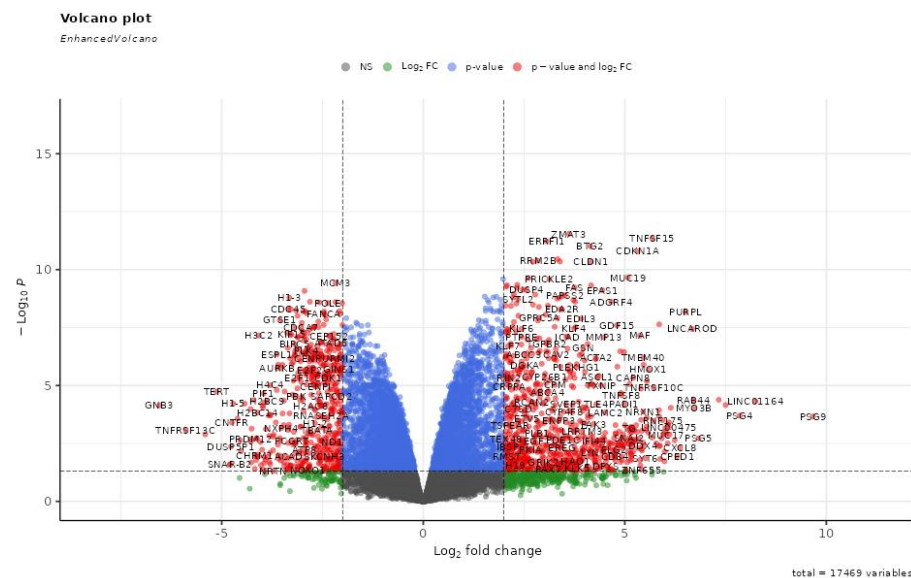


Figure 4.40 **A** volcano plot for the siCHRNA5 treatment in the R273H cell line. *P21* can again be seen as upregulated by the silencing of *CHRNA5*. **B** Regulation of genes in the R273H cell line with the combinatorial treatment.

For the R273H mutant cell line, we did not have a sample for the doxorubicin treatment alone. A FDR of 0.01 proved to be very strict for this mutant cell line, therefore a p-value cutoff of 0.01 was used to filter DEGs. This filtered a total of 694 genes. When plotted against the combinatorial, there was a linear trend with a 0.589 R^2 value.

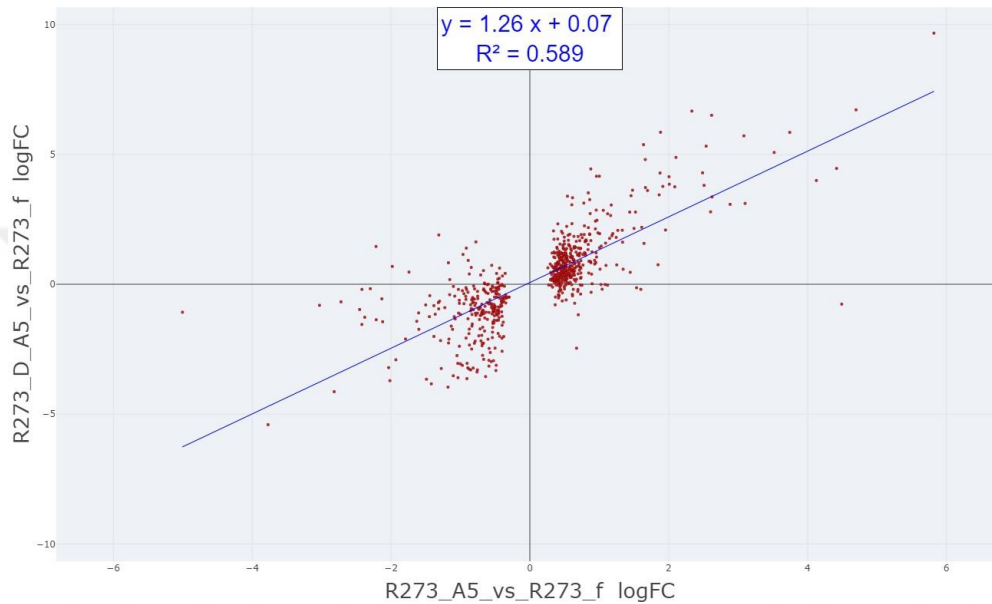


Figure 4.41 Scatter plot of the significant genes in the si*CHRNA5* contrast. The si*CHRNA5* contrast is on the x-axis and the combinatorial is on the y-axis.

With a logFC cutoff of over zero, there were 430 genes upregulated. These genes included *P21*, *TGFA*, *PPARG*, *VEGFC*, *FAS* and *GDF15*. Enrichment for these upregulated genes showed *TP53* and *ELK3* as the enriched TF and showed p53 related pathway and DNA damage response as enriched terms.

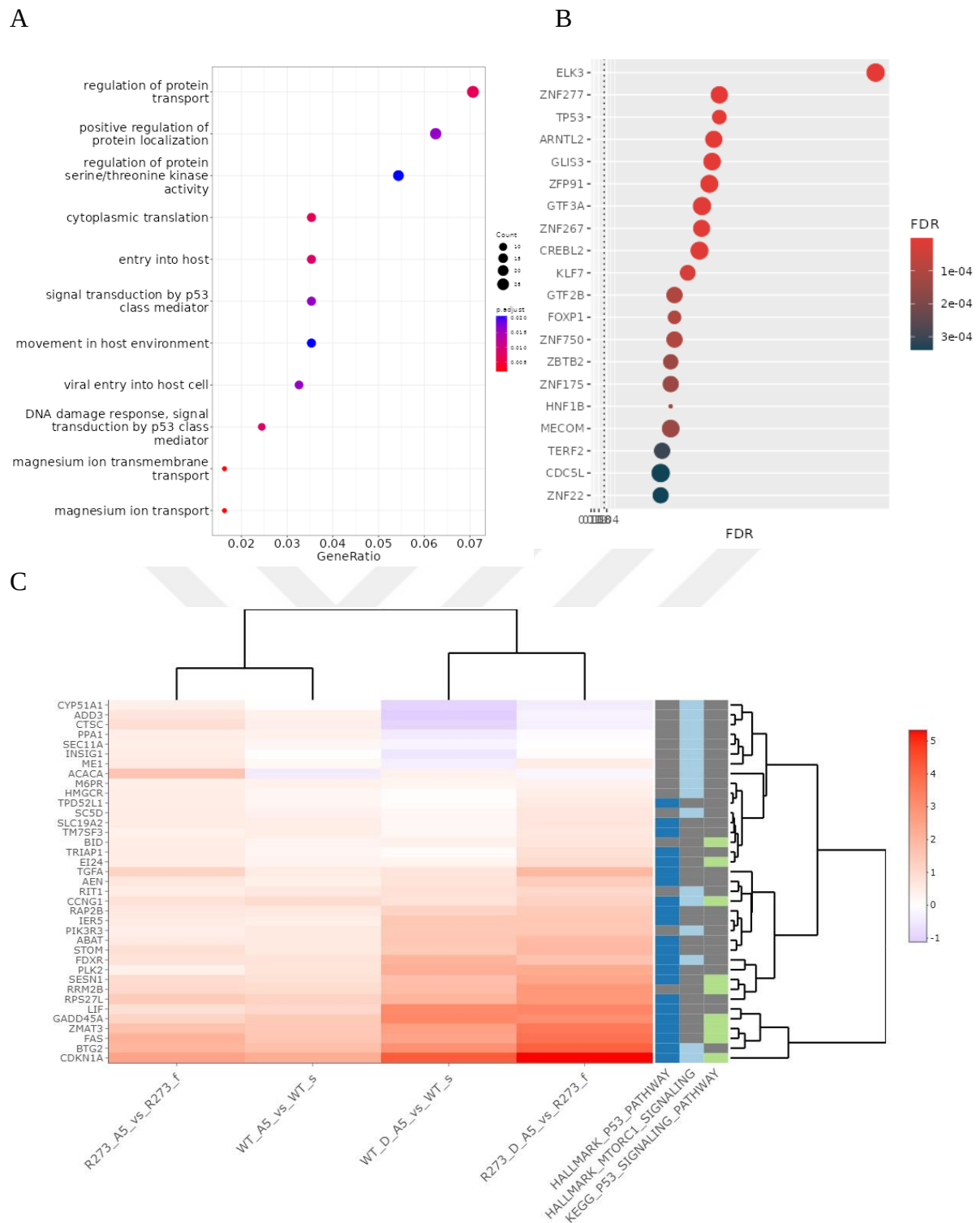


Figure 4.42 **A** Gene Ontology enriched terms for the upregulated genes in the si*CHRNA5* treated R273H cell line. **B** Dorothea based TFE shows *ELK3* and *TP53* as the top enriched TF. **C** Heatmap showing the genes as targets of the hallmark and KEGG p53 pathway and MTORC1 signaling.

The enrichment terms from the MSigDB included p53 targets and MTORC1 signaling pathway. Target genes for MTORC1 pathway appear to be downregulated in the WT contrast. For most of these upregulated genes, the logFC was higher in the combinatorial cell line compared to the si*CHRNA5* alone. *P21* and *FAS* seem to be upregulated more in the combinatorial, but synergism or additiveness could not be concluded due to the lack of the doxorubicin single treatment.

The same cutoffs (p-value < 0.01 and logFC < 0) filtered 264 genes as downregulated. Significant genes included *DRD1*, *GABRP*, *PRRT2*, and *MDK*. No GO:BP or GO:MF term was enriched from this geneset but CC (cellular component) terms showed some cell cycle genes being associated with “spindle pole” and “centriole”.

Since these cutoffs did not prove any conclusive results for the contrasts, we profiled the DREAM targets and the LIN37 dependent *TP53* targets for the combinatorial treatment in the R273H mutant cell line. The LIN37 target genes are significantly downregulated more compared to the si*CHRNA5* treatment but synergism could not be concluded due to the lack of the doxorubicin treatment alone. With this geneset, some of the enriched terms included, DNA replication, mitotic cell cycle transition, and chromosome segregation. The DREAM targets showed a very similar pattern to the LIN37 geneset, with genes downregulated more in the combinatorial, compared to the single si*CHRNA5* treatment. Since these are already a subset of the DREAM targets, the enriched terms included E2F targets, cell cycle checkpoint and G2M checkpoint.

4.2.2.5 WT contrasts

For the comparisons of the WT cell line, the following contrasts were analyzed;

WT_Doxorubicin vs WT_Control;

WT_siCHRNA5 vs WT_Control;

WT_Doxorubicin_siCHRNA5 vs WT_Control;

EV_Doxorubicin vs EV or EV_siCHRNA5 vs EV or EV_Doxorubicin_siCHRNA5 vs EV_control.

The final contrast for the EV cell line was changed according to the WT treatment being analyzed. For example if the effect of doxorubicin alone was being explored, I used the doxorubicin treatment for the EV also and for the other treatments, I used the same treatment groups

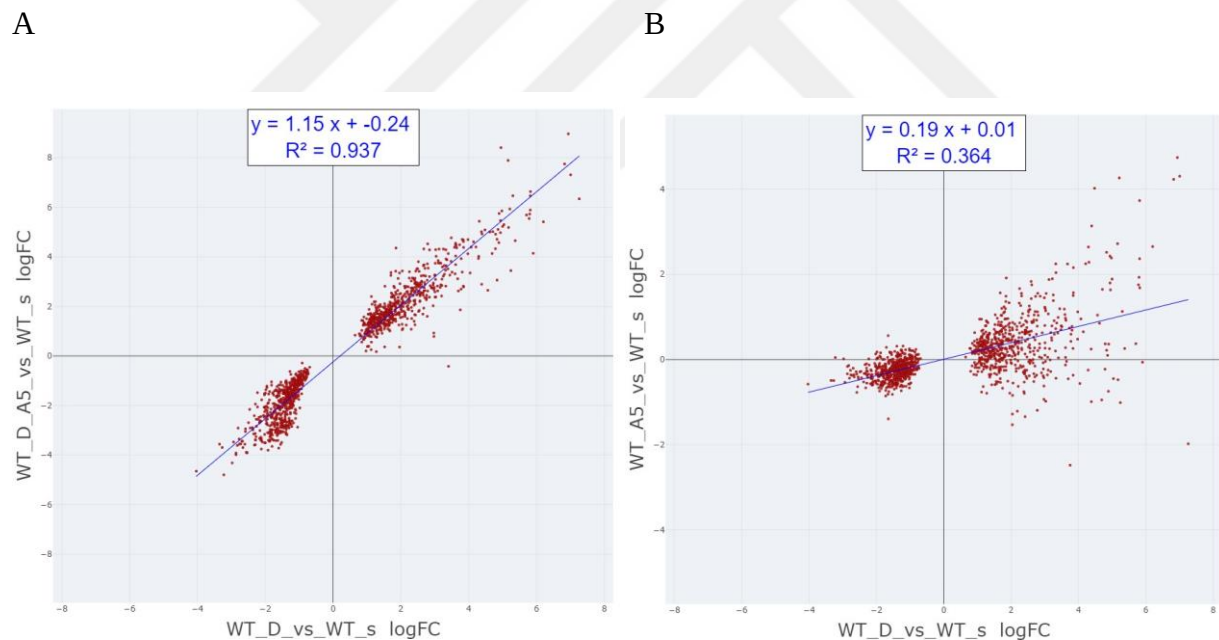


Figure 4.44 DEGs from doxorubicin treated WT contrast. **A** The doxorubicin is plotted on the x-axis and the combinatorial on the y-axis. These appear to highly correlated with one another. **B** The doxorubicin is plotted on the x-axis and the siCHRNA5 on the y-axis.

For the doxorubicin treatment, FDR of 0.01 filtered 1164 genes. These genes were highly correlated within the doxorubicin and the combinatorial treatment. Filtering with a logFC of greater than 0, filtered 590 genes. These upregulated genes enriched various carcinomas from the DisGeNET database, including “Invasive carcinoma of the breast”.

The GO enrichment terms included similar terms seen with other cell lines with the doxorubicin treatment. These included terms such as tissue, cell migration and “response to hypoxia”. The MSigDB enriched “P53 pathway”, “estrogen response” along with “hypoxia” and “apoptosis” (Figure 4.45 A). The TFE showed *TP53* which was expected on doxorubicin treatment. *ESR1* and *ESR2* were also significantly enriched. The heatmap for the “apoptosis” term and the “P53 hallmark targets” was plotted. The cluster of genes at the lower end, show additiveness or synergism because the expression for these genes increases in the combinatorial compared to the single treatment. *HMOX1* for example, which is a NRF2 target, appears to be expressed almost double compared to the single si*CHRNA5* treatment (Figure 4.45 C).

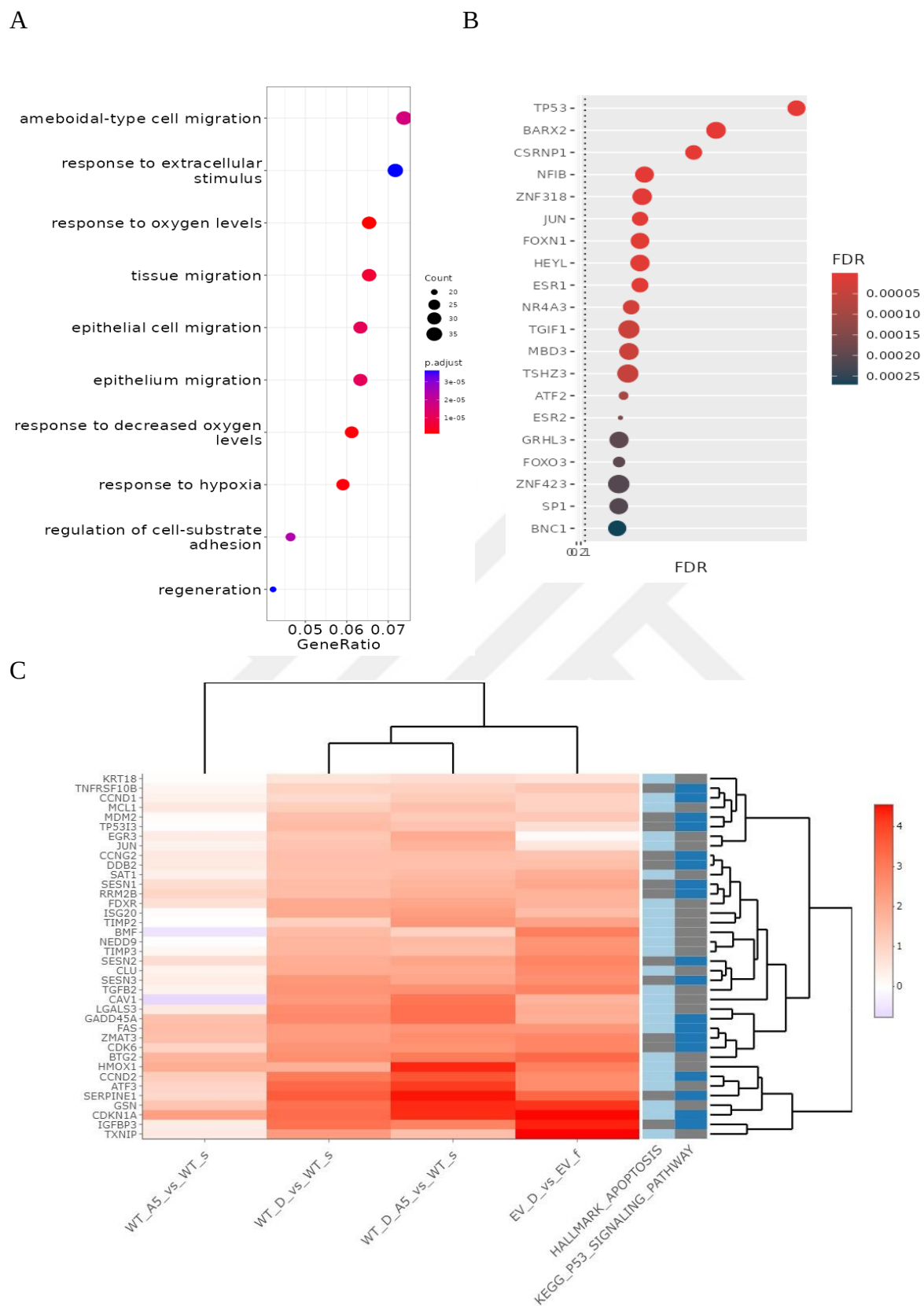


Figure 4.45 **A** GO enrichment for the doxorubicin treated WT upregulated genes. **B** TFE analysis for the same gene set. Shows TF such *TP53*, *ESR1* and *ESR2*. **C**, Heatmap for the “apoptosis” and “P53 signaling pathway”.

Filtering down with a logFC of less than 0, filtered 574 genes. These downregulated genes for the doxorubicin treated WT cell line, enriched multiple GO cell regulatory terms such as “DNA replication”, “chromosome segregation”, “mitotic cell phase transition” (Figure 4.46 A). The transcription factor enrichment showed the E2F family of TFs, along with *MYC* and *FOXM1* (Figure 4.46 B). The MSigDB also showed various cell cycle related terms such as “G2M checkpoint”, “DNA replication”, “E2F targets”. The clustering in the heatmap shows that the doxorubicin treatment is similar in the WT and EV cell lines for these downregulated genes (Figure 4.46 C).



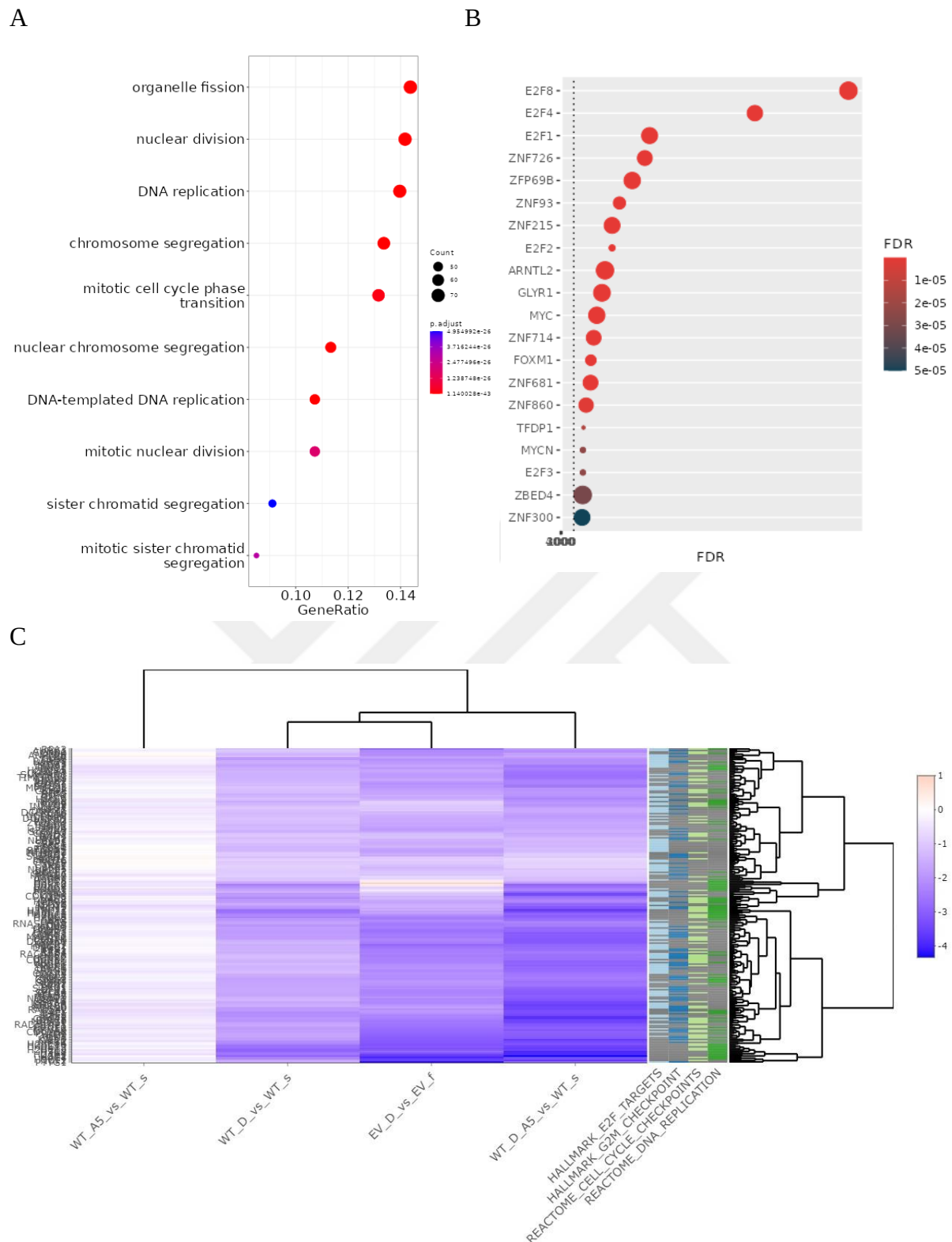


Figure 4.46 **A** GO enrichment terms for these doxorubicin treated WT downregulated genes showed various cell regulatory pathways. **B** TFE shows E2F family of transcription factors along with MYC and FOXM1. **C** Heatmap of the downregulated genes of the E2F targets and cell regulatory terms.

Then I filtered the si*CHRNA5* contrast for DEGs. A p-value cutoff of 0.01 filtered 549 genes. These genes showed a linear trend with the combinatorial treatment, and an R^2 value of 0.66. However, the same filtered genes gave a R^2 value of 0.907 for the doxorubicin treatment against the combinatorial. It appears that in the WT cell line, the doxorubicin treatment on the combinatorial has a very significant effect, even with the genes that are significantly expressed for the si*CHRNA5* treatment. There were 319 upregulated genes ($\logFC > 0$) and 230 downregulated genes ($\logFC < 0$).

The terms enriched by the upregulated genes included “DNA damage response” along with “signal transduction by p53 class mediator”. *TP53* and *ELK3* are the two top enriched TF from the dorothea database. P53 related terms that were enriched from the MSigDB database were plotted as a heatmap. The doxorubicin contrast was clustered together with the combinatorial while the si*CHRNA5* treated WT and EV were clustered together (Figure 4.48 C).

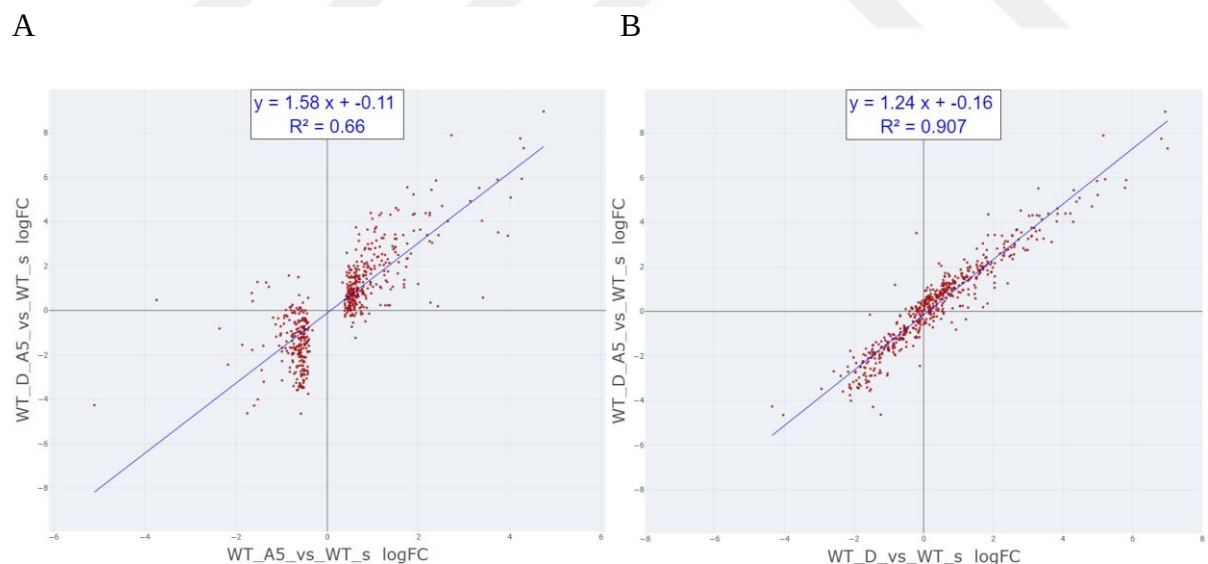


Figure 4.47 Scatter plots for the si*CHRNA5* treated DEGs. **A** the si*CHRNA5* is plotted on the x-axis against the combinatorial on the y-axis. **B** For the same genes differently expressed by the si*CHRNA5*, the doxorubicin against the combinatorial was plotted. This showed a very correlation.

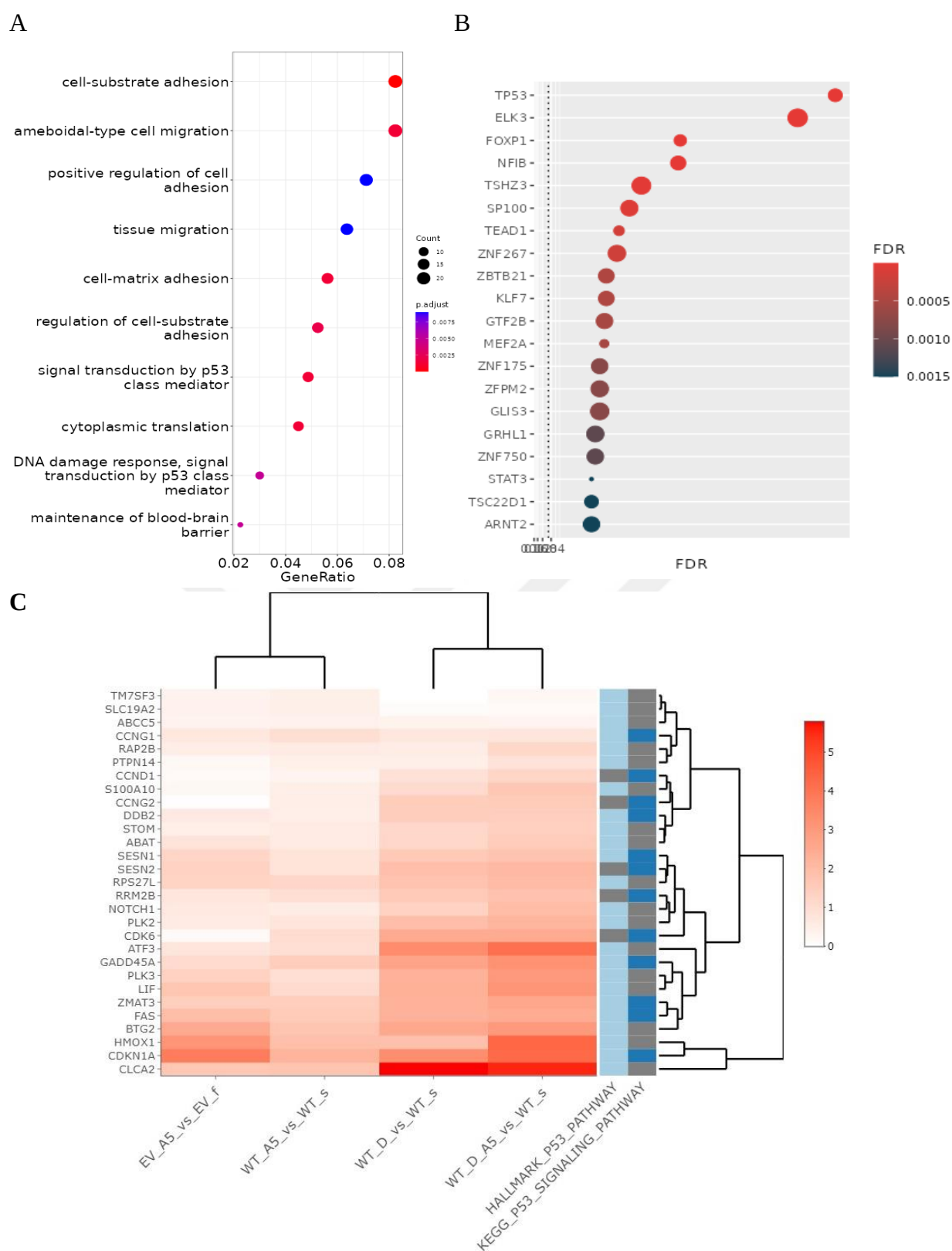


Figure 4.48 **A** GO enriched terms for the upregulated genes of the si*CHRNA5* WT contrast. **B** TFE for the same genes shows *TP53* as the most significantly enriched TF. **C** Heatmap of the upregulated genes based on the MSigDB database terms relating to P53.

The 230 downregulated genes for this si*CHRNA5* contrast enriched various replication pathways, and cell regulatory terms. Accordingly, the Dorothea enriched TF included the E2F family of TFs. The MSigDB database enriched terms such as “E2F targets”, and various other cell regulatory terms. The heatmap shows the downregulated genes plotted for the most significant TF. E2F8 has a low confidence level in the Dorothea database and therefore was not plotted.

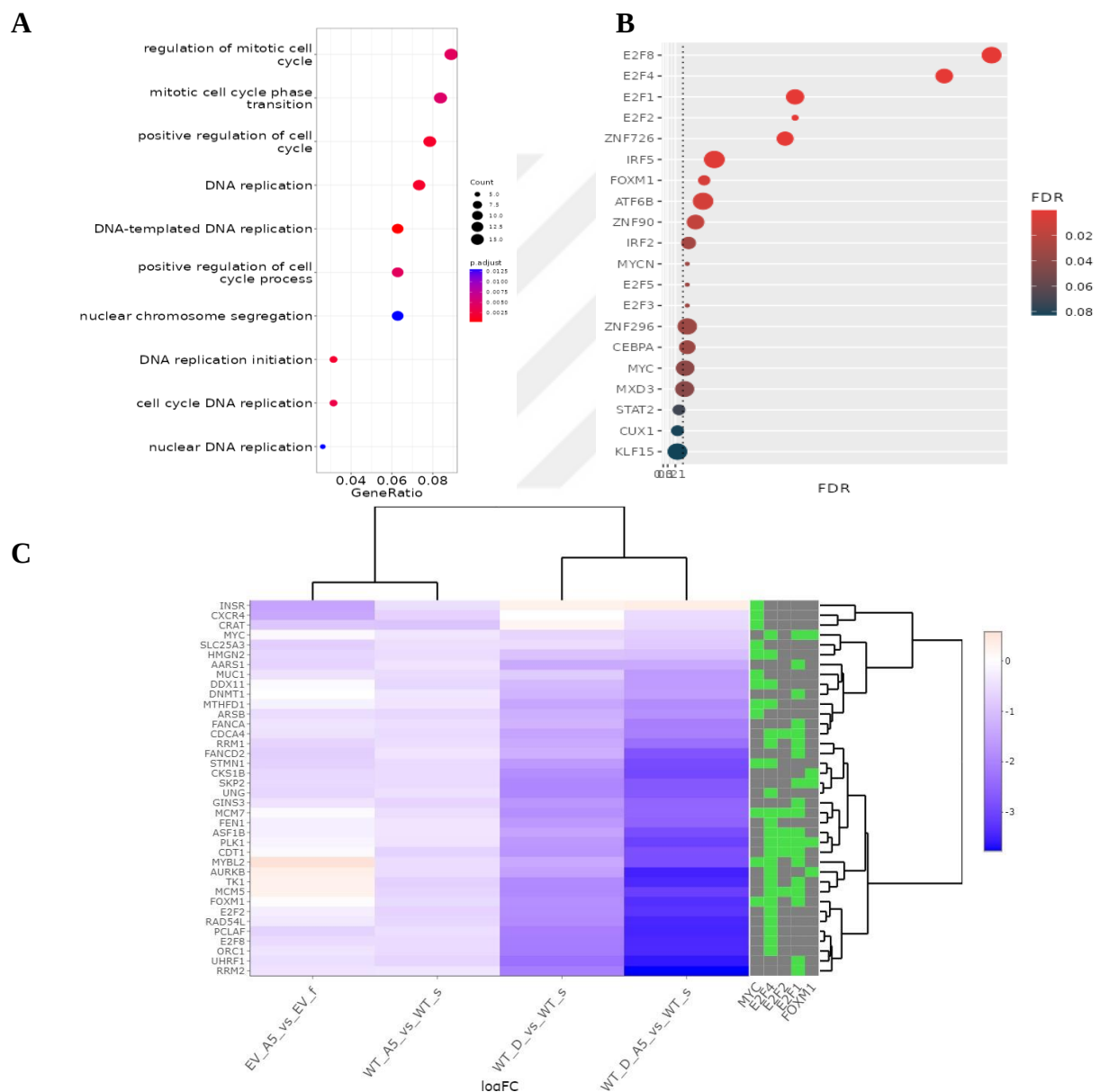


Figure 4.49 **A** Enriched terms for the downregulated genes of the si*CHRNA5* WT contrast. These include many cell regulatory and replication terms. **B** TFE shows the E2F family of transcription factors. **C** Heatmap for the significant TFs, *MYC*, *E2F4*, *E2F2*, *E2F1* and *FOXM1*. All of these TF have a high confidence level in the Dorothea database.

4.2.2.6 Effect of Combinatorial in-between the cell lines

Based on the IGV analysis of EV_Doxorubicin_siCHRNA5 (Appendix B), it suggests that this sample most likely does not belong to EV isogenic MCF7 cell line hence its position in any cluster analysis should not be relied upon.

The final set of contrasts was for comparison of the combinatorial effect between the three different cell lines;

R175H_Doxorubicin_siCHRNA5 vs R175H_control;

R273H_Doxorubicin_siCHRNA5 vs R273H_control;

WT_Doxorubicin_siCHRNA5 vs WT_control.

The cutoffs used for the R175H cell line were 0.01 for the FDR and 2 for the logFC. The upregulated genes filtered were 348. Transcription factors enriched with the upregulated genes included *TP53*, *ESR1*, *JUN* and *FOXN1* (Figure 4.50 B). GO enrichment showed hypoxia terms along with “DNA damage response” and “signal transduction by p53 class mediator” (Figure 4.50 A). This is very similar to the enrichment for the upregulated genes for the doxorubicin treatment in the R175H cell line. In the MSigDB database, estrogen response terms were enriched, along with p53 pathway and apoptosis terms. For these particular targets, the R175H and the R273H cell lines are clustered together on the heatmap (Figure 4.50 C).

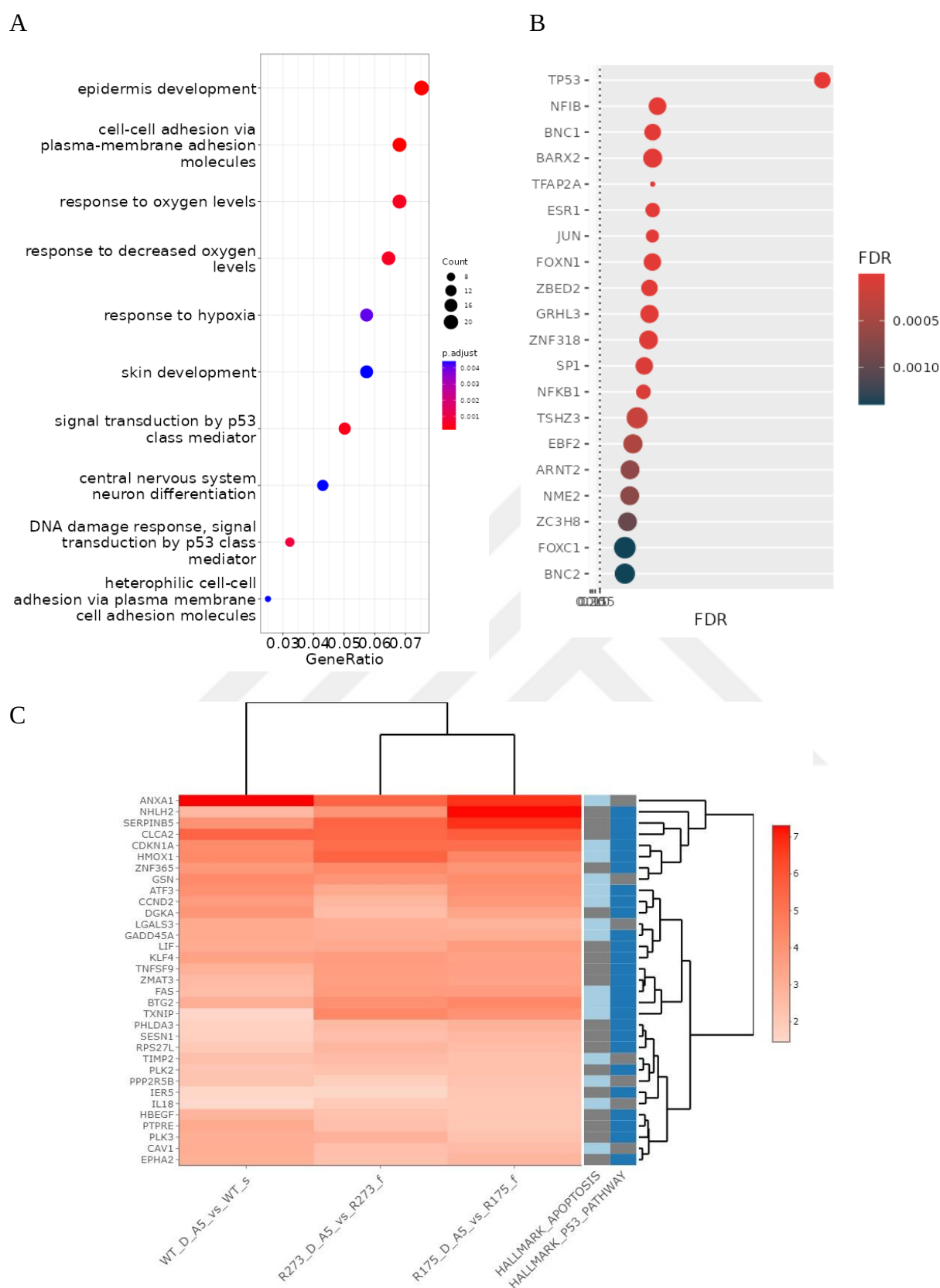


Figure 4.50 **A** GO enrichment for the upregulated R175H combinatorial genes. Terms enriched are very similar to the ones enriched by the doxorubicin treatment in R175H. **B** TFE based on the Dorothea database. **C** Heatmap of the upregulated genes for the Hallmark P53 and apoptosis pathways.

The downregulated genes for this R175H combinatorial treatment were 143. The downregulated genes for R175H combinatorial showed many family members of the E2F TF, which are important regulators of the cell cycle. This has been the trend throughout the combinatorial treatments. The GO enrichment showed terms related to cell regulatory pathways such as “regulation of cell cycle”, “nuclear division” and “DNA replication” (Figure 4.51 A). For these downregulated genes, the “Hallmark E2F targets” and “G2M checkpoint genes” from the MSigDB database, were plotted as a heatmap (Figure 4.51 C). For these target genes, it appears that the combinatorial has a more significant effect in downregulating them for the R273H cell line, compared to the R175H cell line.

For the R273H mutant cell line treated with the combinatorial treatment, 394 genes were upregulated with the same cutoffs for the FDR (< 0.01) and the logFC (> 2). These genes enriched response to hypoxia and some development terms from the GO database. Enriched TF showed *JUN*, *ESR1*, *NFKB1*, along with *TP53*. For these genes, the heatmap was plotted for the apoptosis and the p53 hallmark genes from the MSigDB (Figure 4.52 C). These upregulated genes show a response, which is similar in the mutant cell lines in comparison to the WT.

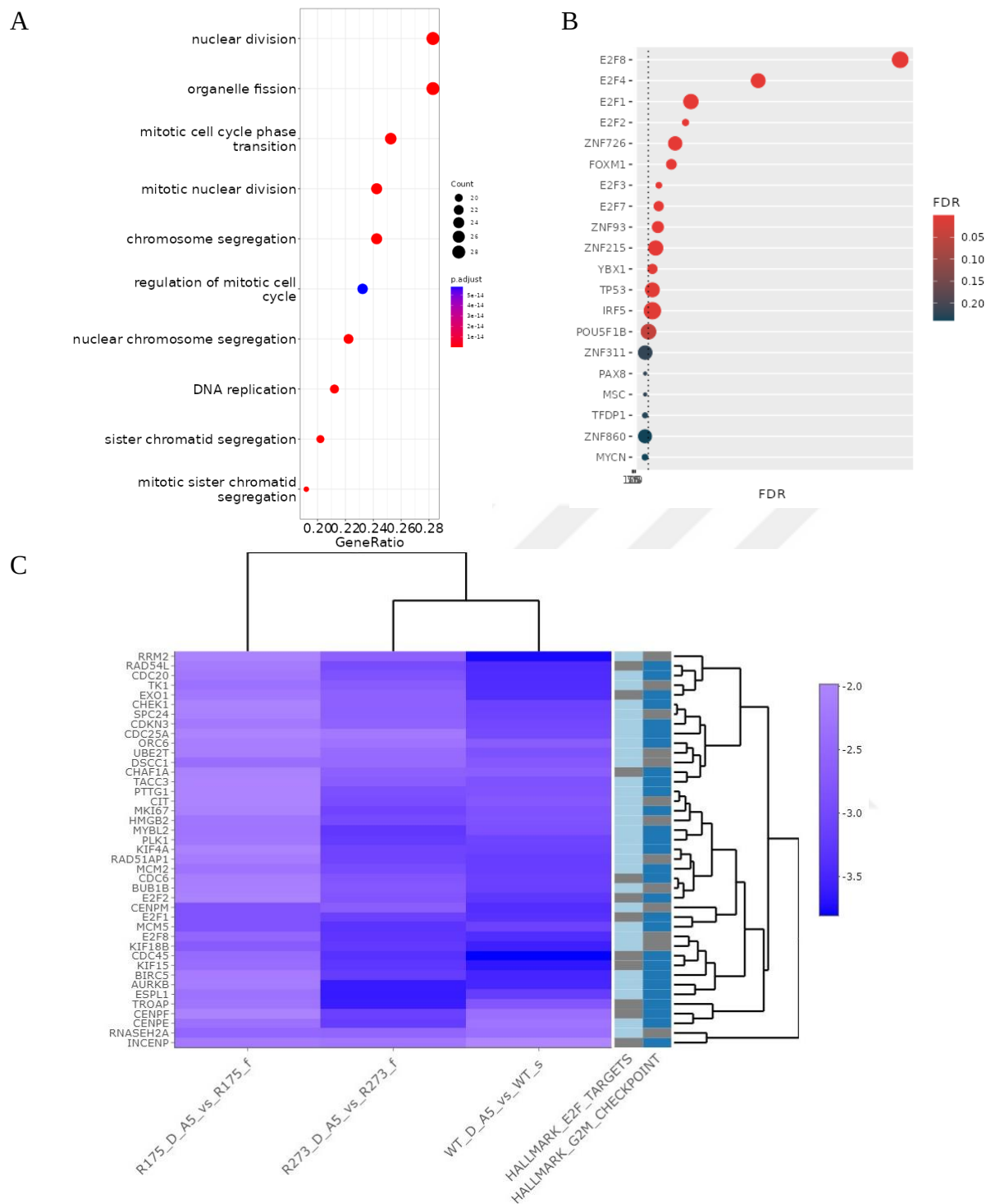


Figure 4.51 **A** GO enriched terms for the downregulated R175H mutant cell line genes. In line with the individual treatments for the R175H cell line, various cell regulatory pathways appear to be enriched. **B** TFE shows the E2F family of TFs. **C** Heatmap for the downregulated genes for the Hallmark E2F and G2M checkpoint genes. The clustering shows an interesting pattern with R273H clustered together with the WT.

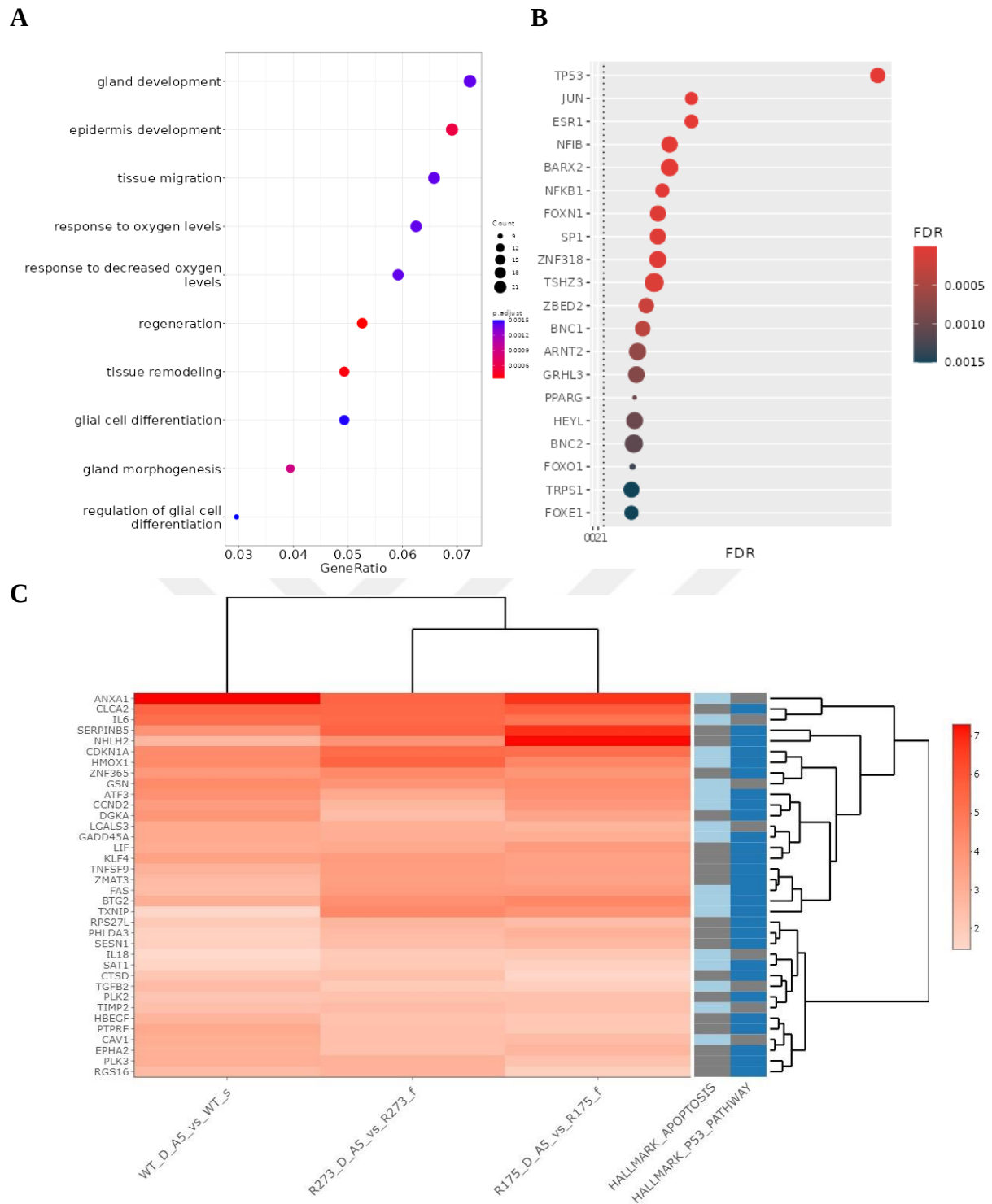


Figure 4.52 **A** GO enriched terms for the upregulated genes in the R273H cell line with combinatorial treatment. **B** TFE based on the dorothea database. **C** Heatmap for the apoptosis and p53 enriched terms from the MSigDB database. The clustering shows the response of these upregulated genes in a similar manner for the mutant cell lines in comparison to the WT overexpressed cell line.

There were 340 genes downregulated from this R273H contrast treated with the combination of doxorubicin and si*CHRNA5*. The enriched terms for these genes were almost the same as the R175H (treated with combinatorial) and in line with the downregulation of the cellular pathways. The enriched transcription factors included the E2F family of TF along with the zinc finger domain factors. The MSigDB enriched hallmark G2M checkpoint genes, E2F targets and other KEGG replication terms. However, here I plotted the TF from the dorothea as the heatmap to observe the trend of these downregulated genes across the cell lines.



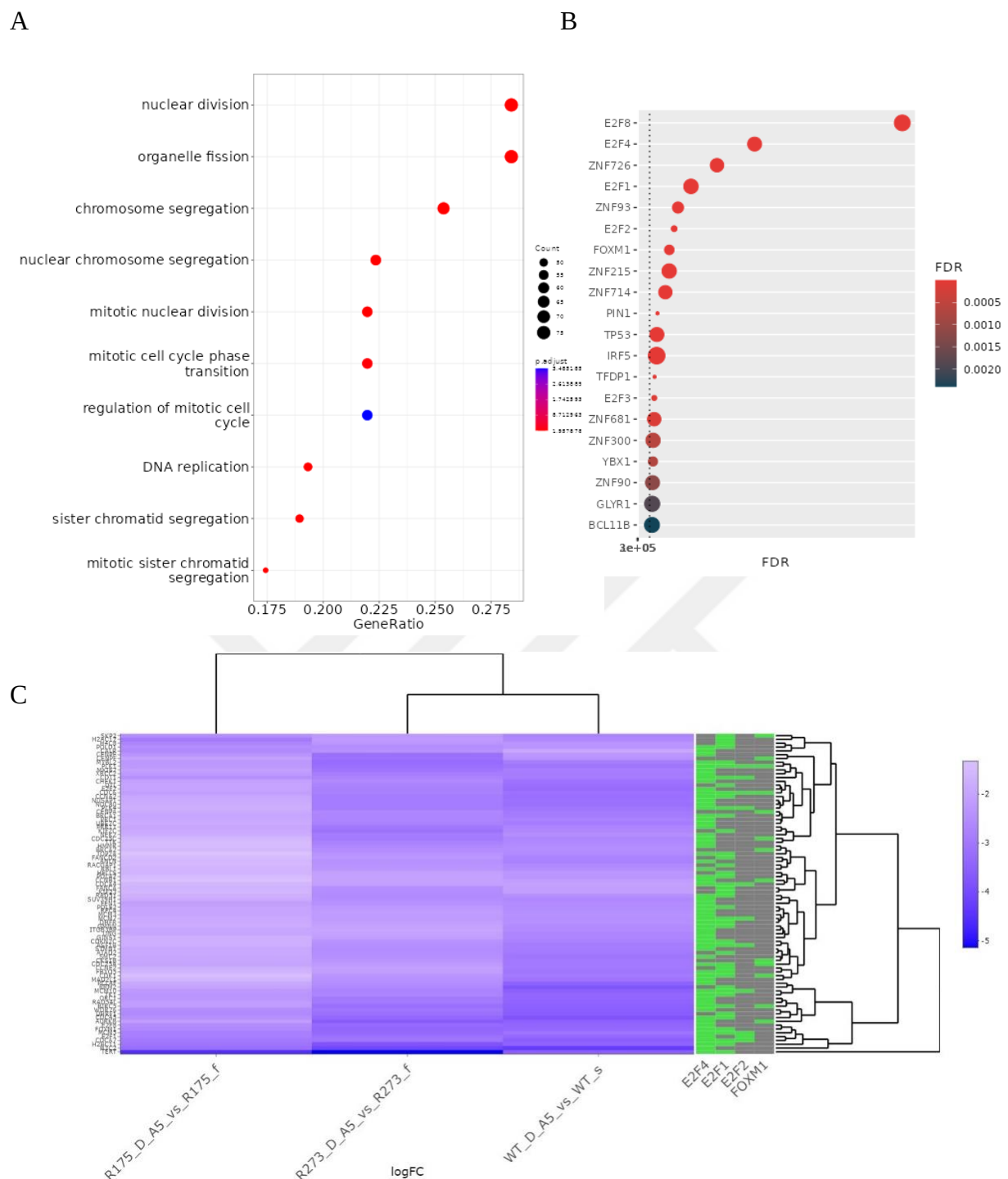


Figure 4.53 **A** GO enrichment terms for the downregulated genes of the R273H mutant. **B** TFE based on the dorothea database. **C** Heatmap of the *E2F4*, *E2F1*, *E2F2* and *FOXM1* TF based on the Dorothea database. The downregulated genes here for the R273H are also affected in a similar manner to that in the WT.

The WT contrast for the combinatorial treatments filtered 410 genes with an FDR value of 0.01 and a logFC cutoff of 2. These upregulated genes for the WT contrast showed tissue and epithelium migration and response to oxygen terms. Similar to the previous contrasts, p53 and apoptosis showed up as enriched terms in the MSigDB. The heatmap for those target genes is plotted. There were 313 genes filtered as downregulated ($\log\text{FC} < -2$). From the DisGeNET database, the disease term of “abnormality of chromosome stability” was enriched for these downregulated genes. GO enrichment showed the same pathways as the previous contrasts and their downregulated genes.



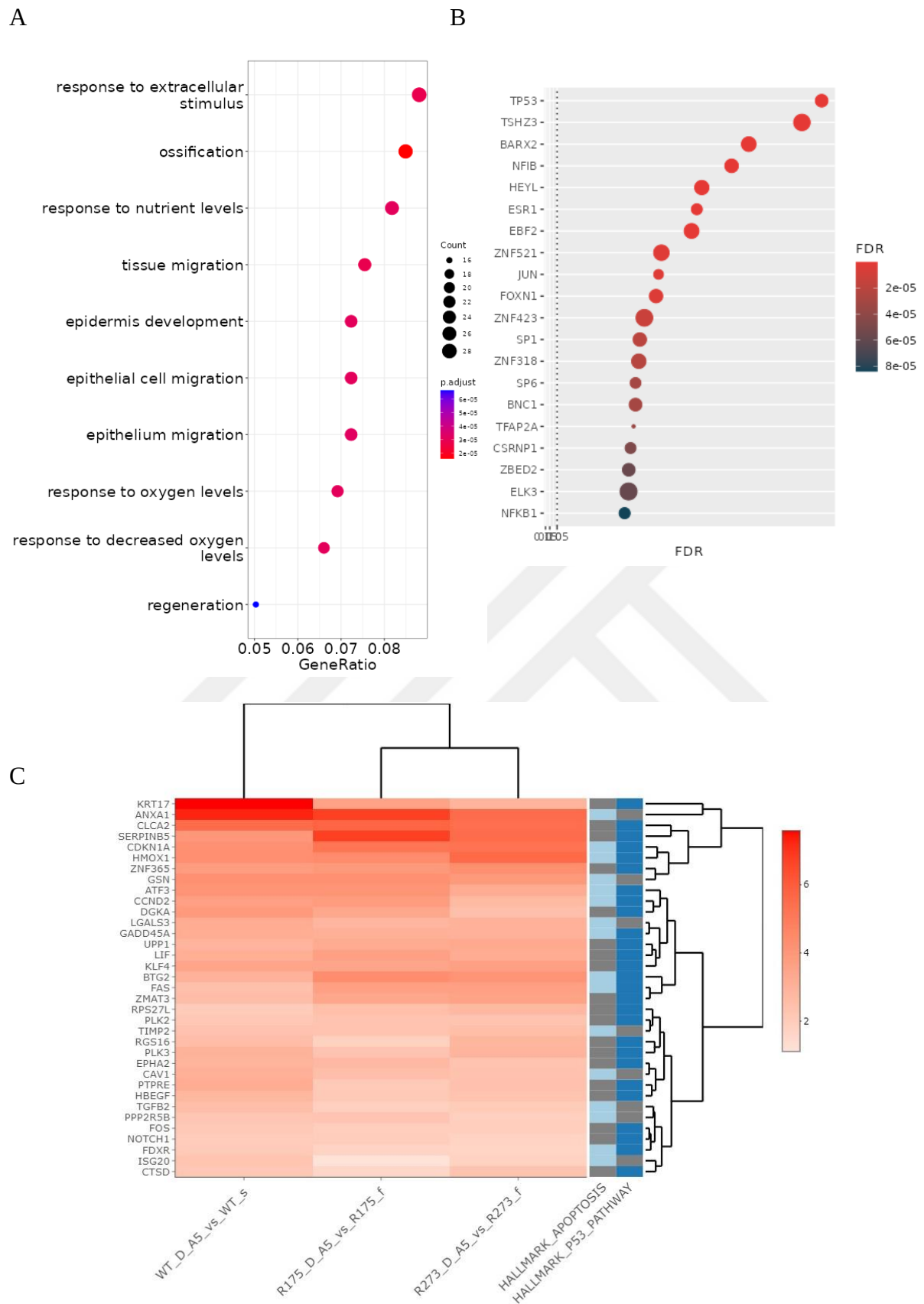


Figure 4.54 **A** GO enrichment terms for the upregulated genes in the WT combinatorial treatment. **B** TFE for these upregulated genes. **C** Heatmap of the p53 and apoptosis genes.

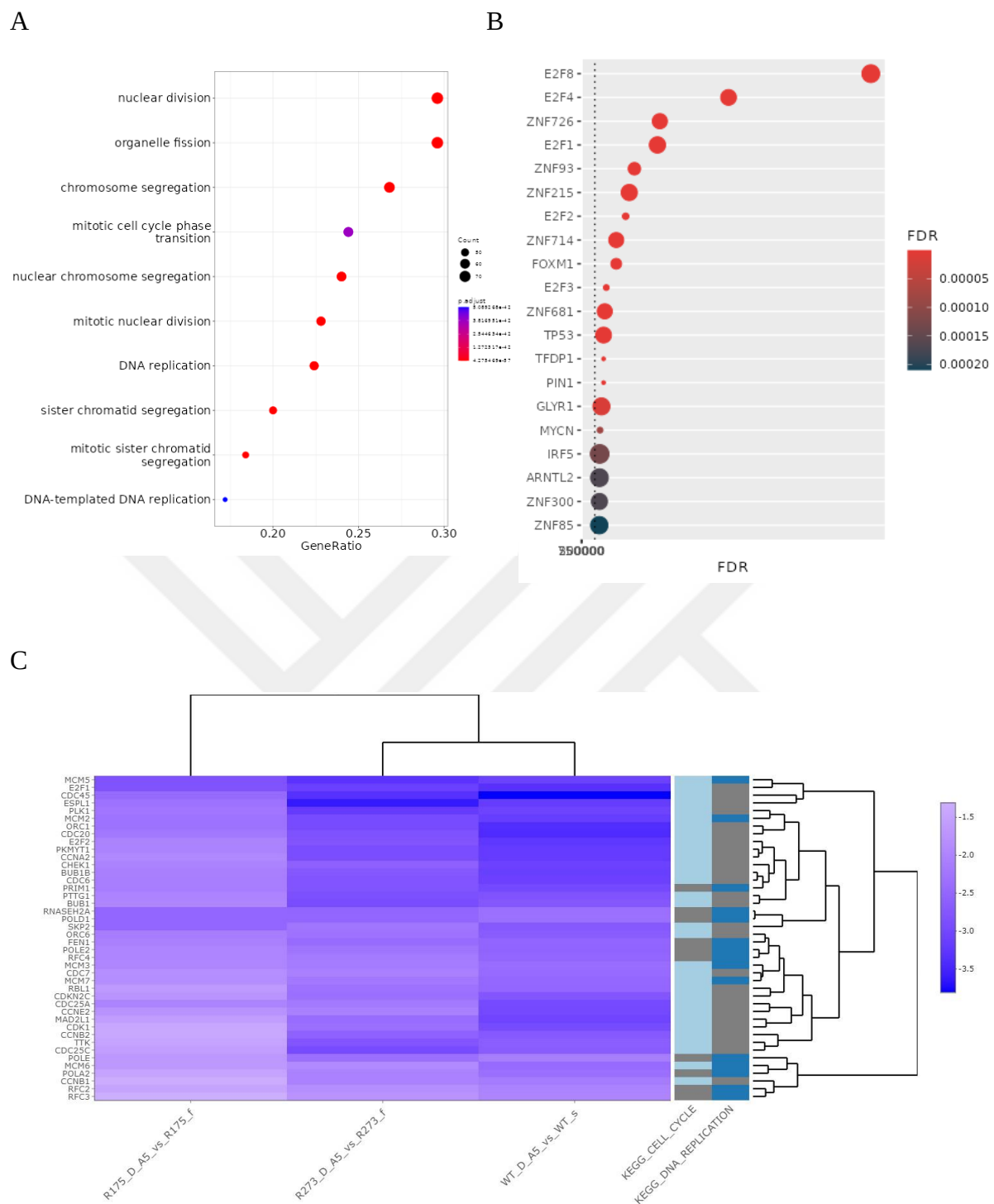


Figure 4.55 **A** GO enrichment terms for the downregulated genes in the WT combinatorial treatment. **B** TFE for these upregulated genes. **C** Heatmap of the KEGG cell cycle and DNA replication terms from the MSigDB database.

For these combinatorial treatments, it was observed that most genes across the contrasts had a similar pattern. The general trend for the treatment was that downregulated genes were more prominent in the WT compared to the R175H and R273H. This trend was also prominent for some p53 targets based on the KEGG database.

4.2.2.7 Cell line status specific signature

To observe genes which could be specific to the overexpression of WT *TP53* in the MCF7 cell line, I filtered for genes that are significantly modulated in wild type cells treated with the combinatorial (*WT_Doxorubicin_siCHRNA5* vs *WT*), as well as in *Nutlin* vs *control* (GSE47042), but not significantly affected in the mutant R273H and R175H cell lines under the same treatments. This approach helps in identifying genes that are specifically responsive in the wild type context but not in cells with *TP53* hotspot mutations.

I filtered the p-values as less than 0.01 for the WT and the nutlin contrast while at the same time, set the filter as above 0.1 for the mutant cell lines, treated with the combinatorial. This showed 126 genes, which is considered as the WT overexpressed signature. *ERBB2* (*HER2*) and four of the *ERBB2* signaling pathway genes, *AGR2*, *BCAR3*, *HIP1R*, and *PTK6* showed up. The *ERBB2* pathway is crucial in breast cancer playing distinct roles in tumorigenesis and cancer progression.

Another important gene is *PARP14*, and its downregulation based on the combinatorial treatment in the WT cells. In the context of cancer, *PARP14* has been implicated in the regulation of macrophage activation and is associated with the survival and proliferation of cancer cells, especially under metabolic stress conditions (Iansante et al., 2015).

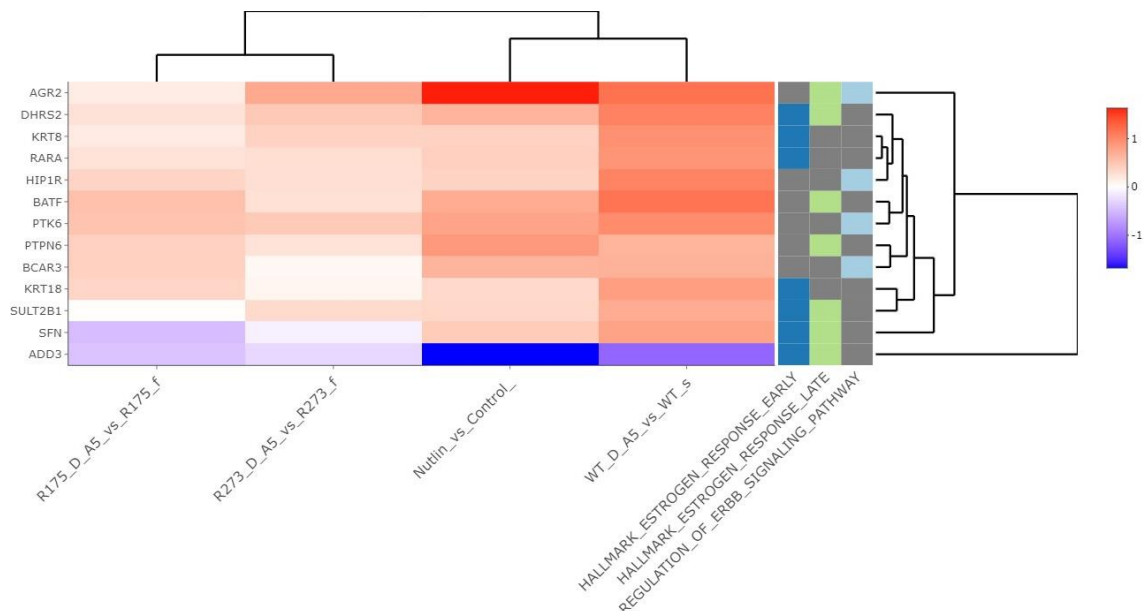


Figure 4.56 Subset of the 126 genes, described as WT specific signature.

Furthermore, as throughout the WT contrast analysis, estrogen response genes also showed overlap.

Similarly, to check for genes overlapping in the mutant cell lines, I filtered significant genes ($p\text{-value} < 0.01$) in both the combinatorial treated mutant cell lines, while I set a filter of $p\text{-value}$ greater than 0.1 for the WT and the nutlin contrast. This filtered 90 genes, specific to both mutations. *AKR1C2* is one interesting gene, which is however upregulated in the mutant signature. It could possibly activate the PI3K/AKT pathway, as seen in oesophageal squamous cell carcinoma by operating as an oncogene (Zhang et al., 2022). This upregulation could potentially cause resistance to the combinatorial treatments as well in these p53 mutant cells. Important downregulated genes in this mutant signature include, *AAMDC*, *CCNC* and *NCOR1*. Roles and implications of these genes are described in the discussion section of this thesis.

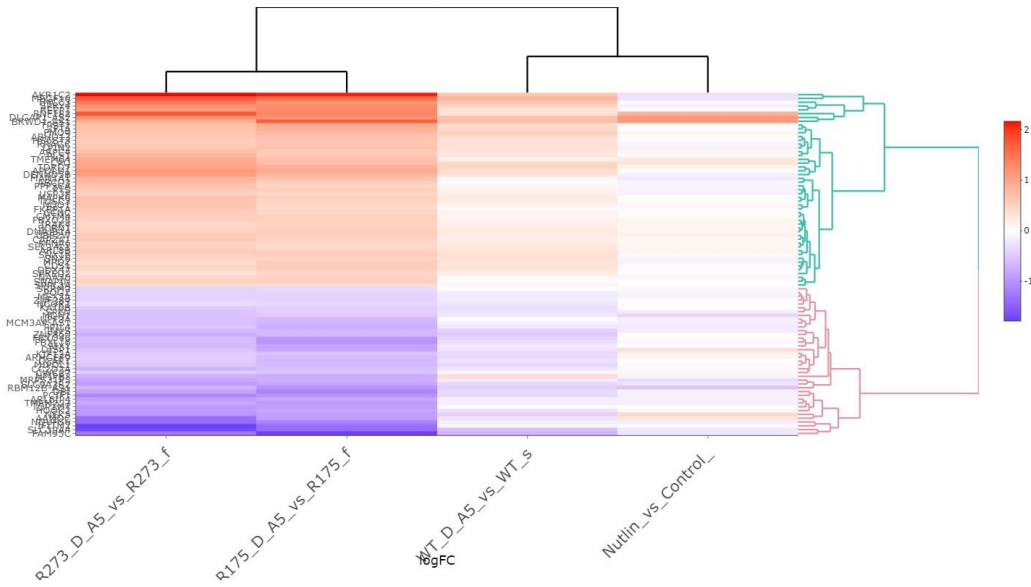


Figure 4.57 Heatmap for the 90 genes, described as the mutant cell line signature.

4.2.2.8 Confirmation of our mutant cell line with online datasets

Since multiple datasets have been incorporated in the TP53-META application, I compared our mutant cell lines with other experiments. The first comparison was of si*CHRNA5* in MCF7 cell line with that of Nutlin. The scatter plot in figure plot shows a linear relationship between the two as expected, with an R^2 value of 0.388.

Next, I compared the WT cell line from the second batch of the doxorubicin data with the GSE47042 Nutlin induced data. I used WT for this comparisons because the GSE47042 used MCF7 cell line with no isogenic mutants. The combinatorial treatment was plotted against Nutlin to see the general pattern and a linear trend was observed here. I then selected the LIN37 dependent *TP53* targets and filtered the meta p-value for 0.01. For these targets the nutlin induced dataset and the combinatorial contrast are clustered together in the heatmap with very similar expression patterns. si*CHRNA5* alone has a very low expression profile for these targets, which is clustered together with Doxorubicin.

To support the claim that LIN37 targets are regulated based on the si*CHRNA5* and the combinatorial treatments, I also plotted si*TP53* vs Control from Data 1 of our lab. Since, *TP53*

is silenced here, the LIN37 targets are upregulated or have no expression. This dysregulated the cell cycle as LIN37 is an integral component of the DREAM complex. Gene targets from other components, such as RB-E2F targets or MMB-FOXM1 targets also showed dysregulation. However, the LIN37 targets showed an overwhelming response based on these treatments.

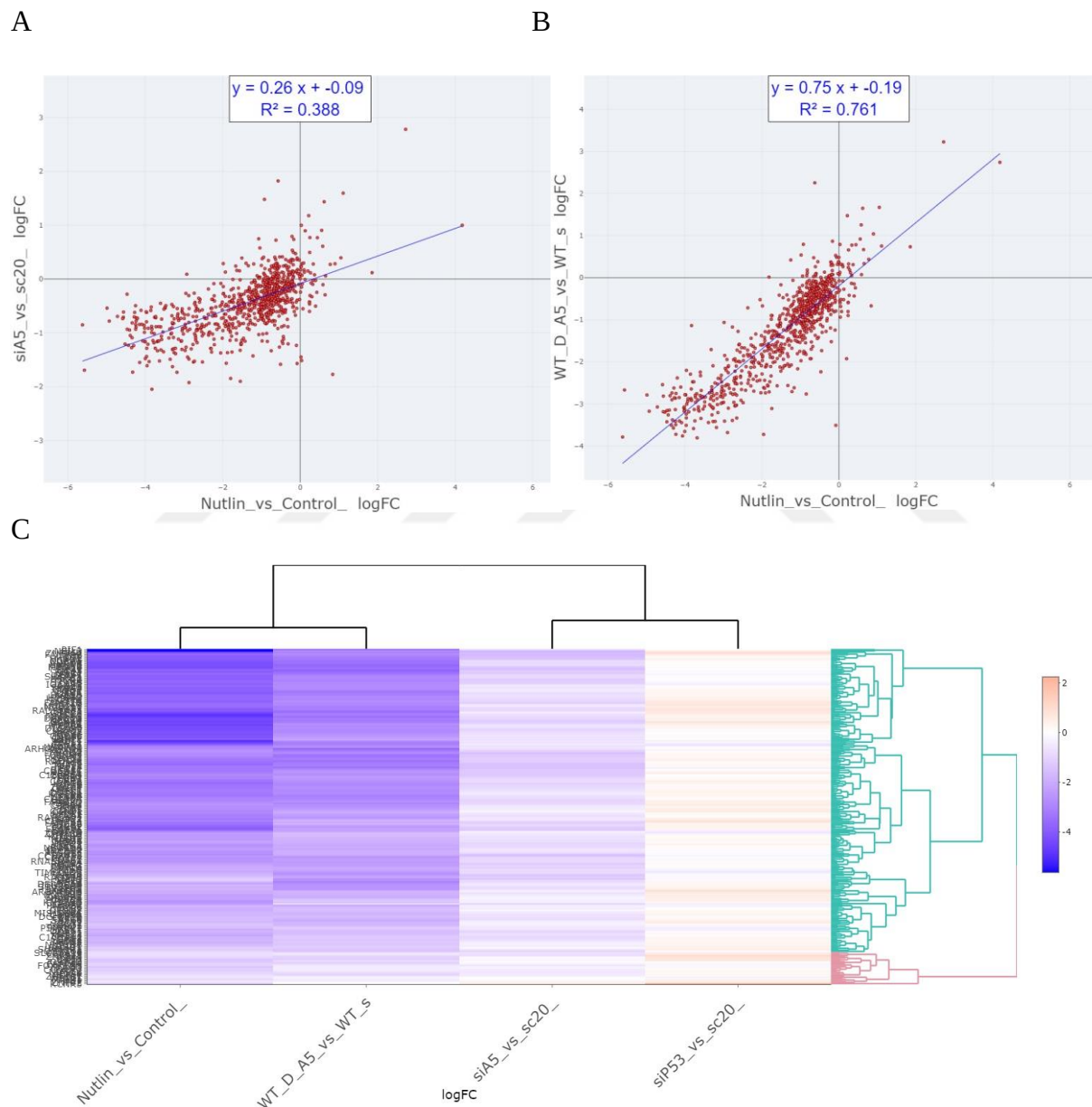


Figure 4.58 **A** Scatter plot of DREAM targets plotted for the Nutlin treatment against the siCHRNA5 (on the y-axis). **B** Scatter plot of DREAM targets plotted for the combinatorial treatment against the siCHRNA5 (on the y-axis). **C** LIN37 targets plotted as a heatmap for three different experiments.

4.3 Synergy Method Calculations

Previously, a synergy method was established in our lab by Dr. Ayse Gokce Keskus (2021). However, some genes appeared to be misrepresented by that method. Therefore, to optimize the synergy calculation, I tried to implement different methods for synergy calculation between treatments. All of these methods were then checked by a randomforest algorithm which predicts the annotation of the genes based on their p-values, logFCs and Synergy scores. Eighty percent of the data was used for the training and from the same dataset twenty percent was used to test and predict.

The first method is the introduction of a sum of the counts of the treatment groups. Since the treatment groups have replicates and to keep the biological variance intact, the columns to sum are chosen randomly. The combinatorial is then compared with this expected column to get a synergy score.

$$\text{Synergy_Score} = \text{Combinatorial vs Expected}$$

This gives us a synergy logFC and a synergy P-value which can then be used to annotate the genes. We categorized a gene as NS (non-significant) when its adjusted p-value fails to reach significance across all four primary contrasts: two individual treatments, their combinatorial effect, and the synergy score. A synergistic gene is classified if it demonstrates significant adjusted p-values for the combinatorial treatment and the synergy score, along with a positive synergy score that is more than the positive logFC of the combinatorial treatment. Similarly, a gene is annotated as antagonistic if it shows a significant p-value for the combinatorial and the synergy score, but the synergy score has a magnitude in the other direction from the combinatorial. However, adding counts did not prove to be that successful due to the variation and the random selection of columns.

Since we were still getting misrepresented genes, I introduced a scaling factor to the raw counts of the treatment groups. For each gene in treatment group 1, I calculate the ratio of

its counts to the sum of counts from treatment groups 1 and 2. This scaling factor represents the proportion of expression attributable to treatment group 1 alone. I repeat the same step for treatment 2 (1 - scaling factor). I then create two new treatment groups, by multiplying the scaling factor I calculated to the combination treatment. This step adjusts the expression levels of the "Combined" treatment group, accounting for the contributions of both treatment groups 1 and 2.

$$SF = siCHRNA5 / siCHRNA5 + siTP53$$

$$SF = siTP53 / siCHRNA5 + siTP53 (1 - SF)$$

$$SFCHRNA5 = combinatorial * SF$$

$$SFTP53 = combinatorial * (1-SF)$$

Contrasts:

SFCHRNA5 vs combinatorial

SFTP53 vs combinatorial

Using limma's statistical framework, I make the contrasts based on these new treatment groups; *Scaled_factor_treatmet vs combination*. A key feature of our approach is that if the logFC is greater than zero for a particular gene, it indicates that the gene is upregulated in the scaled factor treatment compared to the combination. This logFC threshold serves as a measure of synergistic expression, implying that the gene's response is enhanced when both treatments are administered concurrently. Similarly, if the logFC is less than zero, that particular gene is defined as antagonistic, meaning the combined effect of both the treatments is less than expected, than the sum of their effect independently.

To account for variability in the data I use the p values, and if the p value is significant, i.e less than 0.05 we consider our results as significant. However, if the p-value is not

significant, I annotate the genes as insignificant. Furthermore annotating the genes as additive shows that the p value was not significant in the scaled factor contrasts but was significant when we compared the individual treatment with the control. This method proved to be the worst in case of performance with misrepresenting the most number of genes. This misrepresentation is also due to the presence of two contrasts and their combination in the end to come with an annotation.

Since adding the row counts or manipulating them appeared to hinder the results I decided to let the limma statistical framework handle this. In my contrast design matrix, for the synergy contrast I used the following equations separately;

Method 3: Synergy = (Combination – Treatment 1) + (Combination – Treatment 2)

Method 4: Synergy = Combination + Control – Treatment 1 – Treatment 2

Table 4-1. Design matrix for the synergy calculation in Method 3.

Levels	Treatment1	Treatment2	Combinatorial	synergy
Combinatorial	0	0	1	2
Control	-1	-1	-1	0
Treatment1	1	0	0	-1
Treatment2	0	1	0	-1

Table 4-2. Design matrix for the synergy calculation in Method 4.

Levels	Treatment1	Treatment2	Combinatorial	synergy
Combinatorial	0	0	1	1
Control	-1	-1	-1	1
Treatment1	1	0	0	-1
Treatment2	0	1	0	-1

The annotation for both of these methods is based on P-values and not p-adj values. However, for method 3, the number of annotated genes as Synergistic and Antagonistic was the lowest amongst all of these methods (Table 4-3). When compared amongst each other, these two methods provided the least misrepresentations based on the features used for annotation. The features used in training the randomforest model, included the p-values for the combinatorial and their synergy method along with their respective logFCs. The model was trained on our si*CHRNA5* + si*TP53* data. The training and testing data were split 70 to 30.

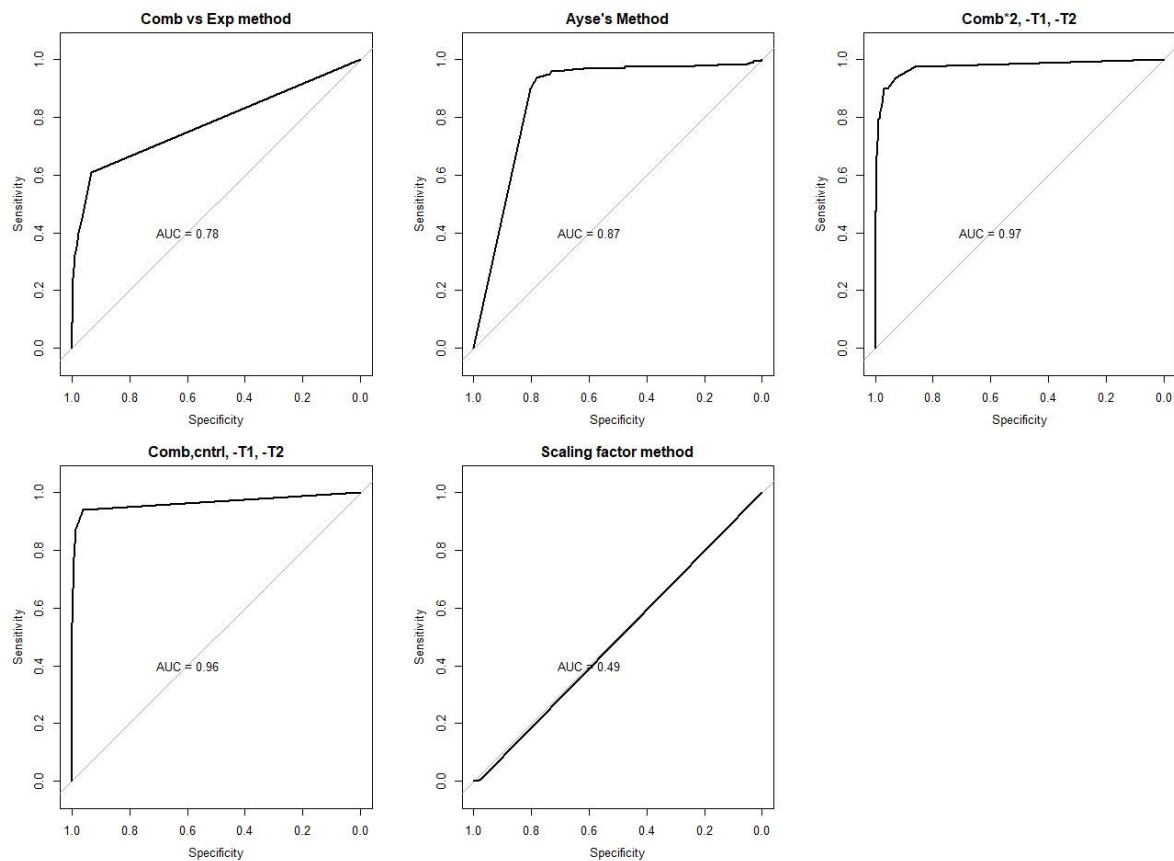


Figure 4.59 AUC curve for the prediction of the model. Method 3 and method 4 gave an AUC of over 0.95, which could mean that this model is overfitting.

Table 4-3. Number of genes for each of the annotations for these methods.

Method	Synergistic	Antagonistic	Additive	NS	other
Method 1 (comb vs exp)	1783	1589	5422	6884	
Method 2 (scaling factor)	1092	1397	6311	7797	
Method 3	320	563	6444	8358	
Method 4 (Comb + cntrl, -T1, -T2)	2651	288	4304	8442	
Ayşe's Method	1834	1476	454	10,128	1786 (Single treatment)

Since the biological variance amongst data makes it difficult to annotate a gene just on the combinatorial logFC and the synergy calculations, we decided to develop a method with 7 different contrasts and assigning a binary value to each to decide if there is a pattern amongst them.

Treatment 1 vs Control

Treatment 2 vs Control

Combinatorial vs Control

Combinatorial vs Treatment 1

Combinatorial vs Treatment 2

Treatment 1 vs Treatment 2

Synergy : Combinatorial + control – Treatment 1 – Treatment 2

This allowed us to look at each cluster separately, based on these contrasts. However, this proved to be more difficult than the previously designed annotations. For these seven contrasts, there were 128 possibilities in total and 100 were present in our dataset. By calculating the pairwise correlation between the genes, within each binary cluster, we calculate the similarity and cluster the binary combinations together based on the genes p-value and

logfc. This is then mapped by UMAP and clustered as shown in Figure 4.58. Table 4-4 shows the binary combinations and where they are clustered. For example, cluster 1 could be considered insignificant genes, since the binary combinations there contain the least significant combinations. However, defining synergistic or antagonistic based on these clusters did not appear to be the correct method.

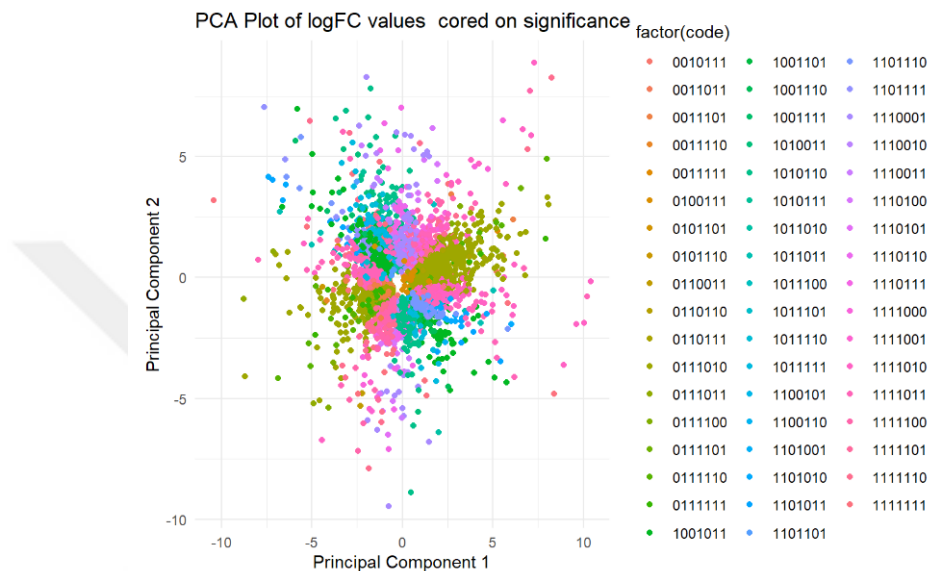


Figure 4.60 PCA of the genes based on their logFC, annotated by their cluster code.

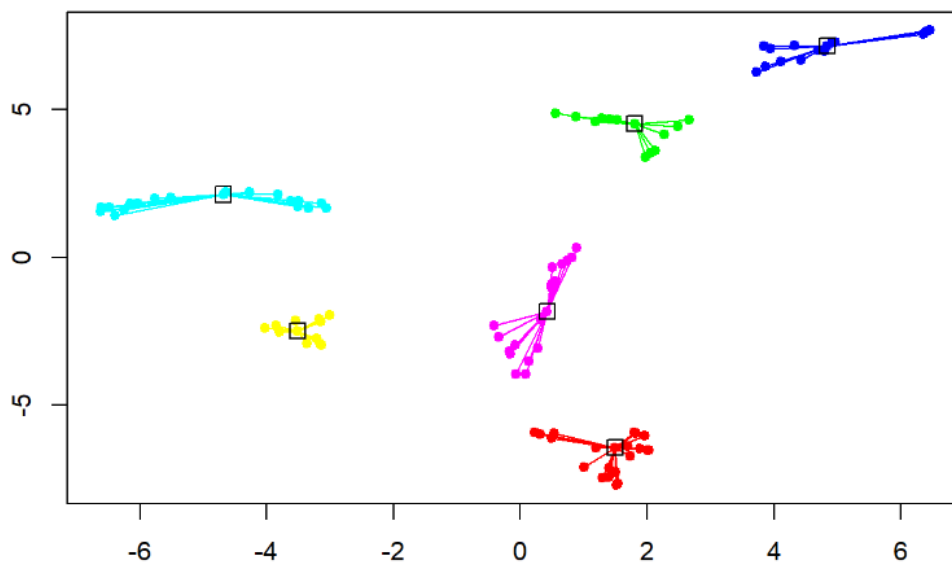


Figure 4.61 UMAP clustering of binary codes to determine the annotations.

Table 4-4. Clustering of the binary codes.

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
0000000	1111000	1100000	1001000	0100100	1010010
1000000	0000100	1110010	1011100	0001100	1101010
0100000	0010100	1011010	0111100	0110110	0000110
0010000	1010100	1010001	1100110	0011110	1000110
1010000	1110100	1111001	0010110	0110001	0100110
0110000	0011100	1110101	0101110	1101001	1110110
1110000	1111100	1101101	0000001	0011001	1001110
0001000	0110010	1111101	0100001	1100101	1101110
0011000	0011010	1000011	1100001	0010101	1011110
0111000	0111110	0110011	0001001	1001101	1000001
0000010	1111110	1010111	1001001	0101101	1110001
1000010	1011011	0110111	0000101	1011101	1010011
0100010		0011111	0100101	0111101	1110011
0001010			0001101	0100011	1101011
1001010			0011101	0001011	1110111
0101010			0011011	0010111	1001111
0111010			0111011		1101111
1111010			0100111		1011111
1010110			0111111		1111111
1001011					
1111011					

5. Conclusion and Discussion

The technological revolution in genomics, characterized by the development and prevalence of RNA-seq technology, has made researchers explore transcriptomics with unparalleled resolution. RNA-seq has been recently used at a very large scale to discover new pathways, novel transcripts and uncover complex gene expression patterns (Z. Wang et al., 2009). The emergence of next-generation sequencing platforms made it possible to obtain vast RNA-seq datasets, leading to the development of numerous computational tools for analysis. These tools have recently evolved from basic analysis tools, to post-transcriptional modification analysis, network analysis and synergy prediction through machine learning (Rahman et al., 2022). The significance of these tools and technologies becomes even more important in the context of cancer research. This is because understanding the intricacies of cancer and its nature is very crucial for the development of targeted therapies. Among the various genes implicated in cancer, *TP53* stands out due to its role as a tumor suppressor and its frequent mutation in a myriad of cancers (Kandoth et al., 2013).

5.1 TP53 META Web Server

In this thesis, I have discussed the development and use of a web application based on R and R shiny, that can be used as a database and for the analysis, functional enrichment and visualization of RNA-seq datasets related to *TP53*. The TP53-META application allows a) for the simultaneous analysis of up to 4 contrasts using either Limma or DEseq2; b) their meta analysis; c) filtering using user defined thresholds or gene lists; d) heatmap visualizations across contrasts/treatments; e) functional enrichment of these contrasts; f) transcription factor enrichment analysis and g) visualization of the data as bipartite gene-treatment regulatory networks.

The TP53-META provides an easy to use interface for the user and easy to follow steps for the analysis. It integrates various RNA-seq analysis steps and makes it easier for the user to analyze the end result in a seamless manner. Its differential expression, filtering by cell regulatory gene lists, and TF enrichment provide a particular focus on *TP53*, which is not present before. The main purpose of developing shiny apps, with a focus on RNA-seq datasets and particularly *TP53*, is to address the critical gap in cancer studies and allowing users who are less familiar with programming to use these functions and tools easily, with reproducible results.

Among the various RNA-seq applications present currently, this application allows the user to perform a meta-analysis of various datasets simultaneously, with a focus on *TP53*, which is one of the most researched gene in the context of cancer genomics. This web application allows to analyze complex in-house datasets with multiple samples and different contrasts which are difficult to analyze one by one. This is also possible with other datasets where the comparisons are done within multiple treatment groups and a single control group is not defined. When surveying the literature, I could not find any webserver that allows the combination of multiple contrasts from different studies.

5.2 Batch effect correction for in-house datasets

The TP53-META application contains our in-house datasets with and without batch effect correction. The reason for using Combat-seq for these datasets is the package's ability to adjust data and still have it as integer values. This allows easy usage for DGE analysis for example in DEseq2, which only works with integer values. However, batch correction was also performed outside the application as an integrated equation in Limma. This method appears to be more reliable as the raw counts are not changed; rather the data is normalized accordingly with the input weights and batches. The clustering for the mutant cell lines without the treatment, after

batch effect correction appeared to be very skewed in one direction (Figure 4.16). Similarly, the doxorubicin treated datasets also clustered separately for the two batches with a big difference in PC1 (Figure 4.23). Even though the R175H sample from the second batch clustered with its replicates from the first batch, the results still appeared skewed, with less variation between the samples in the first batch.

One reason why batch effect correction may not have been a correct solution for these datasets is because our experimental groups across batches do not share identical combinations. This leads to significant analytical problems. This is largely because most methods of batch effect correction assume similar group composition of batches for proper adjustment of the data. When group compositions vary, for example, a few treatment groups within some batches and more of them within others, appropriate correction methods for the modeling of the underlying batch effects cannot be effectively achieved without distortion (Leek et al., 2012).

If the design of an experiment confounds with unevenly assigned batches to different groups of the study, batch effect correction is reduced in efficiency and reliability. Such batch effects will cause a variation in the group distribution (Yu et al., 2023). Therefore, application of correction methods like ComBat, designed to adjust data based on mean and variance estimates from batch-specific distributions, become less effective when batches do not share a uniform group structure. These discrepancies can result in over or under-correction, where the statistical properties of batches are forced to align by probably comprising actual biological variability. Therefore, throughout my comparisons, I used results from a single batch instead of batch corrected results.

5.3 Combinatorial effects of doxorubicin and si*CHRNA5*

This thesis also focuses on the combinatorial treatment of silencing *CHRNA5* along with doxorubicin as a probable treatment option for breast cancer. The *CHRNA5* gene plays a critical

role in the induction of the *TP53* gene. Numerous biological processes, including apoptosis, cell cycle arrest, and the DNA damage response, are impacted by its downregulation. Research indicates that depletion of *CHRNA5* expression in the MCF7 cell line influences the regulation of genes targeted by the *TP53* signaling pathways. This response is similar to the effect observed with *TP53* inducers like nutlin-3a, pointing to a significant role of *CHRNA5* in modulating *TP53* activity (Koker et al., 2018; Shehwana et al., 2021).

Silencing *CHRNA5* in MCF7 and its proposed effect on *TP53* allowed us to focus on p53 hotspot mutations such as R175H and R273H. The combination of doxorubicin and si*CHRNA5* as inducers of p53, is a novel strategy to unravel the effects of these hotspot mutations. I especially observed the cell regulatory pathways – DREAM complex and its components – to decipher the probable mechanisms by which silencing *CHRNA5* can help recover the lost function of *TP53*, especially in a heterogeneous isogenic mutant MCF7 cell line.

5.3.1 Transcriptomic regulation of the isogenic Mutant cell lines

The mutant cell line included the three most prevalent hotspot mutations in *TP53*; R175H, R273H and R248W. The clinical implications for these mutations are discussed in chapter 1 of this thesis. We sent these cell lines for RNA-sequencing and compared their transcriptomic results. The p-value filtered results showed similarity between the R175H and R273H cell lines compared to WT. However, R248W also had a distinct pattern for the filtered genes. I also compared the batch effect corrected results and the first batch results. The logFC of the batch effect corrected results showed smaller variation, which could mean that the ComBat method may have removed some biological variance also. However, there were some genes which had inflated logFCs as well.

For the R175H cell line, *EGR1* was significantly regulated. *EGR1* defines a major disruption of pathways critical for genomic stability and an important disruption of pathways critical to orderly regulation of cell proliferation. However, *EGR1* has been demonstrated to be crucial in both directly inducing apoptosis and, concurrently, activating DNA repair activities as a result of growth factors (B. Wang et al., 2021). The downregulation of *EGR1* hints at a potential mechanism through which the R175H mutation could contribute to tumorigenesis through the impairment of tumor suppressive pathways (Saha et al., 2021). The gene-disease association database (DisGeNET) showed “early stage breast carcinoma” as one of the implicated diseases for these filtered genes in the R175H mutant cell line. Other significantly regulated genes included *TGFB3* and *FGFR3*, both of which are growth factors that regulate cell growth and differentiation. The enrichment analysis linking these genes with terms like “early stage breast carcinoma” as per the DisGeNET database indicates a possible early involvement in cancer development. This association could be indicative of how the R175H mutation influences cancer progression through modifications and alterations in growth factor pathways, crucial in the regulation of cell cycle and apoptosis.

When filtered for the genes significantly dysregulated in R273H cell line, it was observed that R273H cell line was clustered separately compared to the WT and R175H. The differential expression of genes related to steroid response, epithelial and epidermal cell differentiation highlights the R273H mutation's impact on cellular signaling pathways and differentiation processes. The R273H mutation is known to alter the tumor suppressor function of p53, leading to changes in gene expression that may contribute to tumor progression and resistance to apoptosis (Freed-Pastor & Prives, 2012). The upregulation of transcription factors like *EGR1* and *FOS* in the R273H mutant compared to their downregulation in the R175H mutant could suggest a mutation-specific effect on gene regulation, which could influence cell survival and proliferation pathways in a different manner.

Moreover, the enrichment of *MYC* targets in the R273H mutation highlights the role of *MYC* in cancer biology, which is often linked with increased cell cycle progression and tumorigenesis. The upregulation of *MYC* could further amplify the oncogenic potential of the R273H mutation by promoting cellular proliferation (Dang, 2012).

In contrast to these mutant cell lines, the results from the WT cell line showed the E2F targets and estrogen targets as enriched terms (Figure 4.21). In wild-type (WT) cells, estrogen signaling pathways are notably active and possibly linked to the unaltered regulatory function of the *TP53* gene. In its unmutated state, *TP53* regulates genes involved in cell cycle arrest, DNA repair, and apoptosis, which contributes to a balanced cellular response to estrogen. Estrogen binding to its receptor, initiates a cascade of gene activations that typically promote cellular health and proliferation within a controlled framework (Heldring et al., 2007). This balance is crucial as estrogen also contributes to the pathogenesis of breast cancer, where aberrant estrogen signaling can lead to unchecked cell growth (Frasor et al., 2003). In WT, E2F target genes, which are crucial for advancing the cell cycle, are more suppressed compared to the mutant cell lines (Figure 4.21), demonstrating tight control of cell proliferation under normal p53 conditions. This regulation of the E2F targets may act as a protective measure against uncontrolled cell growth, a hallmark of cancerous tissues (Kent & Leone, 2019).

For the confirmation of the cell lines, I also filtered the DREAM targets, LIN37 and the cell regulatory pathway genes. These genes were less regulated in the mutant cell lines compared to the WT (Figure 4.22), which confirmed the disruption of their cell regulatory pathways. This confirmation led to the treatment of the cell lines with doxorubicin and si*CHRNA5* to see their effects alone and in combination and whether this combinatorial treatment is a viable option to target these p53 hotspot mutations.

5.3.2 Transcriptomic regulation of the combinatorial treatments in the mutant cell lines

The comparison of WT, R175H and R273H and their respective combinatorial treatment with their controls, expressed the activation of p53 target genes. The EV combinatorial was removed from the analysis, since IGV analysis suggested that it does not belong to the EV cMCF cell line. Downstream analysis of differential gene expression within the cell lines treated by the combinatorial treatment indicated important changes in the gene expression associated with several biological processes and pathways. For example, an upregulation of *CDKN1A* (*p21*) along with *BAX* in response to the combinatorial treatment across the mutant lines is a reactivation of the p53-dependent pathways, which are one of the key mediators of cellular responses to DNA damage, including apoptosis and cell cycle arrest (Aubrey et al., 2018). p21 is also implicated as a cyclin-dependent kinase inhibitor in cell cycle progression, suggesting its possible compensatory upregulation in a loss-of-function state of *TP53*. In circumstances of DNA damage, p21 works as a crucial mediator in the event of DNA damage, stopping the cell cycle to enable DNA repair while acting under the supervision of tumor suppressor genes (Karimian et al., 2016).

In mutated *TP53*-containing cancer cells, the p21 pathway might become very important since it may make up for the loss of normal p53 activity. The upregulation of p21 is induced in response to the combinatorial (Doxo + si*CHRNA5*) treatments aiming at the reactivation of p53-dependent pathways. This highlights a mechanism in which some sort of control over the cell cycle might be re-established in p53-deficient environments (Karimian et al., 2016).

Another probable mechanism of action for p21, as a CDK inhibitor, is that it causes the hypo-phosphorylation of the RB (retinoblastoma) protein and its related pocket proteins, p130 and p107. These hypo-phosphorylated proteins lead to the assembly of the DREAM complex together with E2F4 and MuvB components. This formation shuffles components from the

MYB-MuvB (MMB) complex that function canonically as a transcriptional activator to form the transcriptionally repressive DREAM complex (Fischer & Müller, 2017). In this repressive state, DREAM binds specifically to DNA elements, cell cycle-dependent element (CDE) and the cell cycle genes homology region (CHR) promoter site. This binding leads to the suppression of cell cycle-related genes, particularly during the G2/M phase, which leads to cell cycle progression being stopped. The action of the p53-p21-DREAM pathway is critical in this context as it makes sure that cells do not proceed to mitosis without proper checks, especially under conditions of DNA damage or stress (Fischer et al., 2016).

5.3.2.1 R175H mutant

It was intriguing to observe a strong association between the combinatorial treatment and the doxorubicin treatment alone in this mutant cell line. Doxorubicin, which is a chemotherapy agent, works by intercalating with DNA strands and inhibiting topoisomerase II, which leads to DNA damage and cell cycle arrest, primarily affecting proliferative cells (Minotti et al., 2004). The significant elevation of p21 after doxorubicin therapy is consistent with the mechanism of action of the treatment, since in response to DNA damage, p21 can prevent the advancement of the cell cycle (El-Deiry et al., 1993).

Furthermore, the observed additive or synergistic effect of si*CHRNA5* treatment suggests that this combination with doxorubicin could restore some p53 functions in *TP53* mutants or bypass the limitations created by the *TP53* mutation. One contradictory gene that was upregulated in the R175H contrasts was the *HMOX1* gene, which responds to cellular stress stimuli. Research has shown that *HMOX1* could be implicated in helping breast cancer cells develop chemo-resistance against doxorubicin (Zhu et al., 2015). The upregulation of this gene throughout all the contrasts of R175H and R273H, with the combinatorial or single (Doxorubicin and si*CHRNA5* both) raises an important research point. Even though this could

be a cell-line dependent effect, since *HMOX1* is also significantly upregulated in the WT vs EV contrast (Data 2), the slight upregulation with doxorubicin and the combinatorial, may be explored further.

In the R175H cell line, doxorubicin therapy led to the downregulation of genes, including *MYC*, *MCM2*, and *E2F8*, which are crucial for DNA replication and cell cycle progression. *MYC* is a well-known proto-oncogene whose overexpression can promote the proliferation of cancer cells by enhancing the transcription of growth-related genes. The downregulation of *MYC* by doxorubicin suggests an interference with this proliferative signal, which is a therapeutic goal in cancer treatment (Dang, 2012). *MCM2* is essential in initiation and elongation phases of DNA replication (Aparicio et al., 2006), while *E2F8* belongs to the E2F TF family, which is critical for controlling the expression of genes necessary for the advancement of the cell cycle. (Chen et al., 2009). Downregulation of both of these genes could potentially reduce the proliferation of cancer cells. Interestingly, genes like *HRAS*, a well-established oncogene, are differentially regulated by si*CHRNA5* as opposed to doxorubicin or the combination. This highlights the specificity of the si*CHRNA5* effect in the mutant R175H cell line, potentially offering a targeted approach to modulate oncogenic signaling pathways in cancer therapy (Baines et al., 2011).

The DREAM complex, acting through LIN37 and other components, serves as a gatekeeper for cell cycle transition from G0/G1 to S phase, predominantly repressing E2F-dependent transcription during quiescence (Litovchick et al., 2007). The synergistic or additive effects of both the treatments reflect the effectiveness of si*CHRNA5* in inducing back the p53 function. Genes like *AURKA* and *AURKB*, involved in mitotic spindle assembly, and *CENPL*, a centromere protein, highlight p53's role in genomic stability (Fischer et al., 2016). The differential regulation of these genes suggests a restoration of functionality in *TP53*'s regulatory network, offering significant therapeutic potential. Moreover, genes such as *TOP2A* and

CDCA8, which are involved in DNA topology and cell cycle progression, respectively, exhibit higher expression in combination treatments compared to individual drug treatments. The observed synergy in the treatment-gene network, particularly in genes significant only in the combinatorial treatment, indicates a therapeutic possibility, where the simultaneous suppression of LIN37 targets by doxorubicin and si*CHRNA5*, despite the presence of the R175H mutation, can be exploited.

5.3.2.2 R273H mutant

The complexity of how p53 modulates cellular processes becomes particularly evident when considering the R273H mutation. p53, through its regulation of targets such as p21 and its effects on the p53 pathway, plays an important role in maintaining genomic stability. Upregulation of p21 in the combinatorial treatment of the R273H cell line is consistent with its role as a CDK inhibitor and a key mediator of p53's tumor-suppressing activity, providing a vital checkpoint in cell cycle progression in response to genotoxic stress (Kastenhuber & Lowe, 2017).

Since there was no single treatment for doxorubicin in the R273H mutant cell line, synergism between the treatments could not be established. In contrast to the single si*CHRNA5* treatment, a greater degree of gene regulation was noted with the combinatorial approach. Genes that exhibited higher expression in the combinatorial treatment compared to si*CHRNA5* alone, such as *TGFA* and *PPARG*, highlight the effectiveness of the combination treatment. These genes are involved in cellular signaling and apoptosis, and their expression levels might reflect the efficacy of the treatment regimen in correcting the imbalance caused by the R273H mutation (D.-H. Li et al., 2023).

One very important result was the downregulation of *DRD1*, which is a dopamine receptor. For the combinatorial treatment, the effect was less compared to the single si*CHRNA5*

treatment in the R273H cell line. Recently it has been established that using *DRD1* antagonists, promotes apoptosis in triple-negative breast cancer cells (Rosas-Cruz et al., 2022). This finding highlights the potential of targeting dopaminergic pathways as part of a therapeutic strategy in breast cancer, particularly for mutations that lead to conventional treatments being less effective.

5.3.2.3 Combinatorial effects across cell lines

In the exploration of combinatorial treatments on various *TP53* mutant cell lines, my analysis showed significant insights into the potential of targeted therapy in restoring p53 functionality. The downregulation of key regulatory genes, such as *CDKN1A* (*p21*) and *BAX*, in the combinatorial treatment cell lines suggests an involvement of p53-dependent apoptotic pathways. *CDKN1A*, a critical CDK inhibitor, plays a crucial role in cell cycle arrest and is directly transcriptionally activated by *TP53*, serving as a cellular stress sensor (Karimian et al., 2016).

These results also emphasize the dysregulation of the cell cycle and DNA repair pathways, with genes implicated in "mitotic cell phase transition" and "chromosome segregation" particularly downregulated in response to the combinatorial treatment. This effect aligns with the previously reported mechanisms of doxorubicin's action, such as DNA intercalation and inhibition of topoisomerase-II-mediated DNA repair, which hinder cell proliferation (Minotti et al., 2004). The enrichment analysis of upregulated genes in R175H and R273H upon combinatorial treatment highlights pathways like hypoxia and p53 class mediator signal transduction, crucial for mediating cellular responses to DNA damage and maintaining genomic stability. These findings necessitate more research into the molecular underpinnings of the interaction of pathways, including estrogen response and hypoxia, which are linked to cancer progression and therapy resistance (Muz et al., 2015).

5.3.2.4 Difference in response of the mutant cell lines to the combinatorial treatment

For the upregulated genes, the mutational cell lines upon combinatorial treatment appeared to respond similarly. This was evident from the clustering in the heatmaps. However, the R273H mutant cell line appeared to cluster with the WT, upon combinatorial treatment, for downregulated genes. This difference may arise from the difference in the structural forms of these hotspot mutations. The R273H mutation occurs at a DNA contact point and primarily affects the DNA-binding ability of p53, leading to impaired transcriptional activation of p53 target genes (Garg et al., 2020). In contrast, the R175H mutation, which is a structural mutation within the DNA-binding domain, has more profound effects on the conformation of p53. This mutation leads to a misfolded protein that not only loses its DNA-binding capability but also gains new functions that promote oncogenesis, such as altered interactions with other cellular proteins (Chiang et al., 2021). Studies have shown that R175H can induce higher levels of chromosomal instability compared to other *TP53* mutations and is more likely to disrupt the 3D nuclear architecture (Samassekou et al., 2014). Between the two hotspot mutations, the R175H mutation is often considered more deleterious due to its broader impact on cellular functions, including more pronounced effects on genomic stability and a higher propensity to drive aggressive cancer phenotypes (Samassékou et al., 2014; Chiang et al., 2021). The combinatorial treatment may have therefore have a less significant effect on the R175H mutant cell line for the downregulated genes, in contrast to the R273H mutant cell line, which was clustered with the WT cell line.

5.3.2.5 Cell line status specific genes

The analysis of differentially expressed genes, after the treatment with either doxorubicin and si*CHRNA5* in wild-type (WT) and mutant cell lines reveals a distinct overexpression signature

in WT cells that includes 126 genes, prominently featuring *ERBB2* and key members of its signaling pathway such as *AGR2*, *BCAR3*, *HIP1R*, and *PTK6*. This demonstrates how the *ERBB2* pathway, which affects several aspects of carcinogenesis and cancer progression, is critical in the development of breast cancer. *ERBB2* is widely studied for its overexpression in certain breast cancer subtypes, hence a critical target for therapeutic interventions (Zhu et al., 2024). The presence of *AGR2*, which is significantly expressed in multiple cancers, and *PTK6*, whose high expression has been linked to detrimental outcomes, is linked to cancer metastasis and cell survival mechanisms (Chevet et al., 2013; Gilic and Tyner, 2020). This could indicate adaptability of cancer cells to treatments, which is also not affected by the combinatorial.

Moreover, the downregulation of *PARP14* in the WT cells under combinatorial treatment also suggests its role in regulating macrophage activation and supporting cancer cell survival under metabolic stress (Iansante et al., 2015). This effect, combined with observed overlaps with estrogen response genes, indicates a broad regulatory mechanism that could be important in the understanding of the combinatorial behavior of the depletion of *CHRNA5* along with its use of doxorubicin in cells expressing WT p53.

The upregulation of *AKR1C2* in mutant cell lines signature of 90 genes, is significant, as its enhanced expression might contribute to oncogenic processes by activating the PI3K/AKT pathway, similar to observations in oesophageal squamous cell carcinoma (Zhang et al., 2022). Upregulation may hint at a potential mechanism for resistance to combinatorial treatments in cells harboring *TP53* mutations, underlining the challenges in targeting mutant p53 tumors.

However, the downregulation of *AAMDC*, *CCNC*, and *NCOR1* in these mutant signatures shows promise. Each of these genes has been implicated in critical regulatory pathways affecting cancer cell metabolism, cell cycle regulation, and gene expression. The reduction in *AAMDC* may impact lipid metabolism and signaling pathways that are vital for

cancer cell survival and proliferation since *AAMDC* is amplified in aggressive estrogen receptor-positive breast cancers (Golden et al., 2021). Similarly, *CCNC* is integral to cell cycle control. It was observed to be increased in various cancer types and its downregulation could be explored further as it interacts with CDK8 (Xu et al., 2011).

5.3.2.6 Response to chemo-resistance

The differential impact of doxorubicin alone compared to its combination with si*CHRNA5* on MCF7 isogenic mutant cell lines, shows interesting insights into the potential mechanisms of drug resistance. The RNA-seq results discussed previously in this thesis highlight distinct patterns of gene expression that may suggest alterations in drug responses.

MYC, which is known for promoting cell proliferation and survival, has been linked to drug resistance in various tumors (Donati and Amati, 2022). In our isogenic mutant cell lines, *MYC* was further downregulated in the combinatorial treatment compared to the single treatments. In the context of cancer therapy, downregulation of *MYC* may be a favorable response, as it might increase the effectiveness of the treatment by limiting the cancer cell's ability to proliferate and survive under stress. This could lead to increased sensitivity to chemotherapy, as the cancer cells may become less capable of overcoming the DNA damage induced by the treatment (Donati and Amati, 2022).

Another important contributor to chemo-resistance is the E2F family of TFs. For the R175H mutant cell line, the downregulated genes enriched various E2F transcription factors (Figure 4.50). In the combinatorial treatment, the downregulation of these factors was seen to a higher level compared to the individual treatments. Similarly, for R273H, the E2F factors were also enriched in the combinatorial treatment. Cell division and tumor suppression depend on E2F, which is largely controlled by the tumor suppressor pRB. E2F activity is increased in cancers when pRB function is compromised, which may help the tumors become resistant to

chemotherapy. This is due to the fact that dysregulated E2F activity has the potential to activate tumor suppressor genes that trigger apoptosis or cellular senescence—processes that cancer cells can resist when faced with treatment pressure. Consequently, E2F downregulation may aid in cancer treatment by reducing or perhaps partially reversing chemo-resistance (Nakajima et al., 2023).

However, one crucial finding was the constant enrichment of the “hypoxia” related terms for the R175H mutant cell line with both doxorubicin and the combinatorial. Hypoxia is commonly associated with tumor adaptation to low oxygen conditions, which can promote resistance to therapy, including chemotherapy. This is because hypoxia-inducible factors (HIFs) can drive the expression of survival genes that help cells deal with the stress induced by chemotherapy (Brown and Wilson, 2004). The combinatorial treatment, even though provides benefits in the synergistic regulation of some crucial genes, it cannot fully reverse resistance mechanisms related to hypoxic adaptations.

5.3.2.7 Confirmation of RNA-seq results through MTT assays

To confirm the efficacy of the depletion of *CHRNA5* in MCF7 mutant isogenic cell lines, and its synergistic/additive effect on reduced cell viability, Cagdas Yanyatan from our lab performed various MTT assays. The si*CHRNA5* treatment alone also reduced the cell viability; however, its effect was far more potent in combination with doxorubicin (Appendix A).

For the R175H mutation, the MTT results demonstrate a significant reduction in cell viability when treated with doxorubicin and si*CHRNA5*, compared to doxorubicin treatment alone. This suggests that the depletion of *CHRNA5* may potentiate the effects of doxorubicin, potentially by enhancing p53-mediated cell death pathways that are otherwise impaired in this mutant cell line. However, the distinct pattern of cell viability in the R175H mutation aligns with RNA-seq data showing that the R175H cell line (compared to the R273H and WT) resists

the induction of p53 pathways or alternate pathways even with the depletion of *CHRNA5* in combination with doxorubicin (Appendix A, Figure A4).

The R273H mutation exhibits a synergistic/additive response to the combinatorial treatment, where the addition of si*CHRNA5* to doxorubicin treatment further decreases cell viability. This is consistent with RNA-seq findings that indicated an upregulation of *MYC* and other proliferative genes under R273H mutation when untreated, which are significantly downregulated upon treatment. The effectiveness of the combinatorial treatment in reducing cell viability further validates the role of depleting *CHRNA5*. These findings provide a compelling argument for the translational relevance of *CHRNA5* as a therapeutic target, especially in cancers harbouring *TP53* mutations. Moreover, the results highlight the importance of considering mutation-specific responses in cancer treatment strategies. The differential impact of treatments on R175H and R273H mutations suggests that personalized approaches to cancer therapy, which consider the specific genetic makeup of a tumour, could enhance treatment efficacy and patient outcomes.

5.4 *CHRNA5* as a possible *TP53* target

CHRNA5 has been implicated as a possible p53 target by Fischer and a possible p53 target independent of LIN37 by Uxa et al. (Fischer, 2017; Uxa et al., 2019). However, upon evaluation on p53 cistrome database, *CHRNA5* was not found enriched in the database. Targetgenereg.org did not show *CHRNA5* as a target gene but showed its downregulation in 33 datasets out of 57.

Not surprisingly, in our doxorubicin mutant datasets, *CHRNA5* was also downregulated along with the nutlin treatment from GSE47042 (also part of targetgenereg.org). The mutant R175H cell line has a lower downregulation of this gene compared to WT (Figure 5.1).

CHRNA5 could be part of a broader p53-dependent regulatory network involved in controlling cell proliferation or apoptosis. *CHRNA5* being a target of p53 independently of LIN37 points to a pathway where p53's regulation of *CHRNA5* does not require the involvement of LIN37, which is a component of the DREAM complex, known for its role in cell cycle control and gene expression suppression. This suggests that p53 can modulate *CHRNA5* through a pathway that might be distinct from the classical cell cycle regulation pathways. Both doxorubicin and Nutlin are known to activate *TP53*. Doxorubicin induces DNA damage, leading to p53-mediated cell cycle arrest and apoptosis. Nutlin, on the other hand, inhibits MDM2, a negative regulator of p53, thereby stabilizing and activating p53. If *CHRNA5* is a *TP53* target, its downregulation following *TP53* activation suggests that *CHRNA5* might be involved in negative feedback regulation or that its suppression enhances the pro-apoptotic functions of *TP53*. By silencing *CHRNA5* further, the load on the *TP53* pathway might be reduced, potentially making it more responsive to upstream signals or existing p53 protein in the cell, even if it is partially functional or mutated.

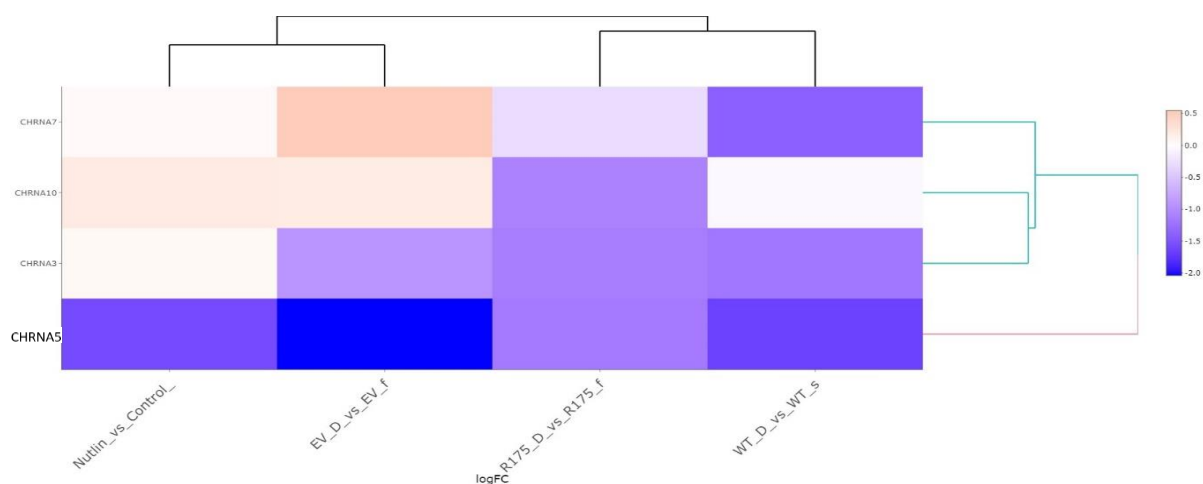


Figure 5.1 Downregulation of the *CHRNA5* gene based on the doxorubicin treatments and the nutlin treatment from GSE47042.

5.5 Synergy Calculations

This thesis builds upon the previously developed synergy method by Dr. Ayse from our lab and aimed to counter the wrong annotation of genes. The method was initially developed to determine synergistic genes from our in house dataset involving si*CHRNA5* and si*TP53*. However, the method extended its use to other datasets as well. To counter the misrepresentation of genes, various methods were explored.

The first re-evaluation method, involved the summation of the individual treatment group counts. However, due to the randomness and variability among replicates, this method did not provide the needed reproducibility. Manipulating raw counts and its comparisons also did not appear to be a correct method because of the doubling of the library size. To counter the inconsistencies of the library size, a scaling method was tried. This scaling factor method adjusts gene expression levels based on their proportional contributions from two distinct treatment groups. This however, involved annotations based on both the separate treatments, and their combination proved to be highly unsuccessful.

One of the approaches involved using Limma's inherent statistical framework, with different factors assigned to the treatment/combinatorial groups, for the synergy score. These methods according to the randomforest benchmark, gave surprisingly high values. This overfitting could be due to the fact that the train and test data was the same and the model learns the intricacies of the data. To have a better understanding and better benchmark of this model, these methods need to be tested on a different dataset, while being trained on another. This can give a more unbiased approach to determine the accuracy of our methods.

5.6 Future perspectives

One of the most important future perspectives in the development of this application is to allow for users to either upload their *TP53* regulated count data to analyze or to integrate as many

datasets relating to *TP53* as possible. This aligns with the fact that this is planned to be a publically available database for everyone to access. Further improvements that could make this application stand out even further include the integration of other NGS datasets such as Chip-seq or scRNA-seq datasets. This would drastically improve the credibility of the results and help users further assimilate with the NGS tools. Another feature which could be added to the webserver would be its ability to directly connect to NCBI GEO and download raw counts for the RNA-seq dataset directly, since only recently NCBI GEO has integrated RNA-seq analysis into GEO as well (Barret et al., 2013).

One other limiting feature of the application is its server capacity. Due to the interactive nature of the application, some analysis have been limited to a certain number of genes. If more than a certain number of genes are chosen, the application takes more time to perform the analysis and render the images. Even though this application is made for the user to go through different analysis side by side, it could be expanded to render some images at the back end and send the results via email. This could be very useful if for example a user wants to visualize a multiple gene lists side by side for these different contrasts.

Doxorubicin's combined effects with si*CHRNA5* offer a strong basis for future investigation into their potential function in preventing the emergence of drug resistance in cancer. The validation of these findings through in vivo studies using animal models with similar *TP53* mutations is critical for determining their physiological relevance and implications in tumor biology. This validation is a crucial step towards confirming the potential therapeutic targets or mechanisms suggested by these *in-vitro* and *in-silico* studies. The nature of *CHRNA5*, being a cholinergic receptor, would prove to be an easy target in-vivo, as its ligand antagonists could be used.

To increase the translational relevance of this research, there is a need to confirm the mutual exclusivity of these p53 hotspot mutations with *CHRNA5* expression and also for WT p53. Hence, there is a need to validate the association observed with clinical data present in TCGA (The Cancer Genome Atlas). By analyzing genomic and transcriptomic data from breast cancer patients, we can assess the prevalence and impact of *TP53* mutations and their correlation with *CHRNA5* expression levels. This will allow us to confirm whether the molecular signatures and pathway activations identified in our MCF7 model is similar to that of the genetic landscape in actual tumors. This comparison will allow us to explore the potential of *CHRNA5* as a therapeutic target, particularly in tumors that harbor WT p53 specific p53 mutations. By understanding the co-occurrence of these genetic alterations in patient samples, we can better understand their combined role in tumor biology and therapy resistance.

Further exploration could be done using time based studies to monitor the efficacy and mechanisms of action of the combinatorial drug treatments over time. Additionally, employing single-cell RNA sequencing could unravel the heterogeneity of cell population responses to treatments, which often remains hidden in bulk analyses. This would allow for a more precise understanding of how different subpopulations within a tumor may respond variably to the same treatment.

Investigating the mechanisms underlying resistance to doxorubicin in *TP53* mutant cells can encourage the development of novel strategies to overcome or circumvent resistance. Using public RNA-seq datasets could provide a platform to identify potential drugs that could be repurposed to target specific pathways altered by *TP53* mutations and could streamline the path to clinical trials and therapeutic development.

6. References

- Afgan, E., Nekrutenko, A., Grüning, B. A., Blankenberg, D., Goecks, J., Schatz, M. C., Ostrovsky, A. E., Mahmoud, A., Lonie, A. J., Syme, A., Fouilloux, A., Bretaudeau, A., Nekrutenko, A., Kumar, A., Eschenlauer, A. C., DeSanto, A. D., Guerler, A., Serrano-Solano, B., Batut, B., ... Briggs, P. J. (2022). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Research*, 50(W1), W345–W351. <https://doi.org/10.1093/nar/gkac247>
- Ahn, H. J., Jung, S. J., Kim, T. H., Oh, M. K., & Yoon, H.-K. (2015). Differences in Clinical Outcomes between Luminal A and B Type Breast Cancers according to the St. Gallen Consensus 2013. *Journal of Breast Cancer*, 18(2), 149. <https://doi.org/10.4048/jbc.2015.18.2.149>
- Aleksander, S. A., Balhoff, J., Carbon, S., Cherry, J. M., Drabkin, H. J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N. L., Hill, D. P., Lee, R., Mi, H., Moxon, S., Mungall, C. J., Muruganugan, A., Mushayahama, T., Sternberg, P. W., Thomas, P. D., Auken, K. Van, ... Westerfield, M. (2023). The Gene Ontology knowledgebase in 2023. *GENETICS*, 224(1). <https://doi.org/10.1093/genetics/iyad031>
- Almende B.V. and Contributors, Thieurmél B (2022). visNetwork:Network Visualization using 'vis.js' Library. R package version 2.1.2, <https://CRAN.R-project.org/package=visNetwork>.
- Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. In <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Aparicio, T., Ibarra, A., & Méndez, J. (2006). Cdc45-MCM-GINS, a new power player for DNA replication. *Cell Division*, 1(1), 18. <https://doi.org/10.1186/1747-1028-1-18>
- Arıcı, B. İ. (2022). EFFECTS OF DEPLETION OF CHRNA5 AND/OR TP53, AND TRANSIENT AND STABLE OVEREXPRESSION OF WILDTYPE OR MUTANT TP53 ON EXPRESSION OF DLK1-MEG3 LOCUS IN MCF7 CELLS.
- Aubrey, B. J., Kelly, G. L., Janic, A., Herold, M. J., & Strasser, A. (2018). How does p53 induce apoptosis and how does this relate to p53-mediated tumour suppression? *Cell Death & Differentiation*, 25(1), 104–113. <https://doi.org/10.1038/cdd.2017.169>
- Baines, A. T., Xu, D., & Der, C. J. (2011). Inhibition of Ras for cancer treatment: the search continues. *Future Medicinal Chemistry*, 3(14), 1787–1808. <https://doi.org/10.4155/fmc.11.121>
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S., & Soboleva, A. (2013). NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Research*, 41(D1), D991–D995. <https://doi.org/10.1093/nar/gks1193>
- Blighe, K., Rana, S., Lewis, M. (2022). EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labelling. R package version 1.14.0, <https://github.com/kevinblighe/EnhancedVolcano>.

- Bodenhofer, U., Kothmeier, A., & Hochreiter, S. (2011). APCluster: an R package for affinity propagation clustering. *Bioinformatics*, 27(17), 2463–2464. <https://doi.org/10.1093/bioinformatics/btr406>
- Brosh, R., & Rotter, V. (2009). When mutants gain new powers: news from the mutant p53 field. *Nature Reviews Cancer*, 9(10), 701–713. <https://doi.org/10.1038/nrc2693>
- Brown, J., Wilson, W. Exploiting tumour hypoxia in cancer treatment. *Nat Rev Cancer* 4, 437–447 (2004). <https://doi.org/10.1038/nrc1367>
- Bykov, V. J. N., Eriksson, S. E., Bianchi, J., & Wiman, K. G. (2018). Targeting mutant p53 for efficient cancer therapy. *Nature Reviews Cancer*, 18(2), 89–102. <https://doi.org/10.1038/nrc.2017.109>
- Bykov, V. J. N., Issaeva, N., Shilov, A., Hultcrantz, M., Pugacheva, E., Chumakov, P., Bergman, J., Wiman, K. G., & Selivanova, G. (2002). Restoration of the tumor suppressor function to mutant p53 by a low-molecular-weight compound. *Nature Medicine*, 8(3), 282–288. <https://doi.org/10.1038/nm0302-282>
- 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>
- Chang, Wi., Cheng, J., Allaire, J. J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2024). *shiny: Web Application Framework for R. R package*. <https://github.com/rstudio/shiny>
- Chellappan, S. P., Hiebert, S., Mudryj, M., Horowitz, J. M., & Nevins, J. R. (1991). The E2F transcription factor is a cellular target for the RB protein. *Cell*, 65(6), 1053–1061. [https://doi.org/10.1016/0092-8674\(91\)90557-F](https://doi.org/10.1016/0092-8674(91)90557-F)
- Chen, H.-Z., Tsai, S.-Y., & Leone, G. (2009). Emerging roles of E2Fs in cancer: an exit from cell cycle control. *Nature Reviews Cancer*, 9(11), 785–797. <https://doi.org/10.1038/nrc2696>
- Chevet, E., Fessart, D., Delom, F., Mulot, A., Vojtesek, B., Hrstka, R., Murray, E., Gray, T., & Hupp, T. (2013). Emerging roles for the pro-oncogenic anterior gradient-2 in cancer development. *Oncogene*, 32(20), 2499–2509. <https://doi.org/10.1038/onc.2012.346>
- Chiang, Y.-T., Chien, Y.-C., Lin, Y.-H., Wu, H.-H., Lee, D.-F., & Yu, Y.-L. (2021). The Function of the Mutant p53-R175H in Cancer. *Cancers*, 13(16), 4088. <https://doi.org/10.3390/cancers13164088>
- Cho, Y., Gorina, S., Jeffrey, P. D., & Pavletich, N. P. (1994). Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science (New York, N.Y.)*, 265(5170), 346–355. <https://doi.org/10.1126/science.8023157>
- Choi, J. K., Yu, U., Kim, S., & Yoo, O. J. (2003). Combining multiple microarray studies and modeling interstudy variation. *Bioinformatics*, 19(suppl_1), i84–i90. <https://doi.org/10.1093/bioinformatics/btg1010>
- Chou, T.-C. (2010). Drug Combination Studies and Their Synergy Quantification Using the Chou-Talalay Method. *Cancer Research*, 70(2), 440–446. <https://doi.org/10.1158/0008-5472.CAN-09-1947>

- Christowitz, C., Davis, T., Isaacs, A., van Niekerk, G., Hattingh, S., & Engelbrecht, A.-M. (2019). Mechanisms of doxorubicin-induced drug resistance and drug resistant tumour growth in a murine breast tumour model. *BMC Cancer*, 19(1), 757. <https://doi.org/10.1186/s12885-019-5939-z>
- Cinar, O., & Viechtbauer, W. (2022). The **poolr** Package for Combining Independent and Dependent ρ Values. *Journal of Statistical Software*, 101(1). <https://doi.org/10.18637/jss.v101.i01>
- Clavero, A. L., Boqvist, P. L., Ingelshed, K., Bosdotter, C., Sedimbi, S., Jiang, L., Wermeling, F., Vojtesek, B., Lane, D. P., & Kannan, P. (2023). MDM2 inhibitors, nutlin-3a and navtemadelin, retain efficacy in human and mouse cancer cells cultured in hypoxia. *Scientific Reports*, 13(1), 4583. <https://doi.org/10.1038/s41598-023-31484-0>
- Comşa, Ş., Cîmpean, A. M., & Raica, M. (2015). The Story of MCF-7 Breast Cancer Cell Line: 40 years of Experience in Research. *Anticancer Research*, 35(6), 3147–3154.
- Dang, C. V. (2012). MYC on the Path to Cancer. *Cell*, 149(1), 22–35. <https://doi.org/10.1016/j.cell.2012.03.003>
- Dias, K., Dvorkin-Gheva, A., Hallett, R. M., Wu, Y., Hassell, J., Pond, G. R., Levine, M., Whelan, T., & Bane, A. L. (2017). Claudin-Low Breast Cancer; Clinical & Pathological Characteristics. *PLOS ONE*, 12(1), e0168669. <https://doi.org/10.1371/journal.pone.0168669>
- Diaz, J. E. L., Ahsen, M. E., Schaffter, T., Chen, X., Realubit, R. B., Karan, C., Califano, A., Losic, B., & Stolovitzky, G. (2020). The transcriptomic response of cells to a drug combination is more than the sum of the responses to the monotherapies. *ELife*, 9. <https://doi.org/10.7554/eLife.52707>
- Donati, G., & Amati, B. (2022). MYC and therapy resistance in cancer: Risks and opportunities. *Molecular Oncology*, 16(21), 3828–3854. <https://doi.org/10.1002/1878-0261.13319>
- Dolgalev, I. (2022). msigdb: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format. R package version 7.5.1, <https://CRAN.R-project.org/package=msigdb>.
- Duffy, M. J., Harbeck, N., Nap, M., Molina, R., Nicolini, A., Senkus, E., & Cardoso, F. (2017). Clinical use of biomarkers in breast cancer: Updated guidelines from the European Group on Tumor Markers (EGTM). *European Journal of Cancer*, 75, 284–298. <https://doi.org/10.1016/j.ejca.2017.01.017>
- Dunn, O. J. (1958). Estimation of the Means of Dependent Variables. *The Annals of Mathematical Statistics*, 29(4), 1095–1111. <https://doi.org/10.1214/aoms/1177706443>
- 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>
- Engeland, K. Cell cycle arrest through indirect transcriptional repression by p53: I have a DREAM. *Cell Death Differ* 25, 114–132 (2018). <https://doi.org/10.1038/cdd.2017.172>
- El-Deiry, W. S., Tokino, T., Velculescu, V. E., Levy, D. B., Parsons, R., Trent, J. M., Lin, D., Mercer, W. E., Kinzler, K. W., & Vogelstein, B. (1993). WAF1, a potential mediator of p53 tumor suppression. *Cell*, 75(4), 817–825. [https://doi.org/10.1016/0092-8674\(93\)90500-P](https://doi.org/10.1016/0092-8674(93)90500-P)

- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>
- Federico, A., & Monti, S. (2020). hypeR: an R package for geneset enrichment workflows. *Bioinformatics*, 36(4), 1307–1308. <https://doi.org/10.1093/bioinformatics/btz700>
- Fischer, M., & Müller, G. A. (2017). Cell cycle transcription control: DREAM/MuvB and RB-E2F complexes. *Critical Reviews in Biochemistry and Molecular Biology*, 52(6), 638–662. <https://doi.org/10.1080/10409238.2017.1360836>
- Fischer, M., Quaas, M., Steiner, L., & Engeland, K. (2016). The p53-p21-DREAM-CDE/CHR pathway regulates G₂/M cell cycle genes. *Nucleic Acids Research*, 44(1), 164–174. <https://doi.org/10.1093/nar/gkv927>
- Fischer, M., Schwarz, R., Riege, K., DeCaprio, J. A., & Hoffmann, S. (2022). TargetGeneReg 2.0: a comprehensive web-atlas for p53, p63, and cell cycle-dependent gene regulation. *NAR Cancer*, 4(1). <https://doi.org/10.1093/narcan/zcac009>
- Fisher, R. A. (1932). Statistical methods for research workers (4th ed.). Oliver & Boyd.
- Frasor, J., Danes, J. M., Komm, B., Chang, K. C. N., Lyttle, C. R., & Katzenellenbogen, B. S. (2003). Profiling of Estrogen Up- and Down-Regulated Gene Expression in Human Breast Cancer Cells: Insights into Gene Networks and Pathways Underlying Estrogenic Control of Proliferation and Cell Phenotype. *Endocrinology*, 144(10), 4562–4574. <https://doi.org/10.1210/en.2003-0567>
- Freed-Pastor, W. A., & Prives, C. (2012). Mutant p53: one name, many proteins. *Genes & Development*, 26(12), 1268–1286. <https://doi.org/10.1101/gad.190678.112>
- 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). <https://doi.org/10.1093/bioinformatics/btx657>
- Ganz, C. (2016). rintrojs: A Wrapper for the Intro.js Library. *The Journal of Open Source Software*, 1(6), 63. <https://doi.org/10.21105/joss.00063>
- Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D., & Saez-Rodriguez, J. (2019). Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Research*, 29(8), 1363–1375. <https://doi.org/10.1101/gr.240663.118>
- Garg, A., Hazra, J. P., Sannigrahi, M. K., Rakshit, S., & Sinha, S. (2020). Variable Mutations at the p53-R273 Oncogenic Hotspot Position Leads to Altered Properties. *Biophysical Journal*, 118(3), 720–728. <https://doi.org/10.1016/j.bpj.2019.12.015>
- Garnier, S., Ross, N., boB Rudis, Filipovic-Pierucci, A., Galili, T., timelyportfolio, O’Callaghan, A., Greenwell, B., Sievert, C., Harris, D. J., Sciaini, M., & Chen, J. J. (2023). *sjmgarnier/viridis*: CRAN release v0.6.3. Zenodo. <https://doi.org/10.5281/zenodo.7890878>
- Geistlinger, L., Csaba, G., & Zimmer, R. (2016). Bioconductor’s EnrichmentBrowser: seamless navigation through combined results of set- & network-based enrichment analysis. *BMC Bioinformatics*, 17(1), 45. <https://doi.org/10.1186/s12859-016-0884-1>
- Gilic, M. B., & Tyner, A. L. (2020). Targeting protein tyrosine kinase 6 in cancer. *Biochimica et biophysica acta. Reviews on cancer*, 1874(2), 188432. <https://doi.org/10.1016/j.bbcan.2020.188432>

- Golden, E., Rashwan, R., Woodward, E. A., Sgro, A., Wang, E., Sorolla, A., Waryah, C., Tie, W. J., Cuyàs, E., Ratajska, M., Kardaś, I., Kozłowski, P., Johnstone, E. K. M., See, H. B., Duffy, C., Parry, J., Lagerborg, K. A., Czapiewski, P., Menendez, J. A., Gorczyński, A., ... Blancafort, P. (2021). The oncogene AAMDC links PI3K-AKT-mTOR signaling with metabolic reprogramming in estrogen receptor-positive breast cancer. *Nature communications*, 12(1), 1920. <https://doi.org/10.1038/s41467-021-22101-7>
- Grohar, P. J., Segars, L. E., Yeung, C., Pommier, Y., D'Incalci, M., Mendoza, A., & Helman, L. J. (2014). Dual Targeting of EWS-FLI1 Activity and the Associated DNA Damage Response with Trabectedin and SN38 Synergistically Inhibits Ewing Sarcoma Cell Growth. *Clinical Cancer Research*, 20(5), 1190–1203. <https://doi.org/10.1158/1078-0432.CCR-13-0901>
- Hänzelmann, S., Castelo, R., & Guinney, J. (2013). GSEA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics*, 14(1), 7. <https://doi.org/10.1186/1471-2105-14-7>
- Hassin, O., Nataraj, N. B., Shreberk-Shaked, M., Aylon, Y., Yaeger, R., Fontemaggi, G., Mukherjee, S., Maddalena, M., Avioz, A., Iancu, O., Mallel, G., Gershoni, A., Grosheva, I., Feldmesser, E., Ben-Dor, S., Golani, O., Hendel, A., Blandino, G., Kelsen, D., ... Oren, M. (2022). Different hotspot p53 mutants exert distinct phenotypes and predict outcome of colorectal cancer patients. *Nature Communications*, 13(1), 2800. <https://doi.org/10.1038/s41467-022-30481-7>
- Heldring, N., Pike, A., Andersson, S., Matthews, J., Cheng, G., Hartman, J., Tujague, M., Ström, A., Treuter, E., Warner, M., & Gustafsson, J.-Å. (2007). Estrogen Receptors: How Do They Signal and What Are Their Targets. *Physiological Reviews*, 87(3), 905–931. <https://doi.org/10.1152/physrev.00026.2006>
- Hientz, K., Mohr, A., Bhakta-Guha, D., & Efferth, T. (2017). The role of p53 in cancer drug resistance and targeted chemotherapy. *Oncotarget*, 8(5), 8921–8946. <https://doi.org/10.18632/oncotarget.13475>
- Hong, F., & Breitling, R. (2008). A comparison of meta-analysis methods for detecting differentially expressed genes in microarray experiments. *Bioinformatics*, 24(3), 374–382. <https://doi.org/10.1093/bioinformatics/btm620>
- Iansante, V., Choy, P., Fung, S. et al. PARP14 promotes the Warburg effect in hepatocellular carcinoma by inhibiting JNK1-dependent PKM2 phosphorylation and activation. *Nat Commun* 6, 7882 (2015). <https://doi.org/10.1038/ncomms8882>
- Issaeva, N., Bozko, P., Enge, M., Protopopova, M., Verhoef, L. G. G. C., Masucci, M., Pramanik, A., & Selivanova, G. (2004). Small molecule RITA binds to p53, blocks p53–MDM-2 interaction and activates p53 function in tumors. *Nature Medicine*, 10(12), 1321–1328. <https://doi.org/10.1038/nm1146>
- Janky, R., Verfaillie, A., Imrichová, H., de Sande, B. Van, Standaert, L., Christiaens, V., Hulselmans, G., Herten, K., Sanchez, M. N., Potier, D., Svetlichnyy, D., Atak, Z. K., Fiers, M., Marine, J.-C., & Aerts, S. (2014). iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLoS Computational Biology*, 10(7), e1003731. <https://doi.org/10.1371/journal.pcbi.1003731>

- Jeffrey, P. D., Gorina, S., & Pavletich, N. P. (1995). Crystal structure of the tetramerization domain of the p53 tumor suppressor at 1.7 angstroms. *Science (New York, N.Y.)*, 267(5203), 1498–1502. <https://doi.org/10.1126/science.7878469>
- Jia, L., Yao, W., Jiang, Y., Li, Y., Wang, Z., Li, H., Huang, F., Li, J., Chen, T., & Zhang, H. (2022). Development of interactive biological web applications with R/Shiny. *Briefings in Bioinformatics*, 23(1). <https://doi.org/10.1093/bib/bbab415>
- Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J. F., Wyczalkowski, M. A., Leiserson, M. D. M., Miller, C. A., Welch, J. S., Walter, M. J., Wendl, M. C., Ley, T. J., Wilson, R. K., Raphael, B. J., & Ding, L. (2013). Mutational landscape and significance across 12 major cancer types. *Nature*, 502(7471), 333–339. <https://doi.org/10.1038/nature12634>
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., & Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic acids research*, 40(Database issue), D109–D114. <https://doi.org/10.1093/nar/gkr988>
- Karimian, A., Ahmadi, Y., & Yousefi, B. (2016). Multiple functions of p21 in cell cycle, apoptosis and transcriptional regulation after DNA damage. *DNA Repair*, 42, 63–71. <https://doi.org/10.1016/j.dnarep.2016.04.008>
- Kastenhuber, E. R., & Lowe, S. W. (2017). Putting p53 in Context. *Cell*, 170(6), 1062–1078. <https://doi.org/10.1016/j.cell.2017.08.028>
- Kent, L. N., & Leone, G. (2019). The broken cycle: E2F dysfunction in cancer. *Nature Reviews Cancer*, 19(6), 326–338. <https://doi.org/10.1038/s41568-019-0143-7>
- Keşküş, A. G. (2021). *IDENTIFICATION OF MODULATORY FUNCTIONS OF TP53, ESTROGEN SIGNALING, AND 14q32.31 mRNA CLUSTER ON CHRNA5 KNOCK-DOWN EXPRESSION PROFILE AND DEVELOPMENT OF syncRgy APP*.
- 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>
- Koker, S. C., Jahja, E., Shehwana, H., Keskus, A. G., & Konu, O. (2018). Cholinergic Receptor Nicotinic Alpha 5 (*CHRNA5*) RNAi is associated with cell cycle inhibition, apoptosis, DNA damage response and drug sensitivity in breast cancer. *PLOS ONE*, 13(12), e0208982. <https://doi.org/10.1371/journal.pone.0208982>
- Kolberg, L., Raudvere, U., Kuzmin, I., Adler, P., Vilo, J., & Peterson, H. (2023). G:Profiler—interoperable web service for Functional Enrichment Analysis and gene identifier mapping (2023 update). *Nucleic Acids Research*, 51(W1). <https://doi.org/10.1093/nar/gkad347>
- Konopka, T. (2023). *umap: Uniform Manifold Approximation and Projection. R package*. <https://CRAN.R-project.org/package=umap>
- Kwong, L. N., Heffernan, T. P., & Chin, L. (2013). A Systems Biology Approach to Personalizing Therapeutic Combinations. *Cancer Discovery*, 3(12), 1339–1344. <https://doi.org/10.1158/2159-8290.CD-13-0394>
- Lacroix, M., & Leclercq, G. (2004). Relevance of Breast Cancer Cell Lines as Models for Breast Tumours: An Update. *Breast Cancer Research and Treatment*, 83(3), 249–289. <https://doi.org/10.1023/B:BREA.0000014042.54925.cc>

- Lambert, J. M., Moshfegh, A., Hainaut, P., Wiman, K. G., & Bykov, V. J. (2010). Mutant p53 reactivation by PRIMA-1MET induces multiple signaling pathways converging on apoptosis. *Oncogene*, 29(9), 1329–1338. <https://doi.org/10.1038/onc.2009.425>
- Lane, D. P. (1992). p53, guardian of the genome. *Nature*, 358(6381), 15–16. <https://doi.org/10.1038/358015a0>
- 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), R29. <https://doi.org/10.1186/gb-2014-15-2-r29>
- Lee, C.-H., Huang, C.-S., Chen, C.-S., Tu, S.-H., Wang, Y.-J., Chang, Y.-J., Tam, K.-W., Wei, P.-L., Cheng, T.-C., Chu, J.-S., Chen, L.-C., Wu, C.-H., & Ho, Y.-S. (2010). Overexpression and Activation of the $\alpha 9$ -Nicotinic Receptor During Tumorigenesis in Human Breast Epithelial Cells. *JNCI: Journal of the National Cancer Institute*, 102(17), 1322–1335. <https://doi.org/10.1093/jnci/djq300>
- Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., & Storey, J. D. (2012). The <tt>sva</tt> package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*, 28(6), 882–883. <https://doi.org/10.1093/bioinformatics/bts034>
- Lehmann, B. D., Bauer, J. A., Chen, X., Sanders, M. E., Chakravarthy, A. B., Shyr, Y., & Pietenpol, J. A. (2011). Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *Journal of Clinical Investigation*, 121(7), 2750–2767. <https://doi.org/10.1172/JCI45014>
- Li, D.-H., Liu, X.-K., Tian, X.-T., Liu, F., Yao, X.-J., & Dong, J.-F. (2023). PPAR γ : A Promising Therapeutic Target in Breast Cancer and Regulation by Natural Drugs. *PPAR Research*, 2023, 1–18. <https://doi.org/10.1155/2023/4481354>
- Li, H., Li, T., Quang, D., & Guan, Y. (2018). Network Propagation Predicts Drug Synergy in Cancers. *Cancer Research*, 78(18), 5446–5457. <https://doi.org/10.1158/0008-5472.CAN-18-0740>
- Li, Z., Hu, P., Tu, J., & Yu, N. (2016). Luminal B breast cancer: patterns of recurrence and clinical outcome. *Oncotarget*, 7(40), 65024–65033. <https://doi.org/10.18632/oncotarget.11344>
- 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>
- Litovchick, L., Sadasivam, S., Florens, L., Zhu, X., Swanson, S. K., Velmurugan, S., Chen, R., Washburn, M. P., Liu, X. S., & DeCaprio, J. A. (2007). Evolutionarily Conserved Multisubunit RBL2/p130 and E2F4 Protein Complex Represses Human Cell Cycle-Dependent Genes in Quiescence. *Molecular Cell*, 26(4), 539–551. <https://doi.org/10.1016/j.molcel.2007.04.015>
- Liu, S., Wang, Z., Zhu, R., Wang, F., Cheng, Y., & Liu, Y. (2021). Three Differential Expression Analysis Methods for RNA Sequencing: limma, EdgeR, DESeq2. *Journal of Visualized Experiments*, 175. <https://doi.org/10.3791/62528>
- Loibl, S., & Gianni, L. (2017). HER2-positive breast cancer. *The Lancet*, 389(10087), 2415–2429. [https://doi.org/10.1016/S0140-6736\(16\)32417-5](https://doi.org/10.1016/S0140-6736(16)32417-5)

- Long, G. V., Stroyakovskiy, D., Gogas, H., Levchenko, E., de Braud, F., Larkin, J., Garbe, C., Jouary, T., Hauschild, A., Grob, J. J., Sileni, V. C., Lebbe, C., Mandalà, M., Millward, M., Arance, A., Bondarenko, I., Haanen, J. B. A. G., Hansson, J., Utikal, J., ... Flaherty, K. (2014). Combined BRAF and MEK Inhibition versus BRAF Inhibition Alone in Melanoma. *New England Journal of Medicine*, 371(20), 1877–1888. <https://doi.org/10.1056/NEJMoa1406037>
- 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>
- Łukasiewicz, S., Czezelewski, M., Forma, A., Baj, J., Sitarz, R., & Stanisławek, A. (2021). Breast Cancer—Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies—An Updated Review. *Cancers*, 13(17), 4287. <https://doi.org/10.3390/cancers13174287>
- Mages, C. F. S., Wintsche, A., Bernhart, S. H., & Müller, G. A. (2017). The DREAM complex through its subunit Lin37 cooperates with Rb to initiate quiescence. *ELife*, 6. <https://doi.org/10.7554/eLife.26876>
- Matoba, S., Kang, J.-G., Patino, W. D., Wragg, A., Boehm, M., Gavrilova, O., Hurley, P. J., Bunz, F., & Hwang, P. M. (2006). p53 Regulates Mitochondrial Respiration. *Science*, 312(5780), 1650–1653. <https://doi.org/10.1126/science.1126863>
- Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L. M., Ding, W., Bell, R., Rosenthal, J., Hussey, C., Tran, T., McClure, M., Frye, C., Hattier, T., Phelps, R., Haugen-Strano, A., ... Skolnick, M. H. (1994). A Strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene *BRCA1*. *Science*, 266(5182), 66–71. <https://doi.org/10.1126/science.7545954>
- Minotti, G., Menna, P., Salvatorelli, E., Cairo, G., & Gianni, L. (2004). Anthracyclines: Molecular Advances and Pharmacologic Developments in Antitumor Activity and Cardiotoxicity. *Pharmacological Reviews*, 56(2), 185–229. <https://doi.org/10.1124/pr.56.2.6>
- Mohell, N., Alfredsson, J., Fransson, Å., Uustalu, M., Byström, S., Gullbo, J., Hallberg, A., Bykov, V. J. N., Björklund, U., & Wiman, K. G. (2015). APR-246 overcomes resistance to cisplatin and doxorubicin in ovarian cancer cells. *Cell Death & Disease*, 6(6), e1794–e1794. <https://doi.org/10.1038/cddis.2015.143>
- Muller, P. A. J., & Vousden, K. H. (2013). p53 mutations in cancer. *Nature Cell Biology*, 15(1), 2–8. <https://doi.org/10.1038/ncb2641>
- Muz, B., de la Puente, P., Azab, F., & Azab, A. K. (2015). The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy. *Hypoxia*, 83. <https://doi.org/10.2147/HP.S93413>
- Nakano, K., & Vousden, K. H. (2001). PUMA, a Novel Proapoptotic Gene, Is Induced by p53. *Molecular Cell*, 7(3), 683–694. [https://doi.org/10.1016/S1097-2765\(01\)00214-3](https://doi.org/10.1016/S1097-2765(01)00214-3)
- Nakajima, R., Zhao, L., Zhou, Y., Shirasawa, M., Uchida, A., Murakawa, H., Fikriyanti, M., Iwanaga, R., Bradford, A. P., Araki, K., Warita, T., & Ohtani, K. (2023). Deregulated E2F Activity as a Cancer-Cell Specific Therapeutic Tool. *Genes*, 14(2), 393. <https://doi.org/10.3390/genes14020393>

- Neve, R. M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F. L., Fevr, T., Clark, L., Bayani, N., Coppe, J.-P., Tong, F., Speed, T., Spellman, P. T., DeVries, S., Lapuk, A., Wang, N. J., Kuo, W.-L., Stilwell, J. L., Pinkel, D., Albertson, D. G., ... Gray, J. W. (2006). A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell*, 10(6), 515–527. <https://doi.org/10.1016/j.ccr.2006.10.008>
- Neyman, J. & Pearson, E.S. On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika* 1928; 20A: 175–240.
- Nguyen, T.-A. T., Grimm, S. A., Bushel, P. R., Li, J., Li, Y., Bennett, B. D., Lavender, C. A., Ward, J. M., Fargo, D. C., Anderson, C. W., Li, L., Resnick, M. A., & Menendez, D. (2018). Revealing a human p53 universe. *Nucleic Acids Research*, 46(16), 8153–8167. <https://doi.org/10.1093/nar/gky720>
- Nielsen, F. (2016). *Hierarchical Clustering* (pp. 195–211). https://doi.org/10.1007/978-3-319-21903-5_8
- Nugoli, M., Chuchana, P., Vendrell, J., Orsetti, B., Ursule, L., Nguyen, C., Birnbaum, D., Douzery, E. J., Cohen, P., & Theillet, C. (2003). Genetic variability in MCF-7 sublines: evidence of rapid genomic and RNA expression profile modifications. *BMC cancer*, 3, 13. <https://doi.org/10.1186/1471-2407-3-13>
- Obermayer, A., Dong, L., Hu, Q., Golden, M., Noble, J. D., Rodriguez, P., Robinson, T. J., Teng, M., Tan, A.-C., & Shaw, T. I. (2022). DRPPM-EASY: A Web-Based Framework for Integrative Analysis of Multi-Omics Cancer Datasets. *Biology*, 11(2), 260. <https://doi.org/10.3390/biology11020260>
- Olivier, M., Hollstein, M., & Hainaut, P. (2010). TP53 Mutations in Human Cancers: Origins, Consequences, and Clinical Use. *Cold Spring Harbor Perspectives in Biology*, 2(1), a001008–a001008. <https://doi.org/10.1101/cshperspect.a001008>
- Ouriq, F., Kwiecinski, M. R., Felipe, K. B., Correia, J. F. G., Farias, M. S., Castro, L. S. E. P. W., Grinevicius, V. M. A. S., Valderrama, J., Rios, D., Benites, J., Calderon, P. B., & Pedrosa, R. C. (2015). DNA Damage and Inhibition of Akt Pathway in MCF-7 Cells and Ehrlich Tumor in Mice Treated with 1,4-Naphthoquinones in Combination with Ascorbate. *Oxidative Medicine and Cellular Longevity*, 2015, 1–10. <https://doi.org/10.1155/2015/495305>
- Petitjean, A., Mathe, E., Kato, S., Ishioka, C., Tavtigian, S. V., Hainaut, P., & Olivier, M. (2007). Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Human Mutation*, 28(6), 622–629. <https://doi.org/10.1002/HUMU.20495>
- Perrier V, Meyer F (2020). fresh: Create Custom 'Bootstrap' Themes to Use in 'Shiny'. R package version 0.2.0, <https://CRAN.R-project.org/package=fresh>.
- Perrier V, Meyer F, Granjon D (2024). shinyWidgets: Custom Inputs Widgets for Shiny. R package version 0.8.3, <https://dreamrs.github.io/shinyWidgets/>, <https://github.com/dreamRs/shinyWidgets>.
- Piñero, J., Ramírez-Anguaita, J. M., Saüch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F., & Furlong, L. I. (2019). The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkz1021>
- Pommier, R. M., Sanlaville, A., Tonon, L., Kielbassa, J., Thomas, E., Ferrari, A., Sertier, A.-S., Hollande, F., Martinez, P., Tissier, A., Morel, A.-P., Ouzounova, M., & Puisieux, A.

- (2020). Comprehensive characterization of claudin-low breast tumors reflects the impact of the cell-of-origin on cancer evolution. *Nature Communications*, 11(1), 3431. <https://doi.org/10.1038/s41467-020-17249-7>
- Quante, T., Otto, B., Brázdová, M., Kejnovská, I., Deppert, W., & Tolstonog, G. V. (2012). Mutant p53 is a transcriptional co-factor that binds to G-rich regulatory regions of active genes and generates transcriptional plasticity. *Cell Cycle*, 11(17), 3290–3303. <https://doi.org/10.4161/cc.21646>
- Rahman, Md. M., Islam, Md. R., Rahman, F., Rahaman, Md. S., Khan, Md. S., Abrar, S., Ray, T. K., Uddin, M. B., Kali, Most. S. K., Dua, K., Kamal, M. A., & Chellappan, D. K. (2022). Emerging Promise of Computational Techniques in Anti-Cancer Research: At a Glance. *Bioengineering*, 9(8), 335. <https://doi.org/10.3390/bioengineering9080335>
- Ray, C., Soderblom, E. J., Bai, Y., Carroll, F. I., Caron, M. G., & Barak, L. S. (2017). Probing the Allosteric Role of the $\alpha 5$ Subunit of $\alpha 3\beta 4\alpha 5$ Nicotinic Acetylcholine Receptors by Functionally Selective Modulators and Ligands. *ACS Chemical Biology*, 12(3), 702–714. <https://doi.org/10.1021/acscchembio.6b01117>
- 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>
- Riley, T., Sontag, E., Chen, P., & Levine, A. (2008). Transcriptional control of human p53-regulated genes. *Nature Reviews Molecular Cell Biology*, 9(5), 402–412. <https://doi.org/10.1038/nrm2395>
- Risso, D., Ngai, J., Speed, T. P., & Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature Biotechnology*, 32(9), 896–902. <https://doi.org/10.1038/nbt.2931>
- 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>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). `edgeR`: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Rosas-Cruz, A., Salinas-Jazmín, N., Valdés-Rives, A., & Velasco-Velázquez, M. A. (2022). DRD1 and DRD4 are differentially expressed in breast tumors and breast cancer stem cells: pharmacological implications. *Translational Cancer Research*, 11(11), 3941–3950. <https://doi.org/10.21037/tcr-22-783>
- Sabapathy, K., & Lane, D. P. (2018). Therapeutic targeting of p53: all mutants are equal, but some mutants are more equal than others. *Nature Reviews Clinical Oncology*, 15(1), 13–30. <https://doi.org/10.1038/nrclinonc.2017.151>
- Sadasivam, S., & DeCaprio, J. A. (2013). The DREAM complex: master coordinator of cell cycle-dependent gene expression. *Nature Reviews Cancer*, 13(8), 585–595. <https://doi.org/10.1038/nrc3556>
- Saha, M. N., Qiu, L., & Chang, H. (2013). Targeting p53 by small molecules in hematological malignancies. *Journal of Hematology & Oncology*, 6(1), 23. <https://doi.org/10.1186/1756-8722-6-23>

- Saha, S. K., Islam, S. M. R., Saha, T., Nishat, A., Biswas, P. K., Gil, M., Nkenyereye, L., El-Sappagh, S., Islam, Md. S., & Cho, S.-G. (2021). Prognostic role of *EGR1* in breast cancer: a systematic review. *BMB Reports*, 54(10), 497–504. <https://doi.org/10.5483/BMBRep.2021.54.10.087>
- Samassekou, O., Bastien, N., Lichtensztejn, D., Yan, J., Mai, S., & Drouin, R. (2014). Different *TP53* mutations are associated with specific chromosomal rearrangements, telomere length changes, and remodeling of the nuclear architecture of telomeres. *Genes, Chromosomes and Cancer*, 53(11), 934–950. <https://doi.org/10.1002/gcc.22205>
- Schaal, C., & Chellappan, S. P. (2014). Nicotine-Mediated Cell Proliferation and Tumor Progression in Smoking-Related Cancers. *Molecular Cancer Research*, 12(1), 14–23. <https://doi.org/10.1158/1541-7786.MCR-13-0541>
- Shahbandi, A., Nguyen, H. D., & Jackson, J. G. (2020). *TP53* Mutations and Outcomes in Breast Cancer: Reading beyond the Headlines. *Trends in Cancer*, 6(2), 98–110. <https://doi.org/10.1016/j.trecan.2020.01.007>
- Shehwana, H., Keskus, A. G., Ozdemir, S. E., Acikgöz, A. A., Biyik-Sit, R., Cagnan, I., Gunes, D., Jahja, E., Cingir-Koker, S., Olmezer, G., Sucularli, C., & Konu, O. (2021). *CHRNA5* belongs to the secondary estrogen signaling network exhibiting prognostic significance in breast cancer. *Cellular Oncology*, 44(2), 453–472. <https://doi.org/10.1007/s13402-020-00581-x>
- Shen, H., Moran, D. M., & Maki, C. G. (2008). Transient Nutlin-3a Treatment Promotes Endoreduplication and the Generation of Therapy-Resistant Tetraploid Cells. *Cancer Research*, 68(20), 8260–8268. <https://doi.org/10.1158/0008-5472.CAN-08-1901>
- Šidák, Z. (1967). Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. *Journal of the American Statistical Association*, 62(318), 626–633. <https://doi.org/10.1080/01621459.1967.10482935>
- Skibinski, A., & Kuperwasser, C. (2015). The origin of breast tumor heterogeneity. *Oncogene*, 34(42), 5309–5316. <https://doi.org/10.1038/onc.2014.475>
- Soussi, T., & Wiman, K. G. (2007). Shaping Genetic Alterations in Human Cancer: The p53 Mutation Paradigm. *Cancer Cell*, 12(4), 303–312. <https://doi.org/10.1016/j.ccr.2007.10.001>
- Spitz, M. R., Amos, C. I., Dong, Q., Lin, J., & Wu, X. (2008). The *CHRNA5-A3* Region on Chromosome 15q24-25.1 Is a Risk Factor Both for Nicotine Dependence and for Lung Cancer. *JNCI: Journal of the National Cancer Institute*, 100(21), 1552–1556. <https://doi.org/10.1093/jnci/djn363>
- Sprang, M., Andrade-Navarro, M. A., & Fontaine, J.-F. (2022). Batch effect detection and correction in RNA-seq data using machine-learning-based automated assessment of quality. *BMC Bioinformatics*, 23(S6), 279. <https://doi.org/10.1186/s12859-022-04775-y>
- Stouffer, S. A., Suchman, E. A., Devinney, L. C., Star, S. A., & Williams, R. M., Jr. (1949). The American soldier: Adjustment during army life. (Studies in social psychology in World War II). Princeton Univ. Press.
- Sullivan, K. D., Galbraith, M. D., Andrysik, Z., & Espinosa, J. M. (2018). Mechanisms of transcriptional regulation by p53. *Cell Death & Differentiation*, 25(1), 133–143. <https://doi.org/10.1038/cdd.2017.174>

- Sun, Y.-L., Patel, A., Kumar, P., & Chen, Z.-S. (2012). Role of ABC transporters in cancer chemotherapy. *Chinese Journal of Cancer*, 31(2), 51–57. <https://doi.org/10.5732/cjc.011.10466>
- Sun, Y., Xia, P., Zhang, H., Liu, B., & Shi, Y. (2015). P53 is required for Doxorubicin-induced apoptosis via the TGF-beta signaling pathway in osteosarcoma-derived cells. *American journal of cancer research*, 6(1), 114–125.
- Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J., & von Mering, C. (2023). The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, 51(D1), D638–D646. <https://doi.org/10.1093/nar/gkac1000>
- Tan, B. S., Tiong, K. H., Choo, H. L., Chung, F. F.-L., Hii, L.-W., Tan, S. H., Yap, I. K. S., Pani, S., Khor, N. T. W., Wong, S. F., Rosli, R., Cheong, S.-K., & Leong, C.-O. (2015). Mutant p53-R273H mediates cancer cell survival and anoikis resistance through AKT-dependent suppression of BCL2-modifying factor (BMF). *Cell Death & Disease*, 6(7), e1826–e1826. <https://doi.org/10.1038/cddis.2015.191>
- Tippett, LHC. The Methods of Statistics An Introduction mainly for Workers in the Biological Sciences. London: Williams & Norgate Ltd; 1931. The Methods of Statistics. An introduction mainly for workers in the biological sciences.
- Toro-Domínguez, D., Villatoro-García, J. A., Martorell-Marugán, J., Román-Montoya, Y., Alarcón-Riquelme, M. E., & Carmona-Sáez, P. (2021). A survey of gene expression meta-analysis: Methods and applications. *Briefings in Bioinformatics*, 22(2), 1694–1705. <https://doi.org/10.1093/bib/bbaa019>
- Tovar, C., Rosinski, J., Filipovic, Z., Higgins, B., Kolinsky, K., Hilton, H., Zhao, X., Vu, B. T., Qing, W., Packman, K., Myklebost, O., Heimbrook, D. C., & Vassilev, L. T. (2006). Small-molecule MDM2 antagonists reveal aberrant p53 signaling in cancer: Implications for therapy. *Proceedings of the National Academy of Sciences*, 103(6), 1888–1893. <https://doi.org/10.1073/pnas.0507493103>
- Truong, T., Hung, R. J., Amos, C. I., Wu, X., Bickeböller, H., Rosenberger, A., Sauter, W., Illig, T., Wichmann, H.-E., Risch, A., Dienemann, H., Kaaks, R., Yang, P., Jiang, R., Wiencke, J. K., Wrensch, M., Hansen, H., Kelsey, K. T., Matsuo, K., ... Spitz, M. R. (2010). Replication of Lung Cancer Susceptibility Loci at Chromosomes 15q25, 5p15, and 6p21: A Pooled Analysis From the International Lung Cancer Consortium. *JNCI: Journal of the National Cancer Institute*, 102(13), 959–971. <https://doi.org/10.1093/jnci/djq178>
- 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). <https://doi.org/10.1016/j.cell.2017.06.010>
- Tufail, M., Cui, J., & Wu, C. (2022). Breast cancer: molecular mechanisms of underlying resistance and therapeutic approaches. *American Journal of Cancer Research*, 12(7), 2920–2949.

- Uxa, S., Bernhart, S. H., Mages, C. F. S., Fischer, M., Kohler, R., Hoffmann, S., Stadler, P. F., Engeland, K., & Müller, G. A. (2019). DREAM and RB cooperate to induce gene repression and cell-cycle arrest in response to p53 activation. *Nucleic Acids Research*, 47(17), 9087–9103. <https://doi.org/10.1093/nar/gkz635>
- Vassilev, L. T., Vu, B. T., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., Kong, N., Kammlott, U., Lukacs, C., Klein, C., Fotouhi, N., & Liu, E. A. (2004). In Vivo Activation of the p53 Pathway by Small-Molecule Antagonists of MDM2. *Science*, 303(5659), 844–848. <https://doi.org/10.1126/science.1092472>
- Vejpongsa, P., & Yeh, E. T. H. (2013). Topoisomerase 2 β : A Promising Molecular Target for Primary Prevention of Anthracycline-Induced Cardiotoxicity. *Clinical Pharmacology & Therapeutics*, 95(1), 45–52. <https://doi.org/10.1038/clpt.2013.201>
- Vousden, K. H., & Lane, D. P. (2007). p53 in health and disease. *Nature Reviews Molecular Cell Biology*, 8(4), 275–283. <https://doi.org/10.1038/nrm2147>
- Vural-Ozdeniz, M., Calisir, K., Acar, R., Yavuz, A., Ozgur, M. M., Dalgic, E., & Konu, O. (2024). CAP-RNAseq: an integrated pipeline for functional annotation and prioritization of co-expression clusters. *Briefings in Bioinformatics*, 25(2). <https://doi.org/10.1093/bib/bbad536>
- Walerych, D., Lisek, K., Sommaggio, R., Piazza, S., Ciani, Y., Dalla, E., Rajkowska, K., Gaweda-Walerych, K., Ingallina, E., Tonelli, C., Morelli, M. J., Amato, A., Eterno, V., Zambelli, A., Rosato, A., Amati, B., Wiśniewski, J. R., & Sal, G. Del. (2016). Proteasome machinery is instrumental in a common gain-of-function program of the p53 missense mutants in cancer. *Nature Cell Biology*, 18(8), 897–909. <https://doi.org/10.1038/ncb3380>
- Wang, B., Guo, H., Yu, H., Chen, Y., Xu, H., & Zhao, G. (2021). The Role of the Transcription Factor EGR1 in Cancer. *Frontiers in Oncology*, 11. <https://doi.org/10.3389/fonc.2021.642547>
- 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>
- Weisz, L., Oren, M., & Rotter, V. (2007). Transcription regulation by mutant p53. *Oncogene*, 26(15), 2202–2211. <https://doi.org/10.1038/sj.onc.1210294>
- WHITLOCK, M. C. (2005). Combining probability from independent tests: the weighted Z - method is superior to Fisher's approach. *Journal of Evolutionary Biology*, 18(5), 1368–1373. <https://doi.org/10.1111/j.1420-9101.2005.00917.x>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag .
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wickham, H. (2023). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.5.1, <https://CRAN.R-project.org/package=stringr>.
- Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation*. R package. <https://github.com/tidyverse/dplyr>
- Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., Mangion, J., Collins, N., Gregory, S., Gumbs, C., Micklem, G., Barfoot, R., Hamoudi, R., Patel, S., Rices, C., Biggs, P.,

- Hashim, Y., Smith, A., Connor, F., Arason, A., ... Stratton, M. R. (1995). Identification of the breast cancer susceptibility gene BRCA2. *Nature*, 378(6559), 789–792. <https://doi.org/10.1038/378789a0>
- Xia, J., Benner, M. J., & Hancock, R. E. (2014). NetworkAnalyst--integrative approaches for protein-protein interaction network analysis and visual exploration. *Nucleic acids research*, 42(Web Server issue), W167–W174. <https://doi.org/10.1093/nar/gku443>
- Xie, Y., Cheng, J., Tan, X. (2023). DT: A Wrapper of the JavaScript Library 'DataTables'. R package version 0.28, <https://CRAN.R-project.org/package=DT>.
- Xie, Y., Allaire, J., Horner, J. (2023). markdown: Render Markdown with 'commonmark'. R package version 1.12, <https://CRAN.R-project.org/package=markdown>.
- Yoon, S., Baik, B., Park, T., & Nam, D. (2021). Powerful p-value combination methods to detect incomplete association. *Scientific Reports*, 11(1), 6980. <https://doi.org/10.1038/s41598-021-86465-y>
- Yu, G. (2022). enrichplot: Visualization of Functional Enrichment Result. R package version 1.16.2, <https://yulab-smu.top/biomedical-knowledge-mining-book/>.
- Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology*, 16(5), 284–287. <https://doi.org/10.1089/omi.2011.0118>
- Yu, G., Wang, L.-G., Yan, G.-R., & He, Q.-Y. (2014). Dose: An R/bioconductor package for Disease Ontology Semantic and enrichment analysis. *Bioinformatics*, 31(4), 608–609. <https://doi.org/10.1093/bioinformatics/btu684>
- Yu, Y., Zhang, N., Mai, Y., Ren, L., Chen, Q., Cao, Z., Chen, Q., Liu, Y., Hou, W., Yang, J., Hong, H., Xu, J., Tong, W., Dong, L., Shi, L., Fang, X., & Zheng, Y. (2023). Correcting batch effects in large-scale multiomics studies using a reference-material-based ratio method. *Genome Biology*, 24(1), 201. <https://doi.org/10.1186/s13059-023-03047-z>
- Zhang, C., Liu, J., Liang, Y., Wu, R., Zhao, Y., Hong, X., Lin, M., Yu, H., Liu, L., Levine, A. J., Hu, W., & Feng, Z. (2013). Tumour-associated mutant p53 drives the Warburg effect. *Nature Communications*, 4(1), 2935. <https://doi.org/10.1038/ncomms3935>
- Zhang, S., Carlsen, L., Borrero, L. H., Seyhan, A. A., Tian, X., & El-Deiry, W. S. (2022). Advanced Strategies for Therapeutic Targeting of Wild-Type and Mutant p53 in Cancer. *Biomolecules*, 12(4), 548. <https://doi.org/10.3390/biom12040548>
- Zhang, X., Wu, Z., Hao, Y., Yu, T., Li, X., Liang, Y., Li, J., Huang, L., Xu, Y., Li, X., Xu, X., Wang, W., Xu, G., Zhang, X., Lv, Q., Fang, Y., Xu, R., & Qian, W. (2022). Aberrantly Activated APOBEC3B Is Associated With Mutant p53-Driven Refractory/Relapsed Diffuse Large B-Cell Lymphoma. *Frontiers in immunology*, 13, 888250. <https://doi.org/10.3389/fimmu.2022.888250>
- Zhang, Y., Parmigiani, G., & Johnson, W. E. (2020). ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genomics and Bioinformatics*, 2(3). <https://doi.org/10.1093/nargab/lqaa078>
- Zhang, Y., Xiong, S., Liu, B., Pant, V., Celii, F., Chau, G., Elizondo-Fraire, A. C., Yang, P., You, M. J., El-Naggar, A. K., Navin, N. E., & Lozano, G. (2018). Somatic Trp53 mutations differentially drive breast cancer and evolution of metastases. *Nature Communications*, 9(1), 3953. <https://doi.org/10.1038/s41467-018-06146-9>

- Zhang, Z. F., Huang, T. J., Zhang, X. K., Xie, Y. J., Lin, S. T., Luo, F. F., Meng, D. F., Hu, H., Wang, J., Peng, L. X., Qian, C. N., Cheng, C., & Huang, B. J. (2020). AKR1C2 acts as a targetable oncogene in esophageal squamous cell carcinoma via activating PI3K/AKT signaling pathway. *Journal of cellular and molecular medicine*, 24(17), 9999–10012. <https://doi.org/10.1111/jcmm.15604>
- Zhu, K., Yang, X., Tai, H. et al. HER2-targeted therapies in cancer: a systematic review. *Biomark Res* 12, 16 (2024). <https://doi.org/10.1186/s40364-024-00565-1>
- Zhu, X.-F., Li, W. E. N., Ma, J.-Y., Shao, N. A. N., Zhang, Y.-J., Liu, R.-M., Wu, W.-B., Lin, Y., & Wang, S.-M. (2015). Knockdown of heme oxygenase-1 promotes apoptosis and autophagy and enhances the cytotoxicity of doxorubicin in breast cancer cells. *Oncology Letters*, 10(5), 2974–2980. <https://doi.org/10.3892/ol.2015.3735>
- Zhumakayeva, A. M., Rakhimov, K. D., Omarova, I. M., Arystan, L. I., & Adekenov, S. M. (2019). Experimental, Clinical and Morphological Analysis of H-Ras Oncoproteins for Locally Advanced Breast Cancer. *Open Access Macedonian Journal of Medical Sciences*, 7(19), 3153–3157. <https://doi.org/10.3889/oamjms.2019.708>

7. Appendix

Appendix A: Experimental confirmation of the combinatorial treatments and its synergistic effect to reverse resistance

Section A1. Cell culture maintenance (carried out by Burcin Aras)

Human embryonic kidney 293 (HEK293) cells and MCF7 human breast cancer cells were cultured as monolayer in Dulbecco's Modified Eagle Medium (DMEM) – low glucose (L0064, Biowest, France) supplemented with L- Glutamine (K0100-670, Cegrogen Biotech, Germany), 10% fetal bovine serum (04-127-1A, Biological Industries, USA) and 1% penicillin/streptomycin (P0100-790, Cegrogen Biotech, Germany). For the generated stable MCF7 cells, puromycin (P8833, Sigma-Aldrich, Germany) was added to complete DMEM at 1 µg/ml concentration during cell selection and maintenance. All cells were passaged according to their confluency or changed medium every other day and grown in a humidified incubator with 5% CO₂ at 37°C.

The development of these stable cell lines and all the relevant materials, methods and results are described previously by Burcin (Arici, 2022).

Section A2. mRNA isolation (carried out by Burcin Aras)

After 72 hours, old media was discarded, and wells were washed once with 1X PBS. RNeasy (Qiagen, Germany) kit was used for mRNA isolation according to manufacturer's instructions. RNA was eluted and dissolved in 35 µl nuclease-free water. 800 ng RNA was used for cDNA synthesis by using the RevertAid RT kit. cDNAs were diluted 1:20 and stored at -200C.

Section A3. Doxorubicin and siCHRNA5 treatment (carried out by Cagdas Yanyatan)

Cell lines were cultured in a medium containing 10% FBS, 2% L-Glutamine, 1% non-essential amino acids, 2% sodium pyruvate, and 1% penicillin-streptomycin. 24 hours after seeding

200,000 cells per well in 6 or 12-well plates, treatments with doxorubicin or si*CHRNA5* were performed. For siRNA treatments, Qiagen's FlexiTube siRNA product catalog number 1027419, containing 5 nmol of siRNA, was used. Both scrambled control and si*CHRNA5* versions of the same product were employed. HiPerfect Transfection reagent, also from Qiagen with catalog number 301705, was used to facilitate transient transfection. The optimal concentration for siRNA was observed to be 20nM. Additionally, doxorubicin was sourced from Saba Pharmaceutical Industry. As a control for the drug, saline solution provided with the drug was administered to cells at the same concentration.

72 hours after the application of the drug or siRNA, MTT assays were initiated and imaging was performed using various microscopes. The 72-hour treatment duration was optimized in prior experiments conducted in our lab; therefore, all experiments involving mutant MCF7 cells were treated for 72 hours.

Section A4. MTT cell viability assay (carried out by Cagdas Yanyatan)

Cells were seeded onto 96-well plates in a 2,000 cells per well concentration and the siRNA treatment was performed as described. On day of the siRNA treatment, cells were introduced to increasing concentrations of doxorubicin. Since doxorubicin is dissolved in DMSO, the same amount of DMSO was added as a control. After 72 hours, an MTT assay was performed. 10 mg of MTT (Invitrogen) was dissolved in 2 ml PBS and filtered to remove undissolved particles. Old media of the cells were gently discarded and replenished with 100 μ l phenol red-free media. 10 μ l of MTT/PBS solution was added to each well. Cells were incubated at 37 $^{\circ}$ C for 4 hours and 100 μ l 10% SDS in 0.01M HCl was added on top of each well. Cells were again incubated for 16-18 hours, and colorimetric reads were taken at 570 nm using the microplate reader.

Analysis of the photographs taken after 72 hours was performed via the ImageJ application, and statistical operations were performed via Graphpad. EVOS Cell Imaging systems live-time imaging microscope and Leica DMI8 M/C/A Inverted microscope were used for these images. The effects of doxorubicin and si*CHRNA5* application were compared in WT and mutant cells. Images taken after 72 hours of application were analyzed via ImageJ, and this data were calculated and compared using the two-way anova method via Graphpad software.

Table A1 Pre-library sample QC results for first batch of doxorubicin and si*CHRNA5* treated data

#	SAMPLE ID	A260/A280	A260/A230	Spectrophotometric Conc. (ng/ul)	Fluorometric Conc. (ng/ul)	RIN
1	EV_Doxo_A5_2	1.7	2.2	64.4	48	8.7
2	EV_Doxo_sc_2	1.7	1.81	25.5	28.3	9
3	EV_Serum_A5_1	2.02	1.93	189.4	176	10
4	EV_Serum_A5_2	2.05	2.02	291	271.5	10
5	EV_Serum_sc_1	2.03	2.03	251.8	244	10
6	EV_Serum_sc_2	1.7	2.31	199.2	84.6	9
7	R175H_Doxo_A5_2	2.05	1.95	248.8	258	9.6
8	R175H_Doxo_sc1	2.05	1.42	177.2	145.5	9.6
9	R175H_Serum_A5_2	2.1	1.8	557.9	540	10
10	R175H_Serum_sc_1	2.13	1.34	544.1	331	9.5
11	R175H_Serum_sc_2	2.13	1.34	617.6	537	9.8
12	R273H_Doxo_A5_1	2.04	1.86	65.4	63.3	9.6
13	R273H_Doxo_A5_2	1.59	1.96	156.6	129	9.1
14	R273H_Serum_A5_1	2.04	2.08	387.3	39.6	10
15	R273H_Serum_A5_2	2.07	2.14	427	422	9.6
16	R273H_Serum_sc_1	2.05	2.03	250.8	260	9.2
17	R273H_Serum_sc_2	1.75	2.15	298.5	330	9.5

Table A2 Pre-library sample QC results for first batch of doxorubicin and si*CHRNA5* treated data.

#	SAMPLE ID	A260/A280	A260/A230	Spectrophotometric Conc. (ng/ul)	Fluorometric Conc. (ng/ul)	RIN
1	R175H_Serum_A5_1	-	-	-	516	10
2	WT_Doxo_A5	2.04	1.89	296	270	10
3	WT_Doxo_SC	2.04	1.87	290.3	232	9.9
4	WT_Serum_A5_1	2.02	1.74	480	436	10
5	WT_Serum_A5_2	2.07	1.49	532.9	439	10
6	WT_Serum_SC_1	2.05	2.02	607.3	520	9.8
7	WT_Serum_SC_2	2.04	2.11	649.8	510	9.8

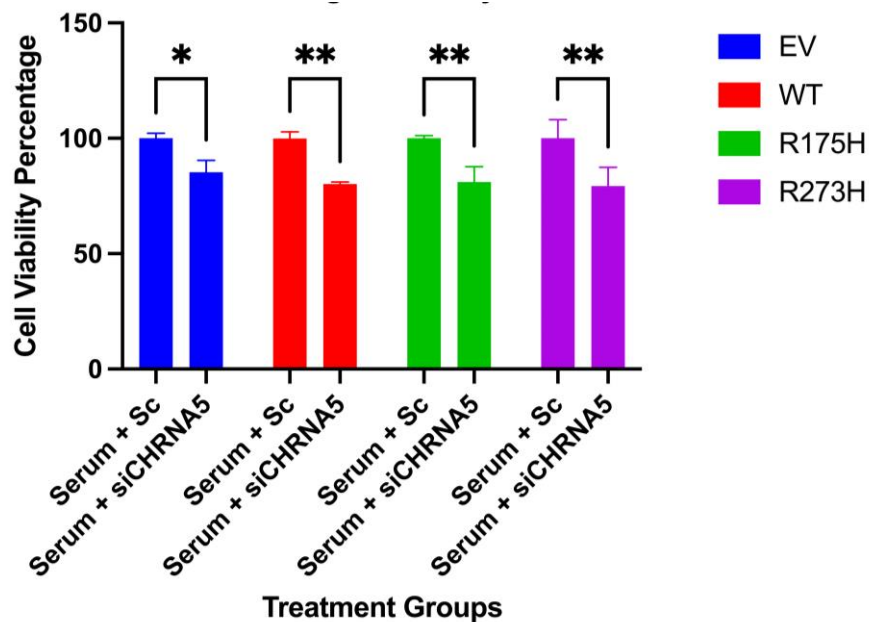


Figure A1 Cell viability using MTT assays for the siCHRNA5 treatment alone. Other than the EV, the depletion of *CHRNA5* leads to a decrease in cell viability of the mutant cell lines and the WT over expressed cell line.

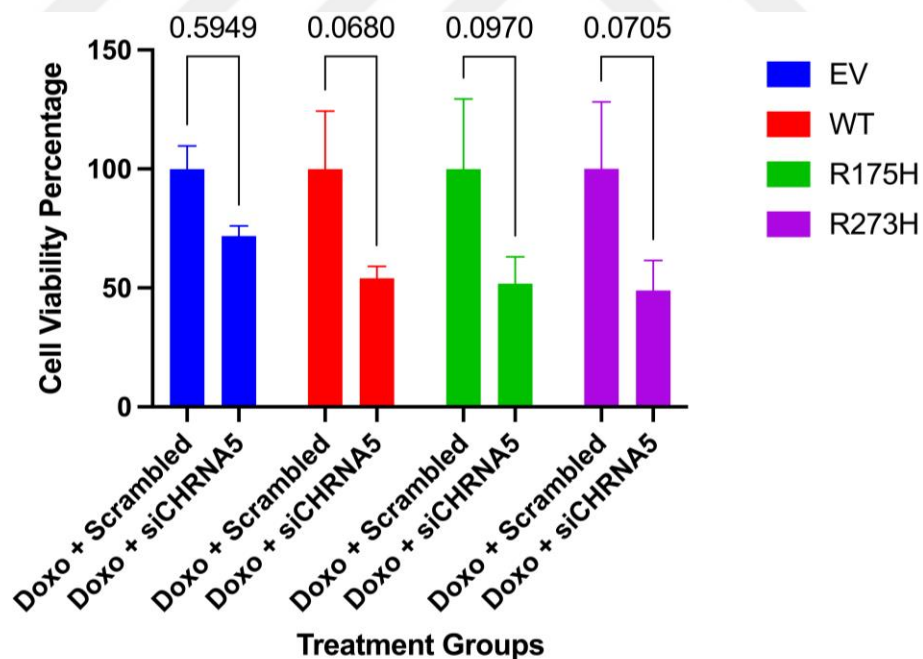


Figure A2 Demonstrates the effects of siCHRNA5 in the presence of doxorubicin. Here, the doxorubicin was set as the control and the effect of the combinatoiral treatment was seen to decrease across all the cell lines. The p-value for the EV did not appear to be significant, while for other cell lines it was below 0.1. The effect of the combinatorial treatment was seen to decrease cell viability across all cell lines, thus reversing resistance observed against Doxorubicin.

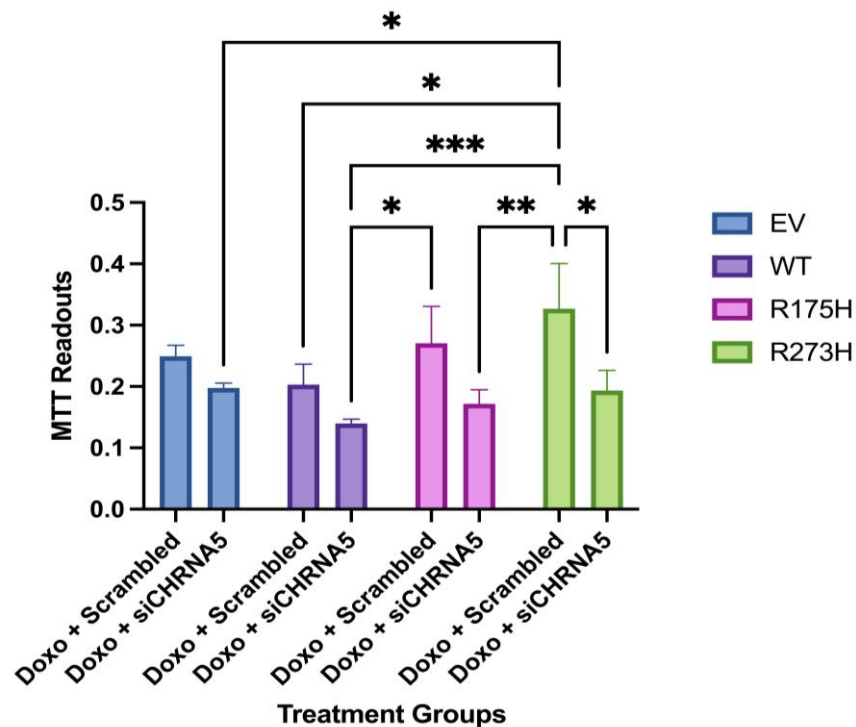


Figure A3 MTT assay raw readouts across the cell lines with doxorubicin treatment and the combinatorial treatment.

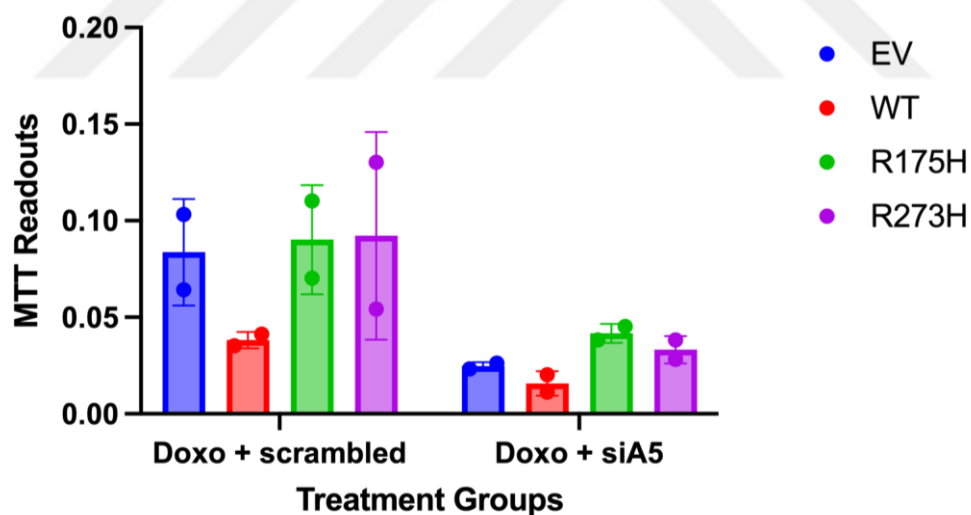


Figure A4 Two-way anova showing the MTT readouts for the doxorubicin data and its combination with siCHRNA5. Here the results show the same response as before with a decrease in drug resistance. The results for R175H show a higher readout which was also observed throughout the RNA-seq data, with R273H having a similar transcriptomic response with the WT upon the combinatorial treatment.

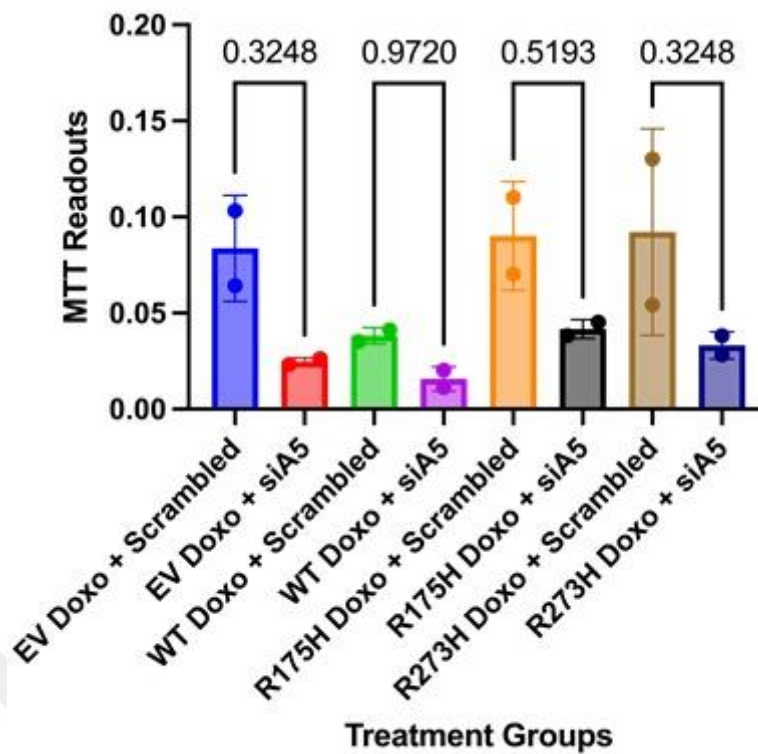


Figure A5 Another MMT assay carried on the same cell lines with the same treatment groups. P-values unfortunately do not show significance, however a consistent pattern of reduced cell viability was observed with the combinatorial treatment.

Appendix B: Exploration of .bam files in IGV

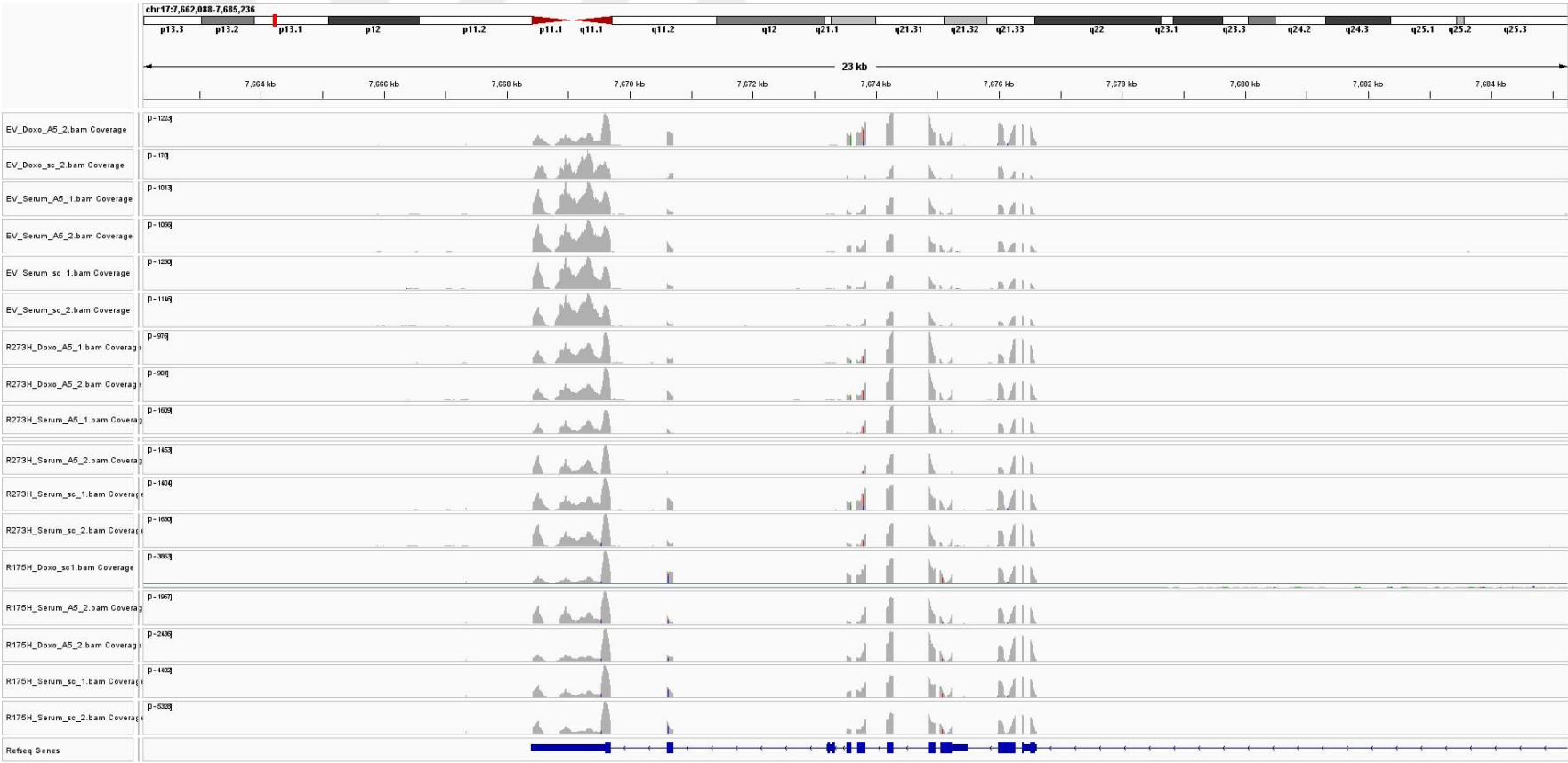


Figure B1 Coverage for all the samples for the *TP53* gene.