

A Study on RNA and Protein Effectors of Transcription Factor Activity and Gene Expression

by

Ayşe Derya Cavga

A Dissertation Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Doctor of Philosophy

in

Chemical and Biological Engineering



**KOÇ
ÜNİVERSİTESİ**

August, 2019

A Study on RNA and Protein Effectors of Transcription Factor Activity and Gene Expression

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a doctoral dissertation by

Ayşe Derya Cavga

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

Prof. Dr. Özlem Keskin Özkaya (Advisor)

Prof. Dr. Attila Gürsoy (Co-Advisor)

Assoc. Prof. Dr. Nathan Lack

Prof. Dr. İ. Halil Kavaklı

Assoc. Prof. Dr. Halil Bayraktar

Asst. Prof. Dr. Saliha Ece Özbabacan

Date

dedicated to my loving family of eleven, who made this possible



ABSTRACT

A Study on RNA and Protein Effectors of Transcription Factor Activity and Gene Expression

Ayşe Derya Cavga

Doctor of Philosophy in Chemical and Biological Engineering

August, 2019

Transcription factors can activate or repress expression of their target genes. Given this critical role, their activity is decisive in health and disease. Transcription factor function is affected by numerous factors that include co-activators, enhancer RNA, and protein oligomerization. Combining functional genomic and proteomic techniques with computational methods, the research presented here investigated how these factors influence transcription factors and modify gene regulation, and the role of a transcriptional repressor in cancer.

Using the androgen receptor (AR) as a model transcription factor, we characterized how the enhancer long non-coding RNA KLK3e affects expression of AR regulated genes. Combining in vitro RNA pulldown and mass spectrometry, we identified proteins that interact with KLK3e. These included known AR regulators such as DDX5, DDX17 and HSPA8 that were known to affect prostate cancer pathogenesis. Our results and computational modeling of the RNA-protein complex suggest that KLK3e RNA may influence AR mediated transcription through acting as a scaffold to gather AR co-activators near the transcription site. Such RNA scaffolding effects may be paralleled in other nuclear receptors and transcription factors, as DDX5 has been implicated to interact with other steroid receptors such as the estrogen and vitamin D receptors.

Next we explored how co-activator proteins influence gene transcription. We used four novel small molecule inhibitors that target different sites on the AR and disrupt co-activator binding. To identify the disrupted interactions, we conducted rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) and incorporated RNA sequencing to understand the effects on gene expression. Inhibition of either a co-activator binding site, dimerization site, DNA binding site, or ligand binding site appeared to have a similar negative impact on the number of protein interactors. Also, all of the inhibitors that target different sites affected the gene expression profile similarly and reduced transcription by AR.

To understand how transcription factor oligomerization occurs, we conducted in silico modeling with the glucocorticoid receptor (GR). Experimental observations suggested that GR interacts with the mineralocorticoid receptor (MR) and can also form homotetramers upon DNA binding. Our computational models explained the variation in experimental observations and demonstrated that GR and MR could interact through several alternative interfaces shared by close relatives of the nuclear receptor family. Predictions demonstrated that ligand or antagonist binding did not prevent interaction but rather changed the interface preference. This can leave different sites available for binding of co-activators, explaining why different ligands can produce different cellular outcomes. We also proposed a mechanism of action for GR tetramerization through the ligand binding domain, and the chromosomal looping model of interaction through the DNA binding domains.

Finally, circadian clock regulation also relies on transcriptional factors. Cryptochrome is a circadian transcriptional repressor, and its deletion was observed to delay cancer and extend the lifespan of p53 mutant mice. With RNA sequencing analyses we investigated the transcriptomic changes following UV damage response upon cryptochrome deletion in a p53 mutant background. Gene Set Enrichment Analysis of differentially expressed genes demonstrated enrichment in IFN- γ immune surveillance and TNF α signaling via NF- κ B. Protein network analysis pinpointed p21, Sirt1 and Jun as key players. Differentially expressed genes also contained a high ratio of non-coding RNAs.

In short, we show that the KLK3e enhancer RNA interacts with AR co-activators, and separately we observed that inhibition of co-activator binding can have the same effect on AR transcription factor activity as inhibition of ligand activation, DNA binding, or dimerization. Our structural models showed how GR might oligomerize with MR through alternative interfaces, and that ligand or antagonist binding can affect interface preference. We also demonstrated how cryptochrome deletion in p53 mutants enhance apoptotic and anti-tumorigenic responses to UV damage at the transcriptome level.

ÖZETÇE

Transkripsiyon Faktörü Aktivitesi ve Gen İfadesinde

RNA ve Protein Etkenleri Üzerine Bir Çalışma

Ayşe Derya Cavga

Kimya ve Biyoloji Mühendisliği, Doktora

Ağustos, 2019

Transkripsiyon faktörleri, hedef genlerin ifadesini arttırabilir veya baskılayabilir. Bu kritik rolleri sebebiyle transkripsiyon faktörlerinin aktivitesi sağlık ve hastalıkta belirleyicidir. Transkripsiyon faktörü aktivitesi, koaktivatörler, yükseltici (enhancer) RNA'lar ve protein oligomerizasyonu gibi birçok faktörün etkisi altındadır. Burada sunulan araştırmada, işlevsel genomik ve proteomik teknikleri bilgisayar analizleriyle birleştirerek, bu etkenlerin transkripsiyon faktörlerini ve gen ifadesini nasıl etkilediğini araştırdık, ayrıca bir transkripsiyonel baskılayıcının kanserdeki rolünü inceledik.

Androjen reseptörünü (AR) bir model transkripsiyon faktörü olarak kullanarak, uzun kodlanmayan yükseltici RNA KLK3e'nin AR tarafından düzenlenen gen ifadesini nasıl etkilediğini karakterize ettik. *In vitro* RNA çekimi (pulldown) ardından kütle spektrometresi uygulayarak, KLK3e ile etkileşen proteinleri belirledik. Etkileştiği tespit edilen proteinlerin DDX5, DDX17 ve HSPA8 gibi AR aktivitesini düzenlediği ve prostat kanserinde önemli roller oynadığı bilinen proteinler içermesi dikkat çekti. Sonuçlarımız ve RNA-protein kompleksinin bilgisayar modellemesi, KLK3e RNA'sının AR aktivitesinde olan etkisini, AR koaktivatörlerini transkripsiyon bölgesinde toplayan bir iskele görevi görerek gerçekleştirebileceğini gösterdi. Bu tür RNA iskele etkilerinin benzerleri, diğer nükleer reseptörler ve transkripsiyon faktörlerinde de yer alıyor olabilir. Örneğin östrojen reseptörü ve D vitamini reseptörü de RNA bağlayıcı protein DDX5 ile etkileşmektedir.

Sonra, koaktivatör proteinlerin gen transkripsiyon profilini nasıl etkilediğini araştırdık. AR'nin farklı bölgelerine bağlanan ve protein etkileşimlerini bozan dört yeni küçük molekül inhibitör kullandık. Bozulan etkileşimleri saptamak için endojen proteinlerin hızlı immünopresipitasyon kütle spektrometre analizi (RIME) ve gen ifadesinin nasıl etkilendiğini anlamak için RNA sekans analizi uyguladık. Bir koaktivatör bağlanma

bölgesinin, dimerizasyonun, DNA bağlanmasının ve ligand aktivasyonun inhibisyonunun benzer sayılarda protein etkileşimini olumsuz etkileyebildiğini gözlemledik. Ayrıca farklı noktaları etkileyen tüm inhibitörler gen ifadesi profilini de benzer şekilde etkiledi ve AR'nin transkripsiyonel aktivitesini düşürdü.

Transkripsiyon faktörü oligomerizasyonunu incelemek için glukokortikoid reseptörü (GR) ile *in silico* modellemeler yaptık. Deneysel gözlemler GR'nin mineralokortikoid reseptörü (MR) ile etkileştiğini, ve ayrıca DNA'ya bağlandığında homo tetramerler oluşturduğunu öne sürmüştü. Bilgisayar modellemelerimiz, deneysel gözlemlerdeki çeşitliliği açıklayarak GR ve MR'nin, nükleer reseptör ailesindeki yakın akrabaları tarafından paylaşılan çeşitli alternatif arayüzleri kullanarak etkileşime girebileceğini gösterdi. Modellerimiz, ligand veya antagonist bağlanmasının etkileşimi önlemediğini, fakat arayüz tercihinin değiştiğini gösterdi. Bu şekilde farklı bölgeler farklı koaktivatörlerin bağlanması için açık kalıyor olabilir, ve bu durum farklı ligandların nasıl farklı hücrel sonuçlar doğurduğunu açıklayabilir. Ayrıca ligand bağlama birimi üzerinden GR tetramerizasyonunu modelledik. DNA bağlama birimi üzerinden tetramerizasyonun kromozomal döngü modeliyle gerçekleşebileceğini gösterdik.

Sirkadiyen saat düzenlemesi de transkripsiyonel faktörlere dayanır. Kriptokrom (CRY) bir sirkadiyen transkripsiyonel baskılayıcıdır ve silinmesinin kanseri geciktirdiği ve p53 mutant farelerin ömrünü uzattığı görülmüştür. p53 mutantlarda kriptokromun silinmesi üzerine UV hasarı tepkisindeki transkriptomik değişiklikleri RNA dizileme analizi ile araştırdık. Diferansiyel olarak ifade edilen genlerin Gen Kümesi Zenginleştirme Analizi (GSEA), IFN- γ immün sürveyansında ve NF- κ B ile TNF α sinyalleşmesinde zenginleşme gösterdi. Protein ağı analizi, p21, Sirt1 ve Jun gibi yeni kilit oyuncular belirledi. Farklı düzeydeki genlerin yüksek oranda kodlamayan RNA içerdiği gözlemlendi.

Kısaca, KLK3e arttırıcı RNA'nın AR ko-aktivatörleri ile etkileşime girdiğini ve ayrıca ko-aktivatör bağlanmasının, AR transkripsiyon aktivitesi üzerindeki etkisinin ligand aktivasyonu, DNA bağlanması ve dimerizasyonun engellenmesiyle benzer olduğunu gösterdik. Yapısal modellerimiz GR'nin alternatif arayüzlerle MR ile oligomerize olabileceğini ve ligand veya antagonist bağlanmanın arayüz tercihinin etkileyebileceğini gösterdi. Ayrıca, p53 mutantlarında CRY silinmesinin UV hasarına apoptotik ve anti-tümörijenik tepkileri transkriptom düzeyinde nasıl arttırdığını gösterdik.

AKNOWLEDGEMENTS

I am very thankful to my advisors Prof. Dr. Özlem Keskin Özkaya and Prof. Dr. Attila Gürsoy for guiding me, looking out for what is best for me, for being such good and kind advisors, and allowing me the freedom to pursue a difficult project of my choice; it was a very challenging journey for all of us but I learned a lot as a scientist thanks to their trust in me. I am also incredibly thankful to Assoc. Prof. Dr. Nathan Lack for all the support, patience, knowledge, and trust he provided for my pursuit of an experimentally challenging project, and for being available for my questions even when he was halfway around the world. I am also grateful to Prof. Dr. İ. Halil Kavaklı for giving me the opportunity to be a part of the Cryptochrome and p53 project, as well as for his valuable forethought and bringing critical points into our attention on my experimental lncRNA project. I would also like to thank Assoc. Prof. Dr. Halil Bayraktar and Asst. Prof. Dr. Saliha Ece Özbabacan for sparing their valuable time to be on my jury and helping me improve my thesis. I would also like to thank Prof. Dr. Ruth Nussinov for her guidance and feedback on my projects, and the energy she radiates. I would also like to thank Assoc. Prof. Dr. Nurhan Özlü and Büşra Akarlar for their help and patience during the mass spectrometry analyses.

I am incredibly grateful for my family, for their endless love and generosity, for taking care of my son and me when I could not, and for making this thesis possible. My mother and father, my mother in law, my husband, my sister, my sister in law, my grandmother and my aunt have been so amazingly supportive throughout these years, I would never be able to thank them enough. I am also thankful to my dear son Halil for being such a good and loving boy, being the joy of my life, and lighting up my days with his hugs and smile. I am also grateful to my brother in law, for being the adoring uncle, and witty brother that he was. He left us too soon...

I am very lucky to have had incredible office and lab mates, past and present, who have become great friends and have been so supportive; Dr. Serena Muratcioğlu, Dr. Sıla Özdemir, Bilgesu Erdoğan, Ayşe Seda Yazgılı, Ali Tuğrul Balcı, Efe Elbeyli, Farideh Halakou, Dr. Emine Güven-Maiorov, Nilay Karahan, Emel Şen Kılıç, Doğuş Doğru,

Deniz Demircioğlu, Alper Başpınar, Cemal Yamak, Kayra Kösoğlu, Asma Hakouz, Meltem Eda Ömür, Damla Övek, Jackson Weako, Sukejna Valjevac, Simge Şenyüz, Zeynep Abalı, and the whole extended COSBI family... I am also very grateful for my friends from the laboratory side; Dr. Fatma Özgün, Dr. Hilal Saraç, Betül Ersoy, Ceren Şeref, Doğançan Özturan, Tunç Morova, Hazal Şentürk, Bengül Çetin, Dr. Luisa Ercolani, Dr. Deniz Uğurlu, Birsu Kölemen, Dr. Burcu Özçimen, Dr. Filiz Şenbabaoğlu, Dr. Ahmet Cingöz, Fidan Şeker, and all the extended KUTTAM family...

I acknowledge TÜBİTAK for supporting me with the 2228-B PhD scholarship (General name: 2211-Yurt İçi Lisansüstü Burs Programı) for four years of my studies, and for funding the lncRNA project with project number 116Z789. I also wish to acknowledge the Koç University Seed Fund that was granted for additional support of the lncRNA project.

TABLE OF CONTENTS

List of Tables	xiii
List of Figures	xiv
Abbreviations	xvii
Chapter 1: Introduction	18
Chapter 2: Literature Review	4
2.1 Transcription Factors	4
2.2 Prostate Cancer and the Androgen Receptor	5
2.2.1 Prostate Cancer Prevalence and Management.....	6
2.2.2 The Androgen Receptor (AR) as a Transcription Factor	7
2.2.3 Structure and domains of the androgen receptor (AR).....	9
2.2.4 Targeting the Androgen Receptor with Small Molecules	10
2.2.5 Key Androgen Receptor Regulated Genes.....	14
2.2.6 The method of rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) and its implementation on the androgen receptor (AR)	17
2.3 Long Non-Coding RNAs and KLK3e	19
2.3.1 Various Roles of Long Non-Coding RNAs.....	19
2.3.2 Long Non-Coding RNA Xist.....	20
2.3.3 Long Non-Coding RNA KLK3e	21
2.3.4 Methods to Study Protein Interactions of lncRNAs	26
2.4 Oligomerization Potentials of the Glucocorticoid Receptor	28
2.4.1 Interactions of the Glucocorticoid and Mineralocorticoid Receptors	30
2.4.2 Glucocorticoid Receptor Tetramerization	31
2.4.3 Oligomerization potentials of other steroid receptors	35
2.4.4 Glucocorticoid Receptor and Prostate Cancer.....	36
2.5 Circadian Clock and Cancer	37
2.5.1 Circadian Clock and the Transcription Factors Involved.....	37
2.5.2 Circadian Clock and Cancer	37
2.5.3 Protein-Protein Interaction Network Analysis	39
Chapter 3: Methods	40
3.1 Identifying protein interactors of the lncRNA KLK3e	41
3.1.1 Reverse-Transcription PCR amplification of KLK3e lncRNA and controls	42
3.1.2 In vitro Transcription of KLK3e lncRNA and controls, with Biotin tags	45
3.1.3 Testing biotinylated RNA binding by Streptavidin Magnetic Beads	46
3.1.4 DHT treatment of cells	47
3.1.5 Nuclear protein extraction from LNCaP cells	48
3.1.6 RNA Pulldown of nuclear protein lysates with KLK3e lncRNA and controls	49
3.1.7 Mass Spectrometry Identification of Pulled-down Proteins.....	51
3.1.8 Modeling of the KLK3e RNA-protein complex	52
3.2 Characterization of Novel Androgen Receptor Antagonists and their Role in Transcription.....	57

3.2.1	Data Analysis of Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) results	57
3.2.2	RNA Sequencing Data Analysis	58
3.2.3	Gene Set Enrichment Analysis (GSEA) of AR targeting molecules	59
3.3	Protein Interaction and Oligomerization Analyses of Glucocorticoid Receptor	60
3.3.1	PRISM computational modelling of protein-protein interactions	60
3.3.2	Energy calculations, hot spot and mutation analyses	61
3.4	RNA-seq analysis of Cytochrome deletion in a p53 mutant background	62
3.4.1	RNA sequence Alignment & Differential Analysis	62
3.4.2	Identification of differentially expressed genes (DEGs)	63
3.4.3	Gene Set Enrichment Analysis (GSEA) of DEGs in UV response in p53-/-Cry1-/-Cry2-/- cells compared to p53-/-cells.....	64
3.4.4	KEGG Pathway Enrichment Analysis.....	65
3.4.5	Protein-Protein Interaction Network Analysis	65
	Chapter 4: Understanding the Role of lncRNA in Androgen Receptor Mediated Transcription	67
4.1	PCR Amplification of KLK3e and controls	67
4.1.1	Sequence confirmation of amplified KLK3e and analysis of matches	68
4.1.2	Sequence confirmation of amplified Xist transcript.....	71
4.2	In vitro Transcription of KLK3e lncRNA and controls.....	72
4.2.1	DTT addition significantly improves in-vitro transcription	72
4.2.2	Optimizing the amount of RNA Polymerase.....	74
4.2.3	Confirming specific binding of biotinylated RNA to streptavidin beads	75
4.2.4	RNA gel showed specific in vitro transcription of KLK3e and control RNAs	76
4.3	Optimization of RNA pulldown from cell lysates	76
4.4	Optimized RNA pulldown from lysate showed protein bands specific for Xist and KLK3e	77
4.5	Mass Spectrometry shows prostate cancer and AR-related proteins bound to KLK3e lncRNA	78
4.6	RNA- protein complex model and proposed mechanism of action.....	85
4.7	Discussion.....	87
4.8	Future Perspectives	88
	Chapter 5: Characterization of Novel AR Antagonists and their Role in Transcription.....	90
5.1	Data analysis of rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) to detect small moleculeinhibited protein interactions of AR	91
5.2	All small molecules tested downregulate key AR-regulated genes.....	93
5.3	Tested molecules generally counteract the effect of DHT on gene expression	94
5.4	Effect of small molecules on AR Hallmark genes.....	95

5.5	A closer look at 22 genes common between DHT-upregulated and AR Hallmark genes	96
5.6	Gene Set Enrichment Analysis (GSEA) of genes affected by each drug shows greatest effect on androgen response genes	97
5.7	Tested small molecules effect common & distinct genes	99
5.8	Discussion & Future Perspectives	101
Chapter 6: Oligomerization Potentials of the Glucocorticoid Receptor.....		102
6.1	Interaction of Glucocorticoid and Mineralocorticoid Receptors	102
6.1.1	Structural predictions support multiple alternative interfaces for MR-GR interaction at the DNA-binding domain (DBD).....	103
6.1.2	Structural predictions support multiple alternative interfaces for MR-GR interaction at the Ligand Binding Domain (LBD) as well	105
6.1.3	GR and MR LBDs can still interact when bound by different ligands, but their interface preference may be affected.....	106
6.2	Glucocorticoid Receptor Tetramerization	109
6.2.1	Glucocorticoid Receptor Tetramerization at the Ligand Binding Domain	109
6.2.2	Glucocorticoid Receptor tetramerization through DNA Binding Domain possible with chromatin looping	110
6.3	Discussion.....	111
6.4	Future Perspectives	112
Chapter 7: Transcriptome analysis of Cytochrome Deletion in p53 Mutants		113
7.1	Confirmation of higher apoptosis in Cry deletion in a p53 mutant background.....	113
7.2	RNA-seq data analysis showed high unique mapping and coverage	114
7.3	Differentially Expressed Genes were confirmed via qRT-PCR	115
7.4	Identification of differentially expressed genes (DEGs) in p53 ^{-/-} Cry1 ^{-/-} Cry2 ^{-/-} cells compared to p53 ^{-/-} cells in response to UV	117
7.5	Gene Set (GSEA) and KEGG Enrichment Analyses Identified Pathways that may Enhance Apoptosis in p53 ^{-/-} Cry1 ^{-/-} Cry2 ^{-/-} MSF Cells upon Exposure to UV	118
7.6	Protein-protein interaction network (PPIN) analysis helps identify key protein interactions responsible for delayed onset of cancer	122
7.7	Correction of zero division errors allows gain of information on non-coding genes	124
7.8	Discussion.....	127
Chapter 8: Conclusion.....		129
Bibliography.....		131
Resume & Publications		13143
Appendix A.....		144
Appendix B.....		158
Appendix C.....		178
Appendix D.....		182

LIST OF TABLES

Table 3-1: PCR primers with T7 promoter extensions	43
Table 4-1: Sequences producing significant matches to the KLK3e transcript.....	69
Table 4-2: Sequences producing significant matches to the Xist transcript.	72
Table 4-3: 41 proteins that specifically bind to KLK3e..	79
Table 5-1: Numbers and compositions of genes affected by each drug, compared to DHT control.....	99



LIST OF FIGURES

Figure 2.1: Estimated number of new cancer cases and deaths in men in 2018, worldwide and in Turkey.....	6
Figure 2.2: The transcription factor androgen receptor (AR) can be activated through androgen dependent or independent pathways.....	8
Figure 2.3: Gene locus and domains of the transcription factor androgen receptor (AR).	10
Figure 2.4: Small molecules targeting different sites on the transcription factor androgen receptor (AR).	11
Figure 2.5: Endogenous AR interacting proteins identified with RIME.	18
Figure 2.6: The various roles of lncRNAs.	19
Figure 2.7: The KLK3e enhancer lncRNA is transcribed from the enhancer/promoter region upstream of the KLK3/PSA gene.	22
Figure 2.8: RT-qPCR shows KLK3e expression increases with DHT treatment in correlation with KLK3.	22
Figure 2.9: Effect of KLK3e silencing on expression of AR target genes, RT-qPCR.	23
Figure 2.10: KLK3e fragments interact with AR and MED1.	24
Figure 2.11: Silencing of KLK3e, followed by Chromatin immunoprecipitation and qPCR (ChIP-qPCR).	24
Figure 2.12: A proposed model for the functioning of the enhancer lncRNA KLK3e	25
Figure 2.13: Capturing endogenous RNA–protein complexes using RAP-MS and ChIRP-MS.	26
Figure 2.14: Published examples of lncRNA Pulldown experiments.	27
Figure 2.15: The superfamily of steroid nuclear receptors and the structural organization of their functional domains.	28
Figure 2.16: GR is dimeric at the nucleoplasm but mainly tetrameric when bound to DNA at the MMTV array.	32
Figure 2.17: Model proposed for formation of GR tetramers.	33
Figure 2.18: Tetrameric human GR LBD crystal structure with PDB ID 1p93, and the two dimer interfaces within.	34
Figure 2.19: Oligomerization potentials of other steroid receptors.	35

Figure 3.1: Project flowchart to identify protein interactors of lncRNA KLK3e....	41
Figure 3.2: Modeling steps of the proposed RNA-protein complex.	53
Figure 3.3: The PRISM: Protein Interactions by Structural Matching tool.....	60
Figure 3.4: Experimental setup and RNA-Seq analysis methodology.	63
Figure 4.1: KLK3e transcript of ~2kb was confirmed to be at the RNA level.....	67
Figure 4.2: KLK3e transcript is of similar sizes with Xist and CherryGFP controls.	68
Figure 4.3: Optimization of in Vitro Transcription additives and clean-up.	73
Figure 4.4: Optimization of the amount of RNA Polymerase.	74
Figure 4.5: Confirmation of specific binding of biotinylated lncRNA on Streptavidin magnetic beads.	75
Figure 4.6: RNA gel shows efficient and clean in vitro transcription of KLK3e and controls Xist and CherGFP.	76
Figure 4.7: Optimized RNA pulldown showed protein bands specific to KLK3e and Xist.....	77
Figure 4.8: Protein-protein interactions within proteins identified to interact with KLK3e.	81
Figure 4.9: RPS4Y1 expression is high in prostate tissue and prostate cancer.	84
Figure 4.10: Proposed RNA-protein model; the KLK3e RNA being transcribed from the enhancer/promoter region upstream of the KLK3/PSA gene can act as a scaffold to bring co-activators of AR near this transcription site.....	86
Figure 5.1: Similarities and differences of drug effects on AR interactions.	91
Figure 5.2: Small molecules tested downregulate key AR regulated genes at levels comparable to control drug Enzalutamide.	93
Figure 5.3: Tested small molecules generally counteract the effect of DHT in DHT regulated genes.	94
Figure 5.4: Effect of small molecules on AR Hallmark genes..	95
Figure 5.5: Effect of small molecules on AR Hallmark genes.	96
Figure 5.6: Gene Set Enrichment Analysis (GSEA)	97
Figure 5.7: Number of genes significantly affected by each drug, compared to DHT control.	100
Figure 6.1: Structural prediction for MR and GR interaction through the classical dimerization domain in the D-loop of the DNA binding domain.....	103
Figure 6.2: GR-MR DNA binding domain interactions structural predictions.	104

Figure 6.3: As indicated by experiments, GR-MR ligand binding domains can interact through alternative interfaces, which are from the same receptor family.	105
Figure 6.4: Binding energy scores of computational predictions of alternative interfaces between GR and MR with different agonist and antagonist combinations and conformations.	107
Figure 6.5: Example interaction structures of MR and GR LBDs when both are agonist bound, and when GR or MR is antagonist bound.	108
Figure 6.6: PRISM predictions suggest alternative tetramerization modes of GR LBD may be possible.....	109
Figure 6.7: PRISM predictions suggest GR tetramerization can be supported through the DBD as well through chromosomal looping.....	110
Figure 6.8: GR tetramerization structural model.....	111
Figure 7.1: Validation of RNA-seq data with qRT-PCR.....	116
Figure 7.2: GSEA and KEGG Pathway analysis of 626 DEGs.....	119
Figure 7.3: RNA FPKM levels of key genes from gene set enrichment analysis MSigDB hallmark gene sets and PPI network analysis.....	121
Figure 7.4: Protein-protein interaction network analysis reveals key genes that interact with p53 associated genes.....	123
Figure 7.5: Elimination of transcripts with zero FPKM causes loss of information, especially on non-coding transcripts.	126
Figure 7.6: Proposed pathway model for involvement of novel candidate genes in the delayed onset of cancer in p53 ^{-/-} Cry1 ^{-/-} Cry2 ^{-/-} mice compared to p53 ^{-/-} mice.	128

ABBREVIATIONS

AR	Androgen Receptor
ARE	Androgen Response Element
BES	Binding Energy Score
BF3	Binding Function 3
CRPC	Castration Resistant Prostate Cancer
CRY	Cryptochrome
DBD	DNA Binding Domain
DHT	Dihydrotestosterone
DNA	Deoxyribonucleic Acid
DTT	Dithiothreitol
ESR	Estrogen Receptor
FDA	U.S. Food and Drug Administration
GFP	Green Fluorescent Protein
GR	Glucocorticoid Receptor
GSEA	Gene Set Enrichment Analysis
IgG	Immunoglobulin G
KLK	Kallikrein
LBD	Ligand Binding Domain
MED1	Mediator 1
MR	Mineralocorticoid Receptor
NES	Nuclear Export Signal
NLS	Nuclear Localization Signal
NTD	N Terminal Domain
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PER	Period
PR	Progesterone Receptor
PRC	Polycomb Repressive Complex
PSA	Prostate Specific Antigen
RIME	Rapid Immunoprecipitation Mass spectrometry of Endogenous proteins
RIP	RNA immunoprecipitation
RNA	Ribonucleic Acid
RT	Reverse Transcription
TF	Transcription Factor
UV	UltraViolet
XIST	X-inactive specific transcript

Chapter 1:

INTRODUCTION

Transcription factors are a family of proteins that bind DNA and control the rate of transcription. These proteins play a key role in both normal physiology and many different diseases. The four projects described in this thesis characterize different transcription factors and their behaviors in cells. The first three projects involve nuclear steroid receptors including androgen receptor (AR) and glucocorticoid receptor (GR), while the fourth project involves Cryptochrome (CRY), which is a transcriptional repressor that regulates the circadian clock. The specific projects are:

Project One: Understanding the role of lncRNA in AR-mediated transcription:

In this project we focused on the long non-coding RNA KLK3e, which was previously proposed to interact with AR and alter transcription of canonical AR-regulated genes including KLK3, which encodes the prostate specific antigen (PSA). To better understand this mechanism, we performed *in vitro* pulldown of KLK3e followed by mass spectrometry, and identified the protein interactors of this lncRNA. We identified various RNA binding proteins including several that had been previously shown to be AR co-activators, and computationally modeled part of the RNA-protein complex. This work provides new insight into how enhancer RNA can bring together co-activators of a transcription factor to influence gene expression.

Project Two: Characterization of Novel AR Antagonists and their Role in Transcription: In this project we investigated the effects of four novel AR antagonists that each work through different targets on the AR. After treatments with these small molecules, the interactome and downstream transcriptome of AR were quantified using rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) and RNA sequencing. We observed that inhibition of ligand binding, co-activator binding, and DNA binding can have similar effects on the AR interactome and AR mediated gene expression. Inhibition of DNA binding and dimerization also had similar effects.

Project Three: Oligomerization potentials of glucocorticoid receptor (GR):

In this work we investigate both the oligomerization state of GR with the mineralocorticoid receptor (MR), and its ability to self tetramerize. Experimental findings implicated GR can interact with MR through alternative interfaces. The literature also proposes that GR tetramerizes upon DNA binding. Our computational analyses provided structural explanations on how these oligomerizations can occur. The results presented here may have implications for other steroid receptors such as the androgen receptor (AR) and progesterone receptor (PR), which are closely related members of the same steroid receptor family with high structural homology. This is also demonstrated by the observation that GR, MR and PR are able to bind androgen response elements on DNA and can substitute for AR in some cases of prostate cancer.

Project Four: Transcriptome Analysis of Cryptochrome Deletion in p53 Mutants:

We investigated the transcriptomic changes that explain the observation that deletion of the circadian transcriptional repressor Cryptochrome in p53 mutant mice delays the onset of cancer. Our RNA-seq analysis demonstrated the roles of pro-apoptotic and anti-tumorigenic players that were previously unidentified, such as Sirt1 and p21 that work together with p73, which was previously identified, as well as drawing attention to the possible role of non-coding RNAs.

In summary, the first project demonstrated that the enhancer RNA KLK3e can interact with several co-activators of AR, explaining its influence on AR-mediated transcription. The second project showed that AR regulated transcription can be inhibited through disrupting interactions of this transcription factor, as well as its activation and dimerization. The third project presented structural explanations to experimental observations on GR; explaining how it can interact with MR through alternative interfaces, and how it can tetramerize. And the final project brought novel players into attention on the relationship between Cryptochrome deletion and increased apoptosis in a p53 background, such as p21, Sirt1 and non-coding RNAs.

Chapter 2 is composed of a Literature Review covering the relevant aspects of transcription factors, prostate cancer, androgen receptor signaling, small molecule

targeting of AR, long non coding RNAs including Xist and KLK3e, GR oligomerization, circadian clock, and its relation to cancer.

Chapter 3 describes the methods utilized to obtain the results described in the subsequent chapters.

Chapter 4 shows the results of the project on understanding the role of lncRNA in AR-mediated transcription using in-vitro KLK3e pulldown and mass spectrometry. Identified interactors include known co-activators of AR such as DDX5 and HSPA8. The RNA-protein complex between KLK3e, DDX5 and AR has been modeled.

Chapter 5 displays the changes novel small molecules induce in the AR interactome and AR regulated transcriptome. Targeting AR at different sites was shown to disrupt AR-mediated gene expression marks, showing efficiencies of novel molecules comparable to the FDA approved AR ligand binding inhibitor.

Chapter 6 exhibits the structural models that explain alternative integration modes between GR and MR, as well as GR tetramerization. Structural models demonstrate how alternative binding sites between GR and MR are possible, and how GR can tetramerize.

Chapter 7 demonstrates the transcriptomic effects of Cryptochrome deletion in a p53 mutant background, and brings into attention key pro-apoptotic and anti-tumorigenic players such as p21 and Sirt1.

Chapter 8 includes the conclusion with an overview of the contributions.

Appendix A contains supplementary results from the KLK3e lncRNA interactors project. **Appendix B** includes supplementary methods and results from the project on small molecules targeting AR. **Appendix C** shares supplementary results on the GR oligomerization project, which are experimental results obtained by our experimental collaborator. **Appendix D** contains supplementary methods and results from the project on RNA-Seq analysis of Cryptochrome deletion in p53 mutant background.

Chapter 2:

LITERATURE REVIEW

2.1 *Transcription Factors*

Transcription factors (TFs) are proteins that bind to specific DNA sequences to regulate gene expression at the loci. The DNA sequence specifically recognized and bound by a transcription factor is called a response element or transcription factor binding site (Wang 2005). In addition to DNA binding domains, TFs also have regions that interact with RNA polymerase II or other transcriptional regulators, which allows them to control the level of transcription of a gene into messenger RNA. Through these crucial roles, TFs can control embryonic development, drive diseases and cancer (Lee and Young 2013), regulate immune responses (Singh et al. 2014), cell differentiation and even de-differentiation and trans-differentiation (Takahashi and Yamanaka, 2016). General transcription factors are found ubiquitously in almost all cells of an organism, while specific transcription factors are only expressed in specific tissues or developmental stages (Latchman 1997).

Transcriptional activators turn gene expression on, while transcriptional repressors turn gene expression off. The expression of most genes are controlled by several different transcription factors. Especially eukaryotic gene expression requires a specific combination of TFs to reach maximal expression levels. For example, a gene may only be transcribed if activators X and Y are present and if repressor Z is absent. This is referred to as combinatorial regulation of transcription (Reményi et al. 2004).

Transcription factors are classified into families based on their DNA binding domain motifs and structures. The main transcription factor families in eukaryotes include (Johnson 1989; Lambert et al. 2018);

- Homeodomain proteins encoded by homeobox genes,
- Basic Helix-Loop-Helix (bHLH)
- Helix-Turn-Helix (HTH)
- Basic Leucine Zipper (bZIP)
- Cys₂His₂ Zinc Finger (ZF)

- Nuclear Hormone Receptor (NHR) Zn_2/Cys_8 Zinc Finger

The largest group of transcription factors in eukaryotes is the nuclear receptor superfamily of structurally related proteins (Mazaira et al. 2018). The DNA binding domains of nuclear receptors are especially highly conserved and have specialized Zinc Finger motifs. This family is distinctive in the sense that they are ligand-dependent transcription factors; upon binding their ligands such as steroid hormones, they are activated and bind DNA to act as transcriptional regulators of gene expression (Evans 1988). Steroid hormones sensed by nuclear hormone receptors include the sex hormones androgen, estrogen and progesterone, as well as glucocorticoids, mineralocorticoids, thyroid hormones, and vitamin D3 (Olefsky 2001).

Of steroid receptors, the glucocorticoid receptor (GR) is expressed in almost every cell type of vertebrates to control metabolism, immune response, and development. The sex hormone receptors, on the other hand, are more tissue-specific and are mostly expressed in gender specific tissues. The androgen receptor, for example, plays a critical role in prostate cancer. These receptors will be examined in more detail in the coming sections.

The circadian clock is also regulated by transcription factors. The primary feedback loop that forms the oscillatory pattern of circadian gene expression consists of the CLOCK and BMAL1 transcriptional activators that belong to the basic helix-loop-helix (bHLH) transcription factor family, and the PER and CRY transcriptional repressor proteins. CLOCK and BMAL1 form a heterodimer in the cytoplasm and translocate to the nucleus to activate the transcription of the core clock *period* (PER1, PER2, and PER3) and *cryptochrome* (CRY1 and CRY2) genes. PER and CRY heterodimers then repress their own transcription by inhibiting the activity of the CLOCK and BMAL1 complex, forming a negative feedback loop. This oscillatory pattern of gene expression repetitively goes up and down with a ~24 hour period, and regulates the expression of many genes important in night and day behavior and metabolism (Lowrey and Takahashi 2004).

2.2 Prostate Cancer and the Androgen Receptor

2.2.1 Prostate Cancer Prevalence and Management

Prostate cancer is the second most widespread type of cancer among men in Turkey and the world, totaling about 1.3 million new cases and causing about 360 thousand deaths in 2018 alone worldwide, as represented in the graphs below in Figure 2.1 (Bray et al. 2018).

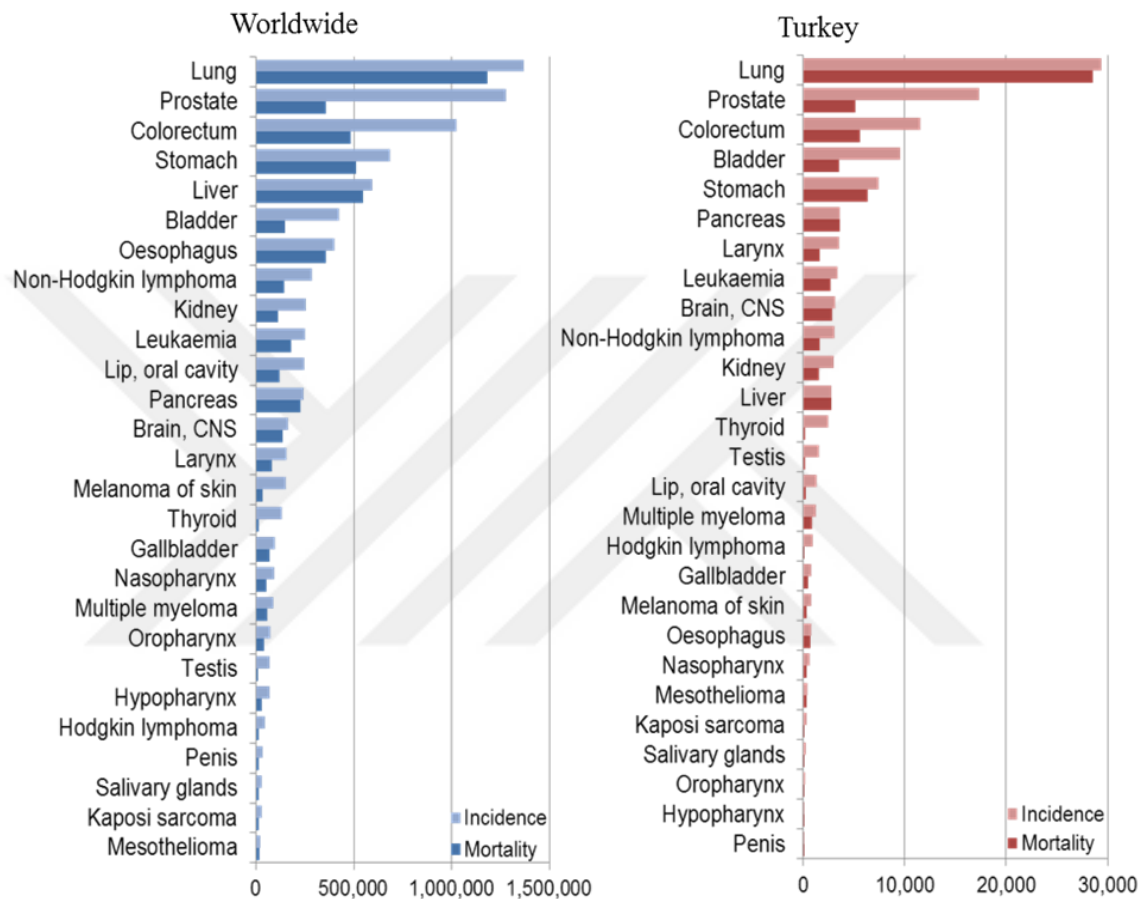


Figure 2.1: Estimated number of new cancer cases and deaths in men in 2018, worldwide and in Turkey. Prostate cancer is the second most common type of cancer among men in Turkey and the world. Data obtained from GLOBOCAN 2018; <http://gco.iarc.fr/today/online-analysis-table> (Bray et al. 2018)

The mortality to incidence ratio in prostate cancer is around 30% in Turkey and 28% worldwide; in other words, approximately 1 in 3 men diagnosed with prostate cancer do not survive it. Its high occurrence rate makes it the fourth leading cause of cancer-related deaths among men in Turkey, and fifth worldwide. (Bray et al. 2018) Prostate cancer most commonly occurs in men over the age of 65, and can severely lower the

quality of life, cause pain, and can become more life-threatening if it metastasizes, most commonly to the bones or lymph nodes. (Ruddon & Raymond, 2007)

Treatment approaches to prostate cancer vary depending on the stage of the cancer at diagnosis, and can include medicinal management, surgery, or radiotherapy (Trachtenberg and Blackledge 2002).

Since androgen receptor (AR) activity plays a major role in the proliferation of prostate cells, metastatic prostate cancer treatment often consists of anti-androgen or androgen repression therapy. Even though this treatment is initially effective, a relapse is common due to the emergence of castration resistant prostate cancer cells, often within 2 to 3 years (Chandrasekar et al. 2015). These cells can proliferate independent of the amount of androgens available to them, through pathways that may include or bypass androgen receptor activity. AR dependent proliferation can occur in the absence of androgens through expression of AR mutants or variants, or activation of AR through androgen-independent pathways, such as the MAPK or Akt pathways, as shown in Figure 2.2. Activation of AR independent survival and apoptosis pathways can also play a role in castration resistant prostate cancer, though AR signaling usually plays an important part (Debes & Tindall, 2004, Nelson et al., 2003).

2.2.2 The Androgen Receptor (AR) as a Transcription Factor

The androgen receptor is a classical steroid hormone receptor. In its inactive form, the androgen receptor resides in the cytoplasm, bound in a protein complex with chaperone proteins such as heat shock protein 90 (HSP90) (Ni et al. 2013). Upon binding its ligand androgens such as testosterone, or the more potent form dihydrotestosterone (DHT), AR dimerizes, detaches from HSP90, and is imported into the nucleus where it binds androgen response elements (AREs) of its target genes, and acts as a transcription factor to control gene expression, as also depicted in Figure 2.2 below (Mooradian et al., 1987). Among the genes regulated by AR, is the kallikrein-related peptidase (KLK3) gene, which encodes the common prostate cancer biomarker prostate specific antigen (PSA), among other AR-regulated genes, some canonical examples of which will be described later in more detail.

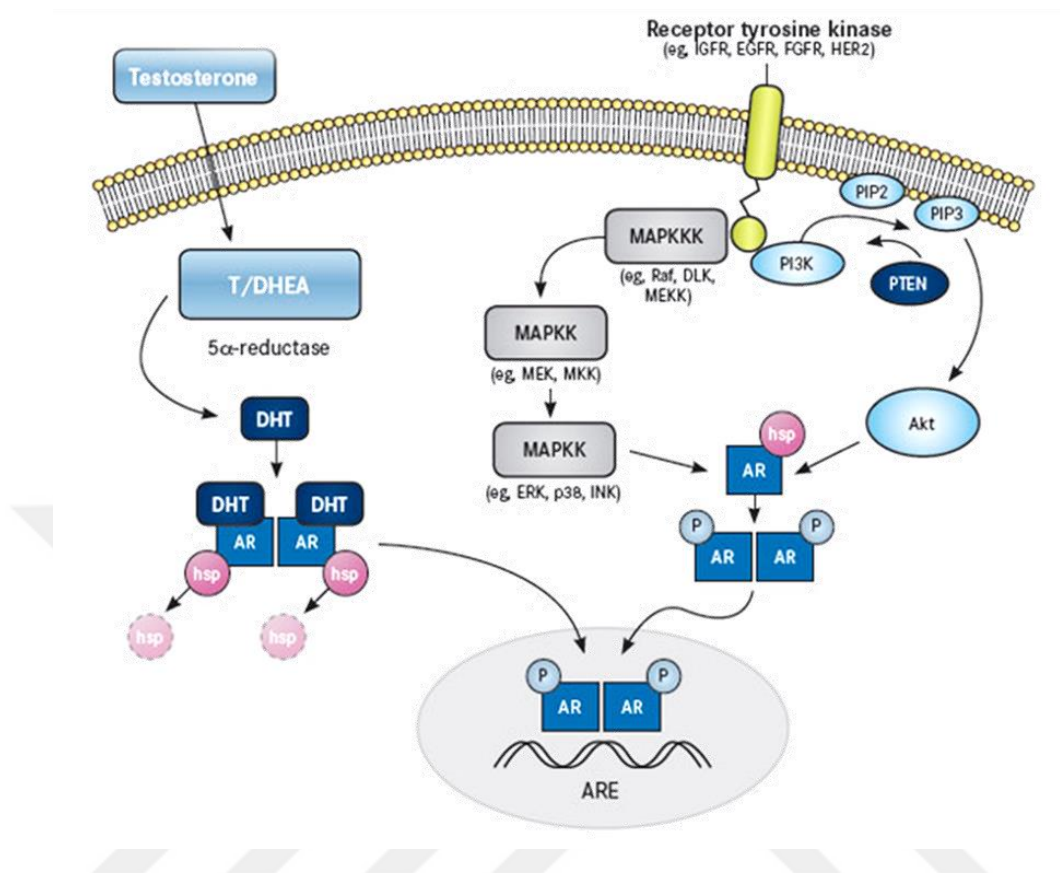


Figure 2.2: The transcription factor androgen receptor (AR) can be activated through androgen dependent or independent pathways. Upon binding its ligand androgens such as dihydrotestosterone (DHT), or being activated through MAPK or AKT signaling, AR dimerizes and localizes to the nucleus, where it binds androgen response elements (AREs) of its target genes, and activates transcription. Figure adapted from (Girling et al. 2007), taken from <https://bit.ly/2Sg0FOF>.

2.2.3 Structure and domains of the androgen receptor (AR)

As depicted in Figure 2.3, the androgen receptor (AR) gene is localized on the long arm of the X-chromosome, and is made up of 8 exons, that encode the androgen receptor protein of 919 amino acids. AR protein structure consists of the N-terminal transcriptional regulation domain, the DNA binding domain (DBD), and the ligand binding domain (LBD). (Davey and Grossmann 2016) This structural organization is seen in other members of the nuclear steroid receptor family as well, such as glucocorticoid receptor (GR), mineralocorticoid receptor (MR), and progesterone receptor (PR), some of which will be studied in the later sections.

The DNA binding domain is highly conserved among steroid receptor family proteins, which contain two zinc fingers that recognize specific consensus sequences on the DNA, called androgen response elements (AREs) in the case of AR. The ligand binding domain is also similar among steroid receptors, and it is where ligands such as androgens bind, and are stabilized through interactions with the NTD. The LBD is also responsible for the interaction of AR with chaperone proteins such as heat shock proteins. The N-terminal domain of the AR is the most variable region, and contains the activation function-1 (AF1) region, which is necessary for the full transactivation of AR. (Jenster et al. 1991; Davey and Grossmann 2016) Crystal structures are available for the DNA binding and ligand binding domains of AR, as well as other steroid receptors, but not the N-terminal domain.

As shown in Figure 2.3, the ligand binding domain contains a second transcriptional activating function (AF2) which is ligand dependent, and takes part in coregulatory binding and interactions with the N-terminal domain. Between the DNA binding domain and the hinge region lies the nuclear localization signal (NLS) for translocation of AR into the nucleus by importins upon being activated by ligand binding and detaching from cytoplasmic chaperones. The ligand binding domain, on the other hand, contains a nuclear export signal (NES) which is required for the translocation of AR back to the cytoplasm after ligand is removed. (Davey and Grossmann 2016)

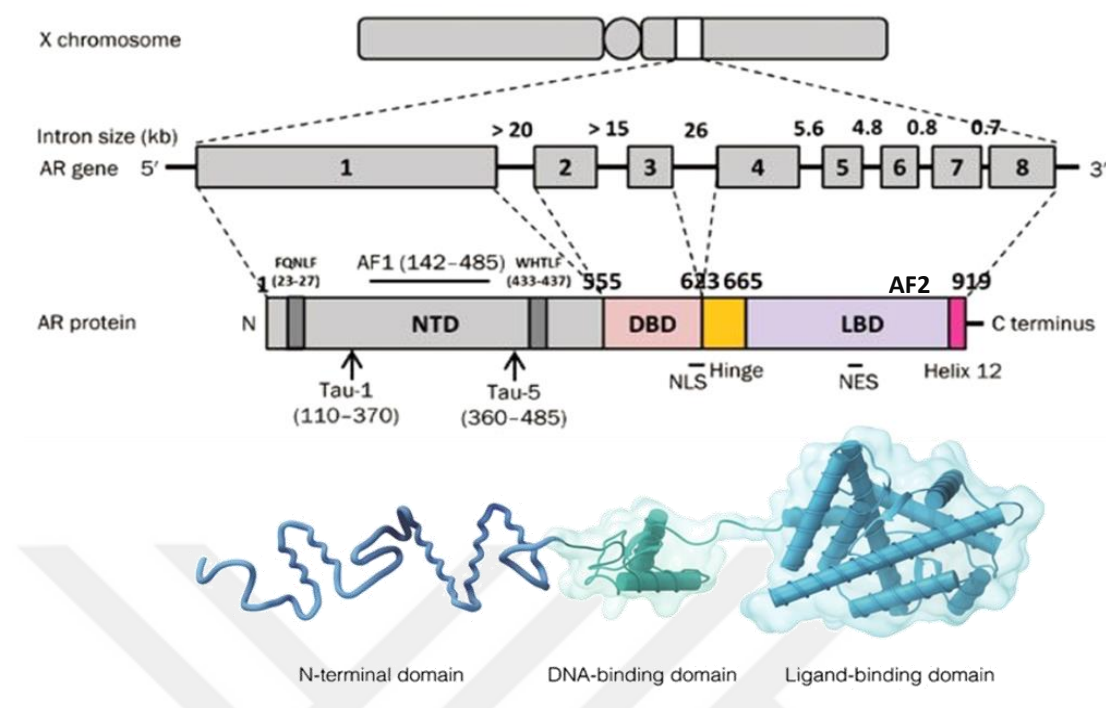


Figure 2.3: Gene locus and domains of the transcription factor androgen receptor (AR). The AR gene is on the X chromosome and encodes a protein of 919 amino acids with 3 main structural domains. The N-terminal domain (NTD), DNA binding domain (DBD), Ligand binding domain (LBD), transcriptional activating function 1 (AF-1), transcriptional activating function 2 (AF-2), nuclear localization signal (NLS), nuclear export signal (NES) and transactivation units (Tau-1 and Tau-5) . (Davey and Grossmann 2016) Figures from: (Tan et al. 2014; Antonarakis et al. 2016)

2.2.4 Targeting the Androgen Receptor with Small Molecules

As mentioned earlier, treatment methods of prostate cancer include androgen deprivation therapy, antiandrogen treatment, chemotherapy, surgical intervention, and radiotherapy. Androgen deprivation therapy fails to be effective after 2 to 3 years, as castration resistant prostate cancer develops. Until 2011, castration resistant prostate cancer (CRPC) was generally treated with chemotherapy, using docetaxel. Then, anti-androgenic oral drugs Abiraterone and Enzalutamide were approved for CRPC treatment. However, these drugs can fail to be effective in some patients who do not respond or develop resistance. Abiraterone is a progesterone derivative. The details below will focus on molecules targeting the AR, and development of alternatives to Enzalutamide.

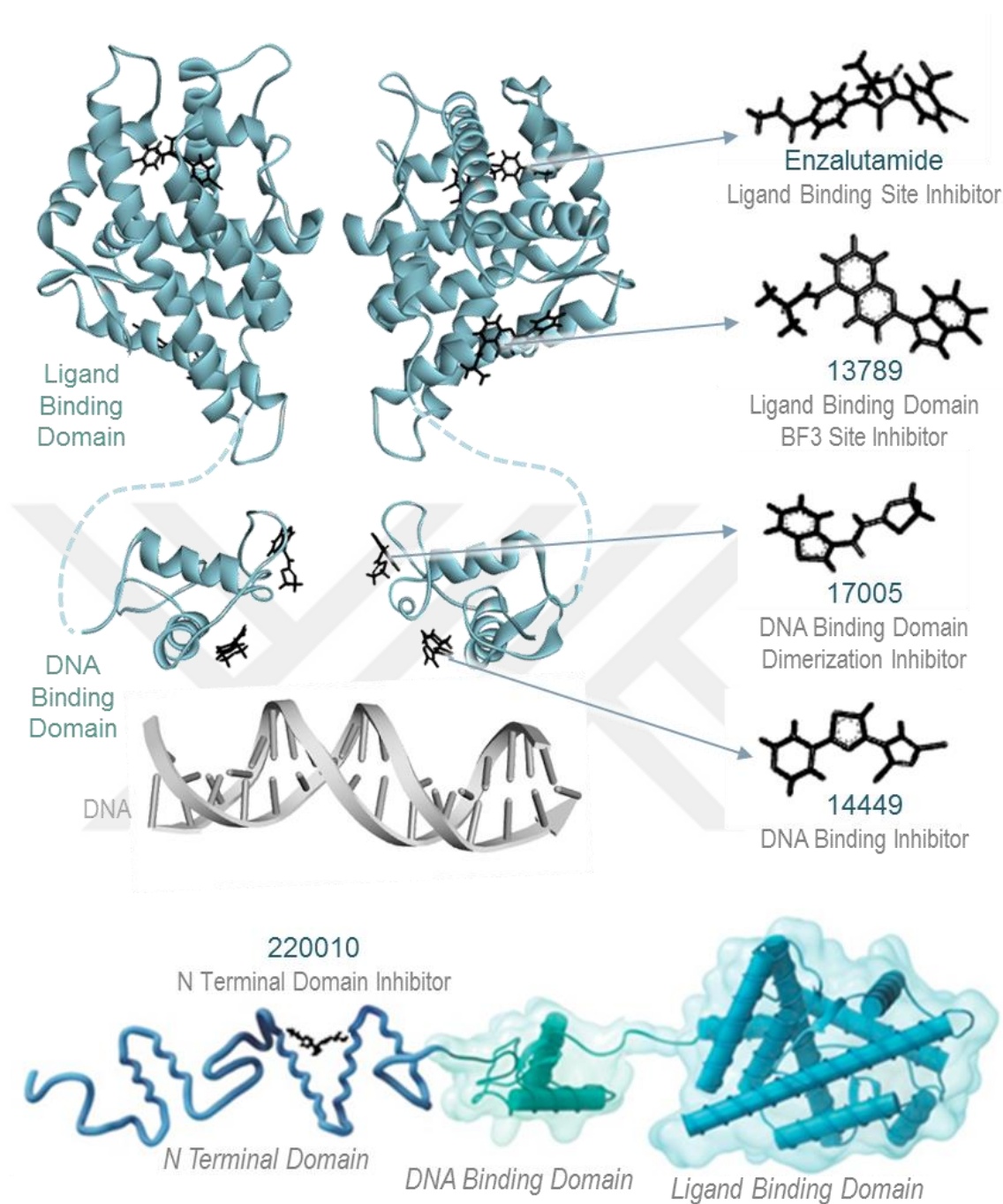


Figure 2.4: Small molecules targeting different sites on the transcription factor androgen receptor (AR). AR ligand binding domain and DNA binding domain crystal structure PDB IDs are 5JJM and 1R4I, respectively. Computational dockings of the small molecules were performed at the Vancouver Prostate Centre. N terminal domain is not crystallized; representative figure is adapted from (Antonarakis et al. 2016). Enzalutamide is an FDA approved AR antagonist, the other four are novel small molecules described in the literature summarized below.

Enzalutamide: An approved ligand binding pocket inhibitor

Enzalutamide (initially named MDV3100) was developed as a competitive inhibitor of the androgen receptor that binds the ligand-binding site. It was shown to inhibit AR nuclear localization, cofactor binding, as well as DNA binding. In a clinical trial of CRPC patients who had not received chemotherapy, Enzalutamide treatment was shown to improve survival rates, decrease radiographic progression and delay the start of chemotherapy in metastatic prostate cancer patients. (Tran et al. 2009; Scher et al. 2010; Beer et al. 2014)

Enzalutamide also outperformed Bicalutamide, another nonsteroidal antiandrogen, with reduced risk of progression and death in castration resistant prostate cancer (CRPC) (Penson et al. 2016). The FDA initially approved Enzalutamide for the treatment of metastatic CRPC, and later expanded this approval for non-metastatic CRPC as well. Enzalutamide is now commonly used for the treatment of prostate cancer worldwide when resistance to androgen deprivation therapy develops, enabling postponement of chemotherapy.

As mentioned above, Enzalutamide binds the ligand binding site of AR, directly competing with ligand binding. However, since resistance can be observed to this approach, scientists were motivated to develop small molecules that target different regions of the AR, which could be used in conjunction with, instead of, or after drugs like Enzalutamide in case of their failure.

Resistance to drugs targeting the ligand binding pocket occur when mutations in the binding pocket left the drugs ineffective, or through the emergence of splice variants that are constitutively active without a ligand binding domain (Taplin et al. 2002; Dehm et al. 2008; Guo et al. 2009; Sun et al. 2010). Hence the small molecules described below either target a different region of the LBD than the binding pocket, or the DBD or NTD. The results section will demonstrate the effects of these inhibitors on AR interactions through rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) analysis, and on gene expression mediated by AR through RNA sequencing analyses.

Small molecule 13789: LBD BF-3 site

Molecule 13789 targets the novel binding function 3 (BF3) site of the AR LBD, which plays a role in its transcriptional activity. It was developed based on compounds that showed significant antiandrogen effects in LNCaP and Enzalutamide resistant cells, through reduction of PSA expression and cell growth (Lack et al. 2011; Gao et al. 2013).

Small molecule 17005: DBD dimerization inhibitor

With the aim of targeting splice variants of AR that lack the LBD as well as full-length AR, molecule 17005 targets the DNA binding domain. Since dimerization was observed to be essential for the transcriptional activity of variants as well as full length AR, (Xu et al. 2015) this molecule selectively targets the DBD dimerization interface of AR. Molecule 17005 was shown to suppress the expression of AR target genes KLK3 (PSA), TMPRSS2, and FKBP5 induced by the synthetic androgen R1881, while not affecting a myc-regulated control gene. The molecule was also shown to specifically target the DBox site (AR 596–600) that controls DBD dimerization, and disrupt dimerization induced FRET signal. (Dalal et al. 2018)

Small molecule 14449: DBD DNA binding inhibitor

Molecule 14449 was designed to target a binding site on the AR DBD. It was optimized to reach an anti-AR effectiveness comparable to Enzalutamide, and was shown to block the interaction of AR with chromatin and reduce cell proliferation of Enzalutamide resistant cells, perhaps through its ability to impede transcriptional activity of LBD truncated AR variants such as AR variant 7. Its efficacy was also shown in vivo through reduction of gene expression and tumor volume in mouse xenografts. (Dalal et al. 2014; Li et al. 2014a)

Small molecule 220010: Targeting the N terminal domain

Molecule 220010 targets the NTD of AR which is critical for its transactivation, and mediates interactions of AR with transcriptional coregulators (Antonarakis et al. 2016).

2.2.5 Key Androgen Receptor Regulated Genes

Several genes are considered canonical androgen response genes. The genes described below all belong to the HALLMARK_ANDROGEN_RESPONSE gene set of the MSigDB (Subramanian et al. 2005), and are also among the genes observed to be under direct androgen regulation in the project described here titled “characterization of novel AR antagonists and their role in transcription,” as well as in the literature, which are described below.

KLK3 / PSA

The KLK3 gene encodes the Prostate Specific Antigen (PSA), also called kallikrein-3. PSA is a serine protease of the kallikrein-related peptidase family. Prostate gland epithelial cells secrete this enzyme to help liquefy semen (Lilja 2003).

PSA is used commonly as a biomarker for prostate cancer in blood tests of men over the age of 40. While the blood serum of healthy males also contains low levels of this antigen, it is often elevated in men with malignancies or other diseases of the prostate. However, the use of PSA as a biomarker is an issue of debate, as it is not a very specific marker, and can be elevated in cases of benign prostate enlargement or inflammation of the prostate gland. (Djulbegovic et al. 2010; Velonas et al. 2013; Ilic et al. 2018)

The KLK3 gene is a canonical example of an AR regulated gene, with several upstream androgen response elements (AREs). Two classical AREs reside in the proximal promoter; AREI, 156 bp upstream, and AREII, 400 bp upstream. Binding of these two response elements is enough to induce standard levels of gene expression at this locus (Cleutjens et al. 1996). The enhancer region of the KLK3e gene that is further upstream also comprises AREIII, 4134 bp upstream, several non-consensus androgen response elements, and binding sites for cofactors of the androgen receptor. Binding of these regions allow maximal gene expression at this region in a prostate specific manner. (Zhang et al. 1997; Keeton et al. 2007)

KLK2

The KLK2 gene encodes the kallikrein-2 protein. Similar to KLK3, KLK2 expression is also very specific to prostate tissue. This gene is also clustered on chromosome 19 with other members of the kallikrein family. This serine protease enzyme is responsible for activating the PSA (KLK3) from its pro-form by cleaving arginine residues (Kumar et al. 1997; Takayama et al. 1997).

TMPRSS2

Transmembrane protease, serine 2 (TMPRSS2) gene encodes a serine protease family protein, which are associated with cellular processes important in health and disease. TMPRSS2 was observed to be upregulated by androgens in cancers of the prostate, and downregulated in prostate cancer cells that have become independent of androgens, i.e. castration resistant. A fusion gene between TMPRSS2 and ERG is observed in about 50% of prostate cancer cases. The TMPRSS2-ERG fusion gene disrupts healthy androgen receptor signaling causing cancer progression. The fusion gene was shown to interfere with prostate lineage-specific differentiation, and trigger de-differentiation of prostate cells through EZH2 (Yu et al. 2010).

FKBP5

FK506 binding protein 5 (FKBP5) belongs to the immunophilin protein family, which function in immune regulation, protein folding and transport. FKBP5 was initially observed to take part in glucocorticoid, progestin and androgen signaling pathways, and affect prostate health, through its regulatory effect on the cellular translocation of steroid hormone receptors between the cytosol and nucleus (Davies et al. 2002). In prostate cancer cell lines, treatment with androgens causes a prompt upregulation of the FKBP5 gene (Nelson et al. 2002). However, unlike previously identified AR regulated genes, FKBP5 does not have a consensus AR binding site at its proximal promoter. Instead, FKBP5 is regulated by androgens through an interaction of AR with an enhancer within the FKBP5 gene (Magee et al. 2006).

STEAP4

Six-Transmembrane Epithelial Antigen of the Prostate 4 (STEAP4) is also known as six-transmembrane protein of prostate 2 (STAMP2) (Korkmaz et al. 2005) or TNF α -induced adipose-related proteins (TIARP) (Moldes et al. 2001). STEAP4 is a transmembrane protein that is a metalloredutase that can reduce iron and copper. STEAP4 is modulated in response to inflammation, and glucose and fatty acid metabolism (Grunewald et al. 2012) .

STEAP4 was observed to be involved in healthy functioning of prostate tissue, as well as prostate malignancies. The expression of STEAP4 was shown to be specifically regulated by androgens in the AR positive prostate cancer cell line called LNCaP, but not in the AR negative cell lines DU145 and PC3. Prostate cancer patient tissue showed that STEAP2 is overexpressed in prostate tumors compared to normal tissue. And ectopic expression of this protein was shown to increase proliferation and colony formation in prostate cancer cells. (Korkmaz et al. 2005)

SGK1

Serum/glucocorticoid-induced protein kinase-1 (SGK-1) is a serine/threonine kinase that is intensely controlled by various stimuli including glucocorticoids, serum, insulin and other growth factors. The signaling pathway leading to SGK1 induction includes PI3K and mTORC2. SGK1, in turn, regulates important functions in the cell such as hormone release, transport, inflammation, neuroexcitation, cell death and proliferation. It regulates several ion channels and ion carriers (Lang and Shumilina 2013).

SGK1gene expression is highly upregulated by androgens, in an AR dependent manner. AR was shown to interact with several positions on the proximal and distal regions of the SGK1 promoter, inducing SGK1 expression upon androgen treatment. A positive feedback loop was observed between SGK-1 expression and AR regulated gene expression. SGK-1 overexpression was able to rescue cells from AR knock-down induced apoptosis, and SGK-1 silencing reduced AR transactivation. (Shanmugam et al. 2007)

2.2.6 The method of rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) and its implementation on the androgen receptor (AR)

The rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) method for analysis of chromatin complexes (Mohammed et al. 2016) is a method developed to identify interactors of a protein without the need for a protein tag or overexpression, which can disrupt true endogenous behavior. The method works especially well for transcription factors.

RIME method utilizes addition of formaldehyde onto cells prior to nuclear protein extraction in order to fixate close protein interactions. Its water solubility and fast permeation into cells and subcellular compartments render formaldehyde easy to use in such applications. More importantly, formaldehyde is small (2.3–2.7 Å) and has no spacer arm. Therefore, it specifically cross-links amino acids that are very close. Also, transient protein-protein interactions are also preserved upon formaldehyde treatment owing to its quick reaction time. (Mohammed et al. 2016)

The protocol states that as few as 10^6 cells maybe enough for successful analysis, although a starting amount of 2 to 10×10^7 cells are recommended for a protein not previously optimized. Cells are fixated using media with 1% (vol/vol) formaldehyde and cross-linked at room temperature for 8 minutes. After this, nuclear extraction is performed. (Mohammed et al. 2016)

Antibodies specific to the target protein of interest are conjugated onto protein beads, and mixed with nuclear lysates from formaldehyde cross-linked cells. The antibodies bind the target protein and its fixed interactors. Washes are performed to eliminate non-specific, unbound proteins. This is followed by on-bead digestion of proteins, desalting, LC-MS/MS and data analysis. (Mohammed et al. 2016)

In order to understand specificity, it is recommended to use at least three biological replicates, technical replicates of the MS, as well as nonspecific IgG as a negative control sample, in place of the antibody targeting the protein of interest, which are subtracted from the target RIME results. (Mohammed et al. 2016)

After the development of this method, it was applied to many transcription factors, such as the androgen receptor. The publication “Endogenous androgen receptor proteomic profiling reveals genomic subcomplex involved in prostate tumorigenesis” (Stelloo et al. 2018) identified 66 proteins in the interactome of AR, which are shown in Figure 2.5 below.

The authors note that some interactors had to be eliminated due to interaction with IgG negative control. One such example was FOXA1, which is a known AR interactor. FOXA1 interacted with AR in their RIME application, but a statistical enrichment over IgG was not detected, and therefore it was eliminated. The interaction between AR and FOXA1 was observable when a co-immunoprecipitation was performed, so the authors noted that this approach was prone to some false negativity. (Stelloo et al. 2018)

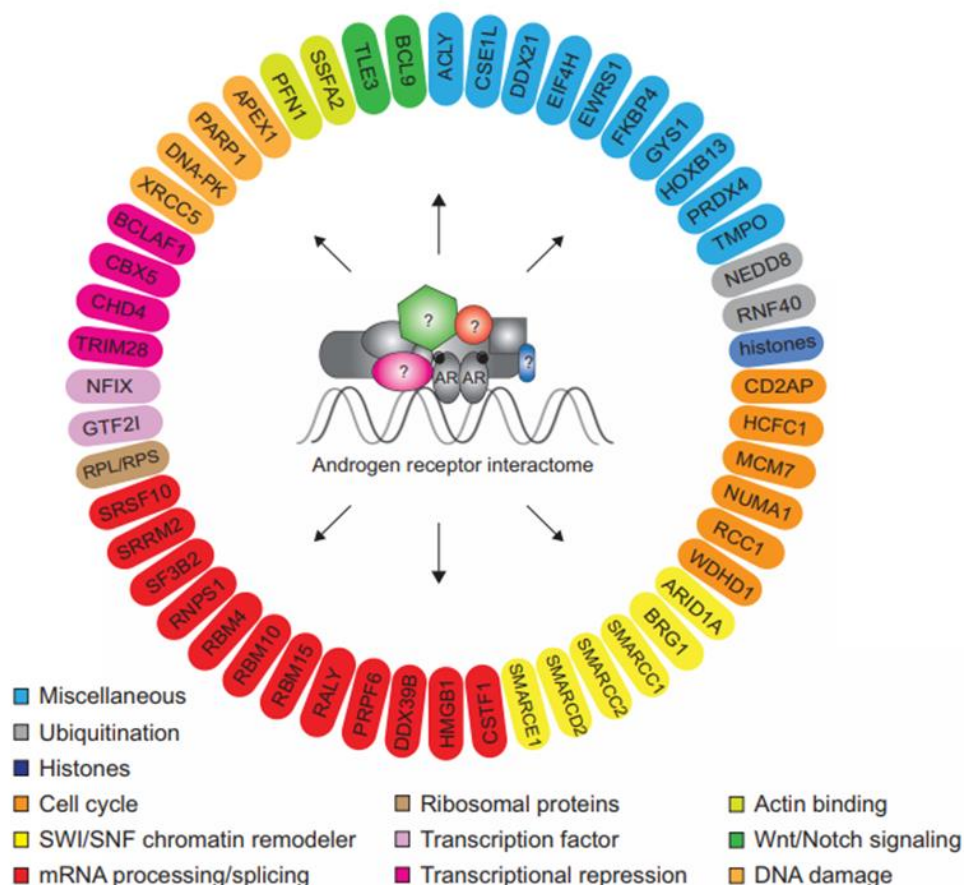


Figure 2.5: Endogenous AR interacting proteins identified with RIME. AR RIME analysis detected both known interactors such as ARID1a, HOXB13, HSP90, FOXA1, PARP1, and novel AR interactors such as TLE3 and TRIM28 (Stelloo et al. 2018).

2.3 Long Non-Coding RNAs and KLK3e

2.3.1 Various Roles of Long Non-Coding RNAs

In 2012, the ENCODE project of the National Institutes of Health (NIH) revealed that more than 70% of the human genome is expressed as non-coding RNAs that are not translated into proteins. Non-coding RNAs that are more than 200 nucleotides long have been defined as long non-coding RNAs (lncRNA). (Mouraviev et al., 2015) Interestingly, these RNAs have much higher species- and tissue-specific expression levels compared to protein-coding messenger RNAs; about one-third seem to have arisen within the primate lineage, and a large fraction of tissue-specific lncRNAs are expressed in the brain (Derrien et al., 2012).

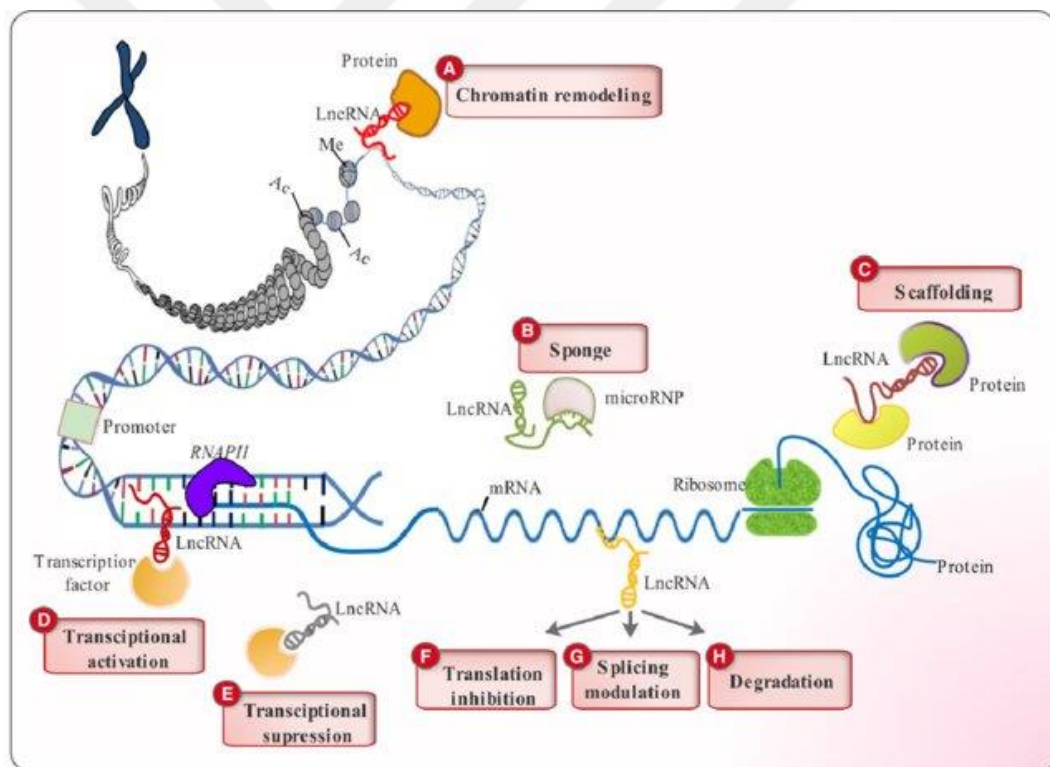


Figure 2.6: The various roles of lncRNAs. Since their identification, lncRNAs have been shown to be involved in important cellular processes such as chromatin remodeling, transcriptional activation and suppression, translation inhibition, RNA splicing and degradation, scaffolding and acting as a sponge to sequester RNA and proteins (Salehi et al. 2017).

Very few of these lncRNAs have since been functionally characterized. It has only been recently understood that lncRNAs can take on a very wide spectrum of important cellular roles, owing to their ability to form three dimensional structures and interact with proteins, DNA and other RNA, acting as scaffolds for various complexes. As demonstrated in Figure 2.6, lncRNAs have regulatory roles both at the transcriptional and translational levels, and play critical roles in cancer since they can influence the levels of many proteins, from tumor suppressors to oncogenes. (Mouraviev et al., 2015)

One of the first major publications on how important an impact lncRNAs can have on cell fate was published in Nature in 2011 (Guttman et al. 2011), and demonstrated that lincRNAs act in the circuitry controlling pluripotency and differentiation. This paper showed that some lncRNAs were directly regulated by pluripotency factors such as OCT4, SOX2 and NANOG. In turn, knocking down lncRNAs in mouse embryonic stem cells (mESC) affected pluripotency at levels comparable to knockdown of ESC regulators, causing escape from pluripotent state or upregulation of differentiation programs. The paper also demonstrated that lncRNAs bind to chromatin reader, writer and eraser proteins, indicating their role in chromatin modification and gene expression. (Guttman et al. 2011)

2.3.2 Long Non-Coding RNA Xist

Perhaps the most dramatic and well-characterized lncRNA function is the role of X-inactive specific transcript (Xist) lncRNA in inactivation of one of the two X chromosomes in cells of female placental mammals. (Penny et al. 1996; Pinter et al. 2012) This is also called dosage compensation, as the level of expression from two copies of X chromosome is brought to the same level as autosomal chromosomes.

Xist lncRNA is encoded by a gene on the X chromosome. Xist is expressed from one of the two X chromosomes in female mammals, and is responsible for its inactivation. In fact, expression of Xist in another chromosome causes its condensation and inactivation. (Penny et al. 1996) Xist was shown to promote chromosome condensation and X inactivation through the recruitment of the Polycomb Repressive Complex (PRC) (Wang et al. 2001) which includes Suz12, EED or Ezh1/2. The Repeat-A region of Xist

at the C-terminal domain was shown to be responsible for binding Suz12, a component of the PRC (Dorsselaer et al. 2010).

Xist has also been established to interact with heterogeneous nuclear ribonucleoprotein K (hnRNP-K) (Pintacuda et al. 2017), heterogeneous nuclear ribonucleoprotein U (hnRNP-U) (Hasegawa et al. 2010; Chu et al. 2015) and the nuclear matrix protein MATRIN-3 (Chu et al. 2015).

Since Xist was a well-studied lncRNA, it was also used in the development of methodologies to understand lncRNA function. These methods called RAP-MS (McHugh et al. 2015) and Chirp-MS (Chu et al. 2015) will be described later in detail.

2.3.3 Long Non-Coding RNA KLK3e

In a 2014 publication, scientists have demonstrated that a long non-coding RNA, termed KLK3e, is expressed from the enhancer region upstream of the canonical AR-regulated gene KLK3, which encodes the prostate specific antigen (PSA) (Hsieh et al. 2014). Such lncRNAs located at enhancers are also called enhancer RNAs, hence the title of the publication; “Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation.”

The paper initially demonstrates that a lncRNA is being transcribed from the enhancer region upstream of KLK3, as depicted in Figure 2.7, and that the levels of this RNA is highly correlated with KLK3 and KLK2 expression in LNCaP cells (Hsieh et al. 2014), as partially shown in Figure 2.8. KLK2 is a gene that is also in the kallikrein locus, downstream of KLK3, and they are often expressed together, as explained in the key AR-regulated genes section. In this paper, KLK3e lncRNA is shown to increase KLK2 gene expression through enabling an interaction between the KLK3 enhancer and the more distal KLK2 promoter. This enhancer RNA KLK3e contains the core enhancer element derived from the androgen response element III (ARE III), and it is observed that this element is necessary for interaction with AR and Mediator 1 (MED1), regulating AR dependent gene expression. (Hsieh et al. 2014)

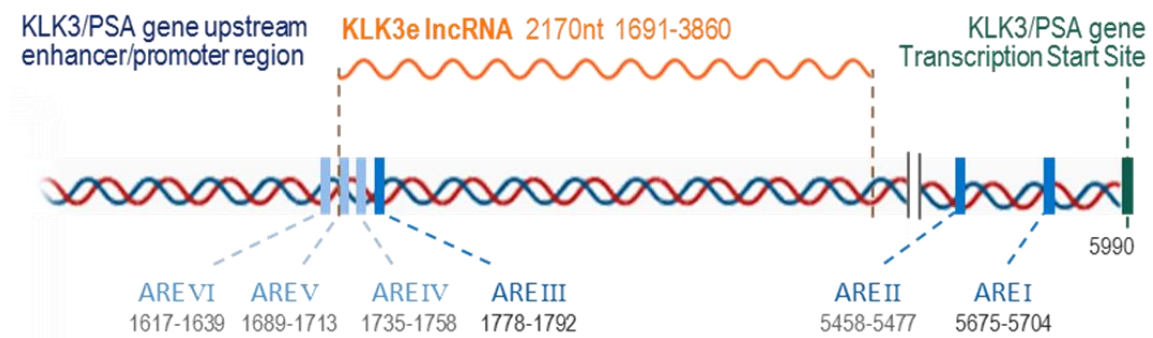


Figure 2.7: The KLK3e enhancer lncRNA is transcribed from the enhancer/promoter region upstream of the KLK3/PSA gene. Androgen Response Elements (AREs) are shown. Importantly, KLK3e carries the core enhancer element derived from ARE III, which is required for the interaction of AR and Mediator 1, as described later. KLK3e sequence from (Hsieh et al. 2014). Sequence match obtained using NCBI Blast tool (Altschul et al. 1990), ARE site information obtained from NCBI record NG_042790.1.

KLK3e lncRNA is upregulated by DHT, in correlation with KLK3

One of the initial main findings in the paper is that KLK3e expression increases with DHT treatment in a time-dependent manner, and in correlation with the level of KLK3, (Hsieh et al. 2014) as shown in Figure 2.8 below.

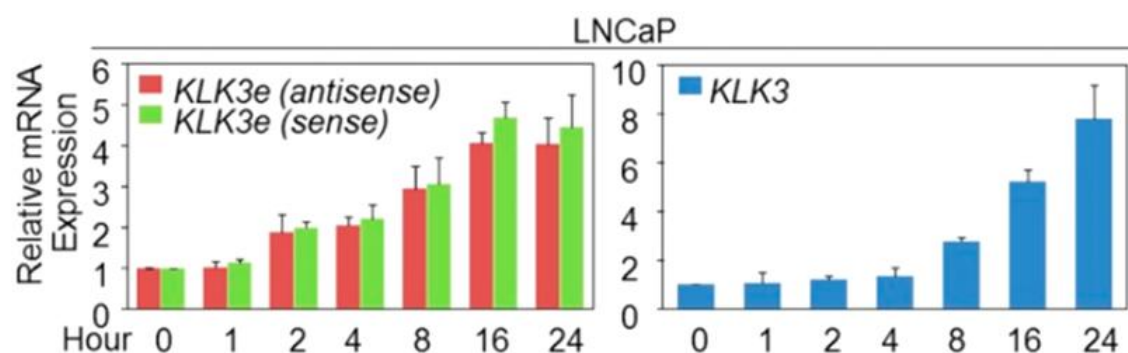


Figure 2.8: RT-qPCR shows KLK3e expression increases with DHT treatment in correlation with KLK3. Antisense KLK3 eRNA (red), sense KLK3 eRNA (green), KLK3 (blue) upon 10nM DHT treatment in a time-dependent manner in LNCaP cells (Hsieh et al. 2014).

KLK3e lncRNA selectively regulates the KLK locus and other AR target genes

The paper also investigated the effect of KLK3e silencing with siRNA on the expression of genes regulated by AR, using RT-qPCR. As shown in Figure 2.9 below, KLK3e was shown to selectively influence AR regulated genes on the KLK locus and other chromosomes (Hsieh et al. 2014).

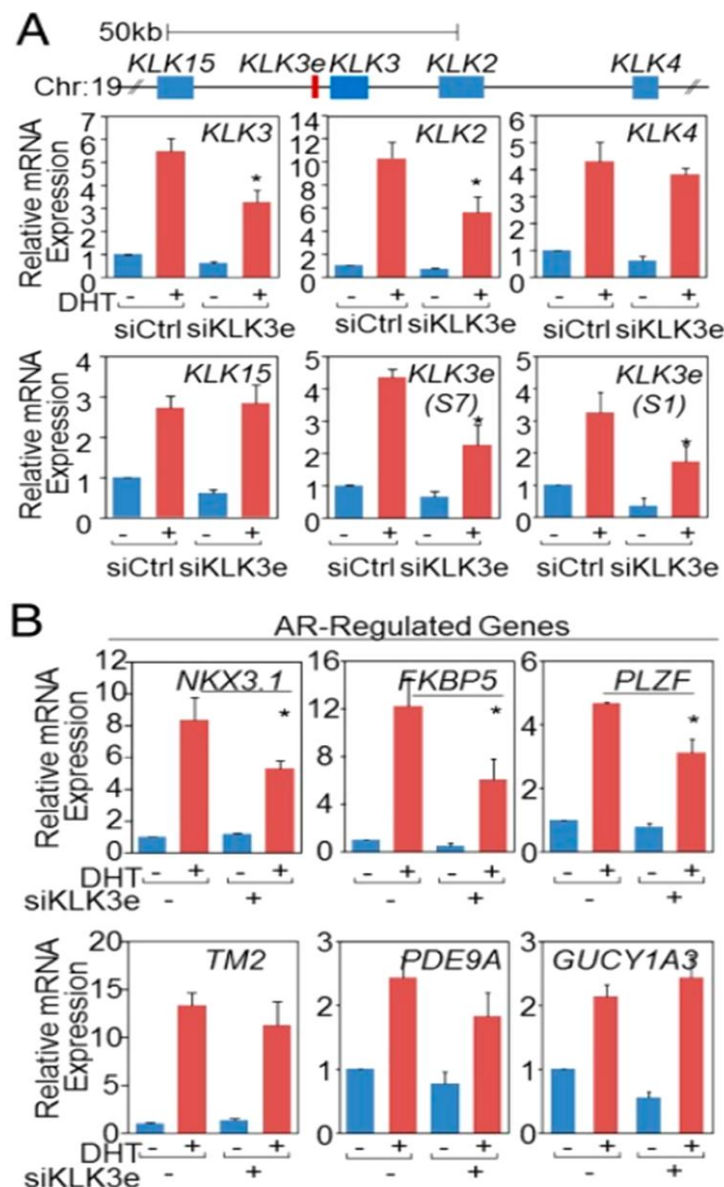


Figure 2.9: Effect of KLK3e silencing on expression of AR target genes, RT-qPCR.

A) KLK3e silencing with siRNA was shown to cause a reduction in KLK3 and KLK2 levels, while not significantly affecting KLK4 and KLK15. B) KLK3e silencing selectively affected AR regulated genes, reducing NKX3.1, FKBP5, and PLZF, whereas TM2, PDE9A, and GUCY1A3 expression was not affected (Hsieh et al. 2014).

RIP-qPCR Assay shows KLK3e interaction with AR and Med1

As shown in Figure 2.10, an RNA Immunoprecipitation qPCR (RIP- qPCR) analysis where they immunoprecipitated AR and performed RT-qPCR on the bound RNA showed KLK3e RNA was among RNA interacting with AR protein (Hsieh et al. 2014).

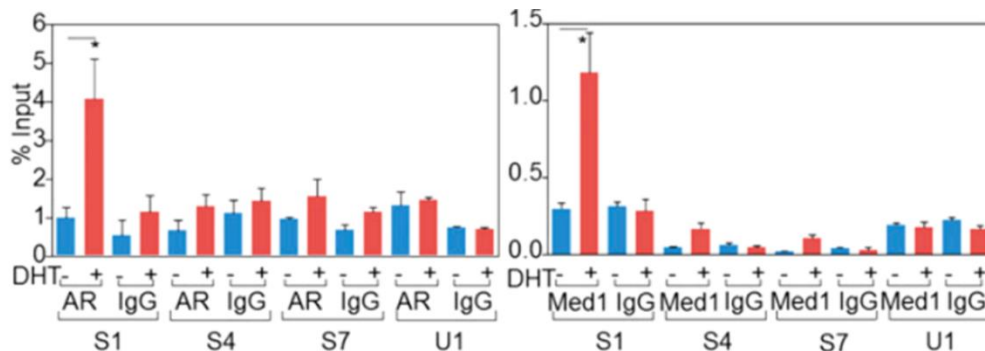


Figure 2.10: KLK3e fragments interact with AR and MED1.

RNA immunoprecipitation (RIP) of AR or MED1 was performed using LNCaP cells treated with vehicle negative control or 10nM DHT for 4 hours (Hsieh et al. 2014).

KLK3e lncRNA regulates target promoters together with AR and Med1

Silencing of KLK3e lncRNA, was shown to reduce AR and RNA Polymerase II (POLII) binding at the KLK2 promoter, similar to MED1 silencing (Hsieh et al. 2014), as shown in Figure 2.11 below.

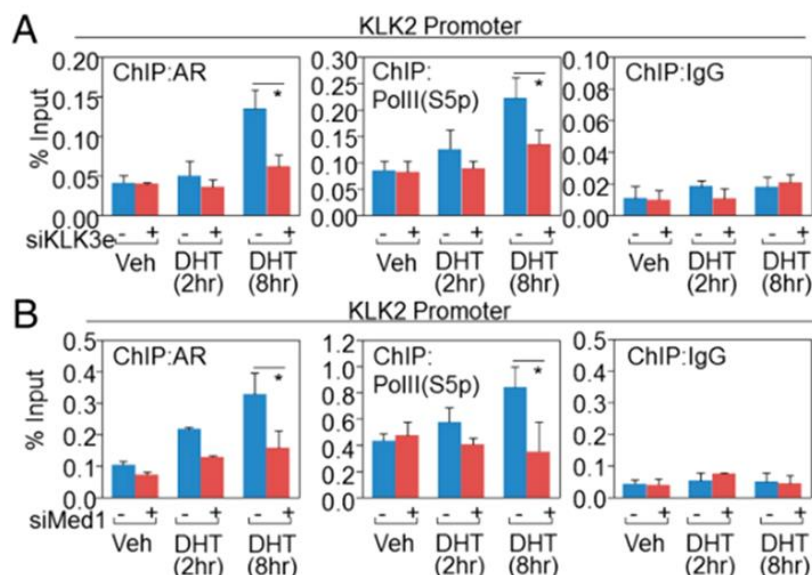


Figure 2.11: Silencing of KLK3e, followed by Chromatin immunoprecipitation and qPCR (ChIP-qPCR). AR and PolII were bound at the KLK2 promoter, and this binding was affected by KLK3e lncRNA (Hsieh et al. 2014).

KLK3e lncRNA chromosomal looping model

In the light of the findings above, the authors propose the model shown in Figure 2.12 below. KLK3e enhancer lncRNA is proposed to interact with AR, MED1 and POLII at the enhancer region of KLK3 and promote chromosomal looping with the promoters of KLK3 and KLK2 (Hsieh et al. 2014). The eRNA could also promote trans activation of distant genes, some of which regulated by AR have been identified, as shown in Figure 2.7B earlier.

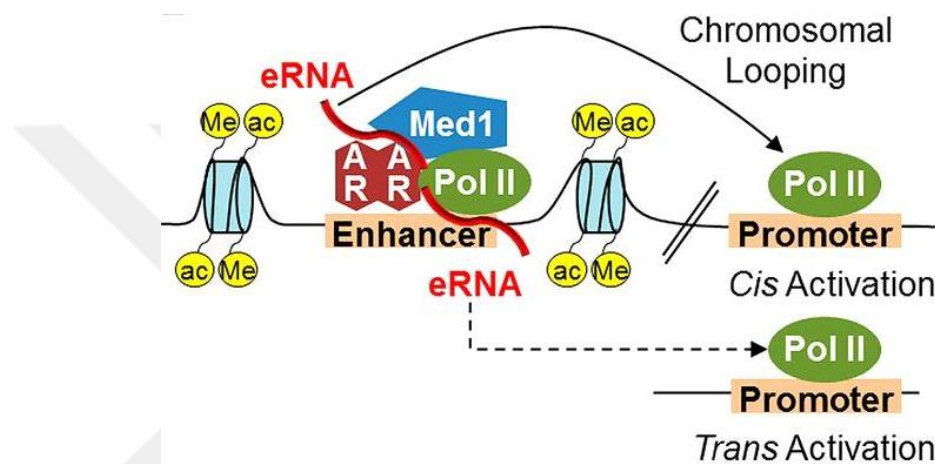


Figure 2.12: A proposed model for the functioning of the enhancer lncRNA KLK3e through interacting with AR, MED1 and POLII, and promoting chromosomal looping between enhancer and promoter regions (Hsieh et al. 2014).

However, they have not investigated what other proteins may be in this complex, and whether the KLK3e lncRNA may be forming alternative complexes with other proteins. It is also possible that the RNA is not directly interacting with AR and RNA PolII; there may be other proteins that mediate this complex.

Identification of proteins that directly interact with KLK3e could be achieved using RNA pulldown followed by mass spectrometry. In the project titled “Understanding the role of lncRNA in AR-mediated transcription,” we identified the protein interactors of this lncRNA to understand the mechanism by which it enhances the transcriptional activity of AR. Results are shown in Chapter 4.

2.3.4 Methods to Study Protein Interactions of lncRNAs

There are two main types of methods to study the protein interactions of lncRNAs. The first group targets the endogenous interactome of a lncRNA, however, may not be feasible for the capture of lncRNAs that are expressed in very little numbers, as most are. The preferred method for less abundant lncRNAs is to in-vitro transcribe it with a biotin tag, and perform RNA pulldown to bind interacting proteins, which can then be identified using mass spectrometry.

Endogenous RNA capture followed by mass spectrometry; RAP-MS and ChIRP-MS

These methods were both published in 2015 and developed using Xist. Chirp-MS (Chu et al. 2015) uses formaldehyde and RAP-MS (McHugh et al. 2015) uses UV treatment to crosslink the endogenous RNA with its interacting proteins in the cell. The RNA is then hybridized with tagged probes which can be captured using streptavidin coated beads. Bound proteins are then eluted and identified using mass spectrometry (Roth and Diederichs 2015). These steps are summarized in the depiction below in Figure 2.13.

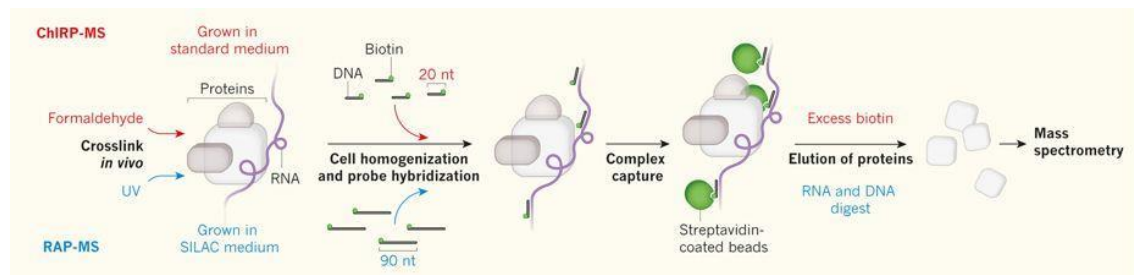


Figure 2.13: Capturing endogenous RNA–protein complexes using RAP-MS and ChIRP-MS. Taken from “Rap and chirp about X inactivation” (Roth and Diederichs 2015).

Pulldown with in-vitro transcribed RNA followed by mass spectrometry

A useful method in identifying protein interactors of lncRNAs that are less abundant in the cell is to in-vitro transcribe biotinylated RNA, pulldown the interacting proteins, and identify them using mass spectrometry. Several examples of applications of this method have been published, as shown in Figure 2.14 below. A detailed flowchart of this methodology that was applied is presented in the Methods section, in Figure 3.1.

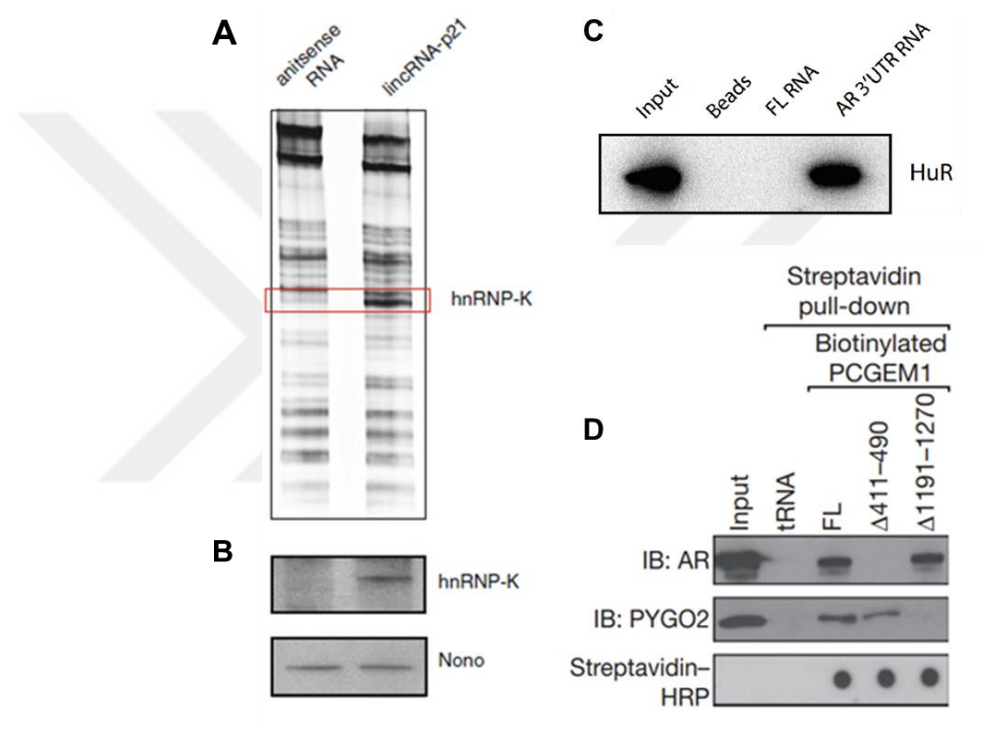


Figure 2.14: Published examples of lncRNA Pulldown experiments. A) Protein gel staining shows different bands between the negative control and the lncRNA, which can be identified using mass spectrometry. B) Western Blotting can be used to detect proteins known to be interacting with the lncRNA, in comparison to a negative control RNA. (Marín-Béjar ve Huarte, 2014) C) In a protocol published in 2016, the HuR protein was detected to be pulled down by the 3' UTR RNA of the AR (Bai vd., 2016). D) PCGEM1 (prostate-specific transcript) lncRNA can bind AR when in full length (FL), and this interaction is lost upon deletion of the 411-490 region (Yang vd., 2013).

2.4 Oligomerization Potentials of the Glucocorticoid Receptor

Traditionally, nuclear steroid receptors such as the androgen receptor (AR) and glucocorticoid receptor (GR) are often thought to function as transcription factors in the form of homodimers. However, steroid receptors have high sequence and structural similarities in their ligand binding and DNA binding domains. Therefore, it is not surprising that new evidence is rising to implicate cross-interactions between members of this protein family when they are expressed in the same tissue. For example, a recent publication in Nature has identified that interaction between the estrogen receptor (ESR) and progesterone receptor (PR) plays a role in breast cancer (Mohammed et al. 2015).

A depiction of the structural domains of nuclear steroid receptor family members can be found in Figure 2.15 below. The project titled “Oligomerization Potentials of the Glucocorticoid Receptor” focuses on the interaction with the glucocorticoid receptor (GR) with the mineralocorticoid receptor through their ligand and DNA binding domains through computational structural interaction prediction. The project also investigates tetramerization of glucocorticoid receptor, which was recently observed to occur upon DNA binding (Presman et al. 2016).

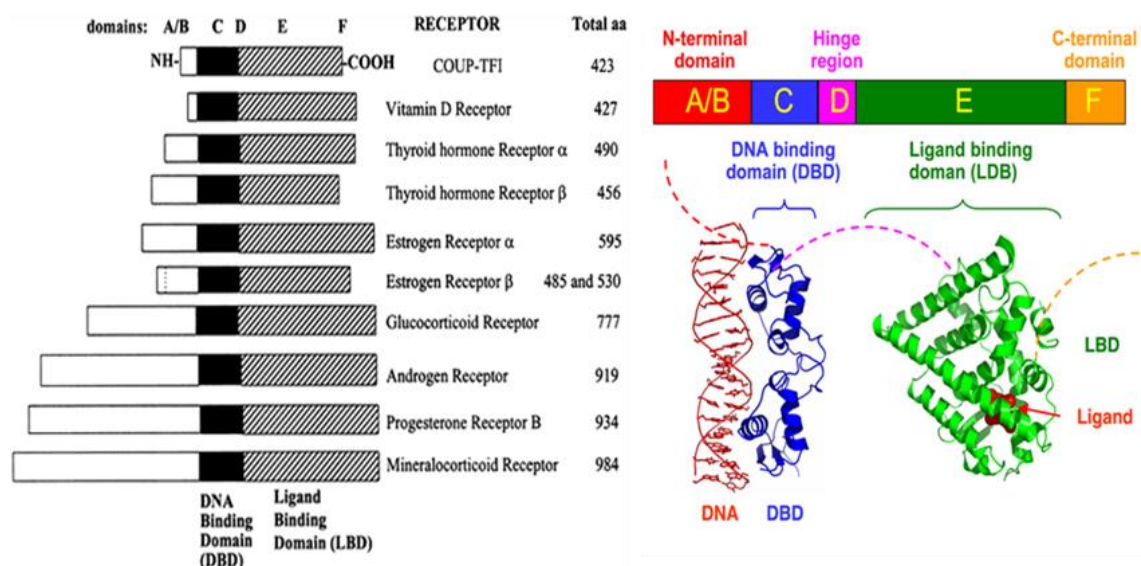


Figure 2.15: The superfamily of steroid nuclear receptors and the structural organization of their functional domains. Adapted from (Wahli and Martinez 2018) & Wikimedia Commons & PyMol

The glucocorticoid and mineralocorticoid receptors are two closely related members of the steroid nuclear receptor family that are able to bind the same ligands, such as the endogenous corticosterone, or the synthetic dexamethasone (Gomez-Sanchez and Gomez-Sanchez 2014).

The Glucocorticoid Receptor (GR)

The glucocorticoid receptor (GR) is present in almost every cell type of vertebrates. Cortisol and corticosterone are the two endogenous glucocorticoid hormones, secreted by adrenal glands above the kidneys. Activation of GR by these ligands is associated with wakefulness, stress, and anti-inflammatory properties. In the human body, these hormones are released in basal levels under the control of the daily circadian rhythm; cortisol levels are higher in the morning, and are low at night. Failure of reduction of cortisol by night time due to factors such as stress can interfere with sleep. (Chan and Debono 2010; Kadmiel and Cidlowski 2013)

Some key functions of glucocorticoid binding to GR are; activation of glucose synthesis in the liver, upregulation of anti-inflammatory proteins, and suppression of inflammatory proteins. Because of the anti-inflammatory roles of GR, synthetic corticoids such as Prednisone are used very commonly in the clinic to treat acute immune responses or auto-immune diseases. (Kadmiel and Cidlowski 2013)

The Mineralocorticoid Receptor (MR)

The major ligand of the mineralocorticoid receptor is aldosterone, but the mineralocorticoid receptor is able to bind cortisol and corticosterone with the same binding affinity, and can also bind progesterone (Gomez-Sanchez and Gomez-Sanchez 2014). The mineralocorticoid receptor is mainly expressed in the kidneys and distal colon, and its activity regulates salt and water homeostasis of the body, mainly through acting on epithelial sodium channels (Fuller et al. 2000).

2.4.1 Interactions of the Glucocorticoid and Mineralocorticoid Receptors

As expected from their ability to bind the same hormone ligands as described above, the ligand binding domains (LBD) of the glucocorticoid receptor (GR) and mineralocorticoid receptor (MR) are structurally highly similar. Their DNA binding domains (DBD) are also highly homologous and their DNA binding site recognition is identical. (Reul and De Kloet 1985; Arriza et al. 1987; Pearce and Yamamoto 1993)

For both GR and MR, dimerization through the DBD was observed to be mediated by the D-loop in the second zinc finger motif. (Dahlman-Wright et al. 1991; Hudson et al. 2014) However, observations with other nuclear steroid receptors show cross interactions with the possibility of alternative interfaces. For example the estrogen receptor and Thyroid/Retinoid-X Receptor (TR/RXR) can interact through different interfaces. (Fawell et al. 1990; Hirst et al. 2006)

Their similar dimerization modes and high homology lead to the investigation of heterodimerization between GR and MR, which has since been demonstrated, expanding the regulatory roles of glucocorticoid hormones in cells that express both receptors (Trapp et al. 1994). This interaction has been studied mainly in vitro (Trapp et al. 1994; Liu et al. 1995; Derfoul et al. 1998; Ou et al. 2001; Planey et al. 2002), with some observations in living cells or tissue (Nishi 2004; Mifsud and Reul 2016).

Alternative dimerization interfaces between GR and MR have been suggested (Liu et al. 1996; Bledsoe et al. 2002). Another publication has demonstrated in live cells that the GR LBD mediates oligomerization with MR in the cytoplasm that is involved in their cotransport into the nucleus (Savory et al. 2001). Considering that allosteric effects are possible in many proteins, allowing communication between domains, it is possible that these proteins have alternative conformations and alternative modes of interaction depending on the cellular environment. Proteins may also interact through more than one domain when they come in close vicinity.

Understanding how GR and MR interact will shed light on how they regulate downstream responses in tissues where they are expressed together, such as the kidney,

immune cell types, hippocampus, adipocytes and osteoblasts (Armanini et al. 1988; Beavan et al. 2001; Han et al. 2005; Marzolla et al. 2012; Farman et al. 2017).

2.4.2 Glucocorticoid Receptor Tetramerization

In a PNAS publication in 2016, it was observed that GR is predominantly a tetramer when bound to its target DNA (Presman et al. 2016), challenging the traditional thinking that these receptors act only in dimers. They have demonstrated that binding the DNA causes allosteric changes in the protein, which results in tetramerization. The main observations in this paper will be summarized, and the results section will show computational modeling analyses to help explain how this tetramerization might occur at the structural level.

The N&B assay and MMTV array

This research utilized mouse 3617 cells which contains multiple copies of a GR responsive promoter structure; this system is called the mouse mammary tumor virus (MMTV) array. When GR was transiently expressed in these cells with a green fluorescent protein (GFP) tag, the GR responsive elements that were bound with GR would give an enriched GFP signal. They then applied number and brightness (N&B) analysis in the nucleus or at the MMTV array. With this method, the molecular brightness (ϵ) of molecules within a pixel can be measured and used to estimate relative oligomerization states of a protein, based on fluctuations caused by their movement in and out of a frame. If there are oligomers of different sizes, the final brightness reading ϵ will be an average combination of the population. (Presman et al. 2016)

GR forms tetramers when bound to DNA

As shown in Figure 2.13, GR is dimeric at the nucleoplasm as expected, but mainly tetrameric when bound to DNA at the MMTV array. Also interestingly, inhibitor (RU486) bound GR has less ability to tetramerize compared to GR bound by its agonists, corticosterone and dexamethasone. (Presman et al. 2016)

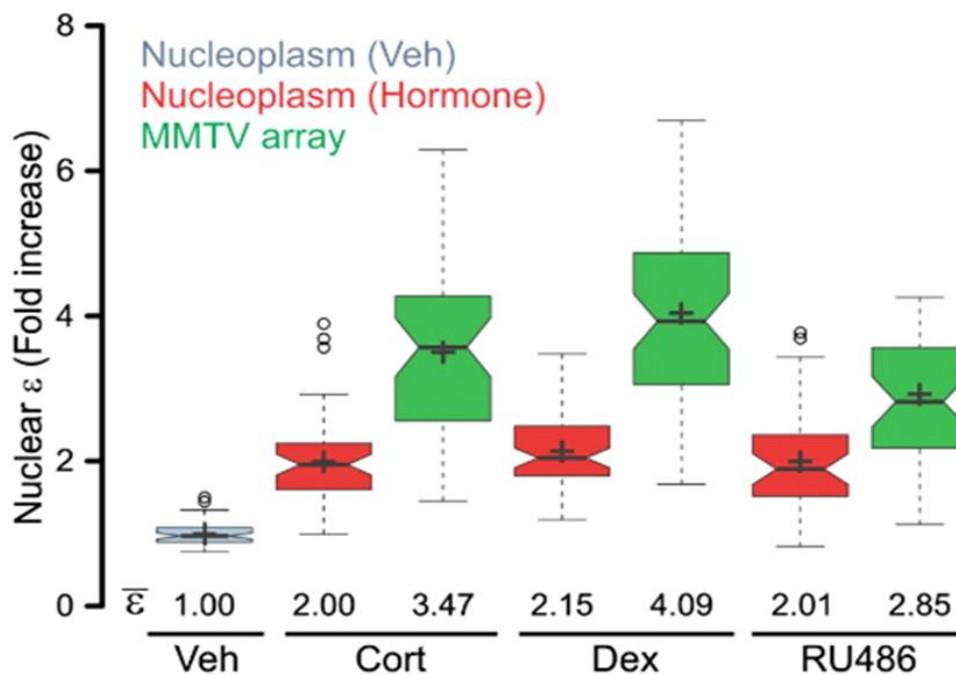


Figure 2.16: GR is dimeric at the nucleoplasm but mainly tetrameric when bound to DNA at the MMTV array. Corticosterone and dexamethasone bound GR is able to tetramerize more, compared to inhibitor (RU486) bound GR. Figure from (Presman et al. 2016)

GR tetramerization is independent of dimerization

Even when dimerization interface amino acids were mutated, and GR was monomeric in the nucleus, it was still tetrameric at the MMTV array, when it was bound to chromatin. This demonstrated that tetramerization was independent of dimerization, therefore GR was forming a ‘dimer of dimers’ (Presman et al. 2016).

The ligand binding domain is necessary for tetramerization

They also investigated which functional domains were necessary for dimerization. When the LBD was deleted, GR was unable to form tetramers, even with a DNA-bound mimic mutation. The N terminal domain was shown to contribute to dimerization, but was not necessary for tetramerization (Presman et al. 2016).

The authors then propose that the ‘dimer of dimers’ tetramerization of GR could occur through two alternatives; the two dimers may bind the same DNA strand, or there may be chromosomal looping involved where each dimer binds a different response element on the DNA that come close due to chromosomal looping that occurs during transcription (Presman et al. 2016), as depicted below in Figure 2.17.

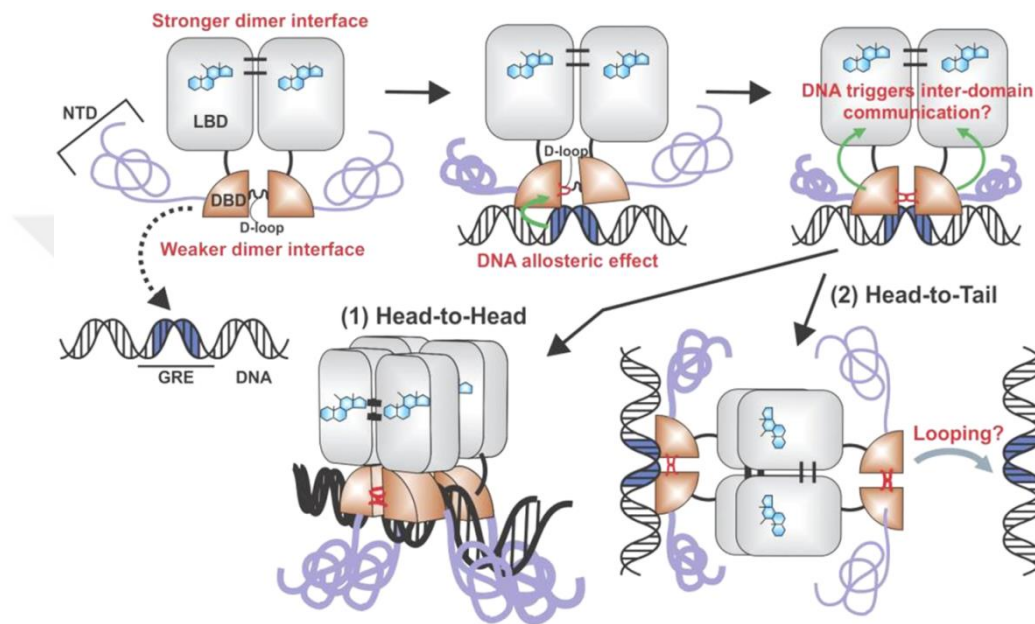


Figure 2.17: Model proposed for formation of GR tetramers. (Presman et al. 2016)

GR crystal structure tetramer

It is important to note that one crystal structure deposited in the protein data bank (PDB) (Berman et al. 2000), of the ligand binding domain of the glucocorticoid receptor, is in a tetramer form; 1p93 (Kauppi et al. 2003). Moreover, this tetramer consists of two dimer interfaces, both of which are paralleled in other GR LBD crystal structures. The 1p93AC interface can also be seen in the 1m2zAD dimer structure. The 1P93AB/CD interface can also be seen in 3cldAB, 3e7cAB, and 4p6xCG dimers of GR crystal structures. Of note, the hot spot amino acids of these interfaces have been confirmed to be the same using HotRegion (Cukuroglu et al., 2012). The crystal structure and the two dimer interfaces within are shown in Figure 2.15.

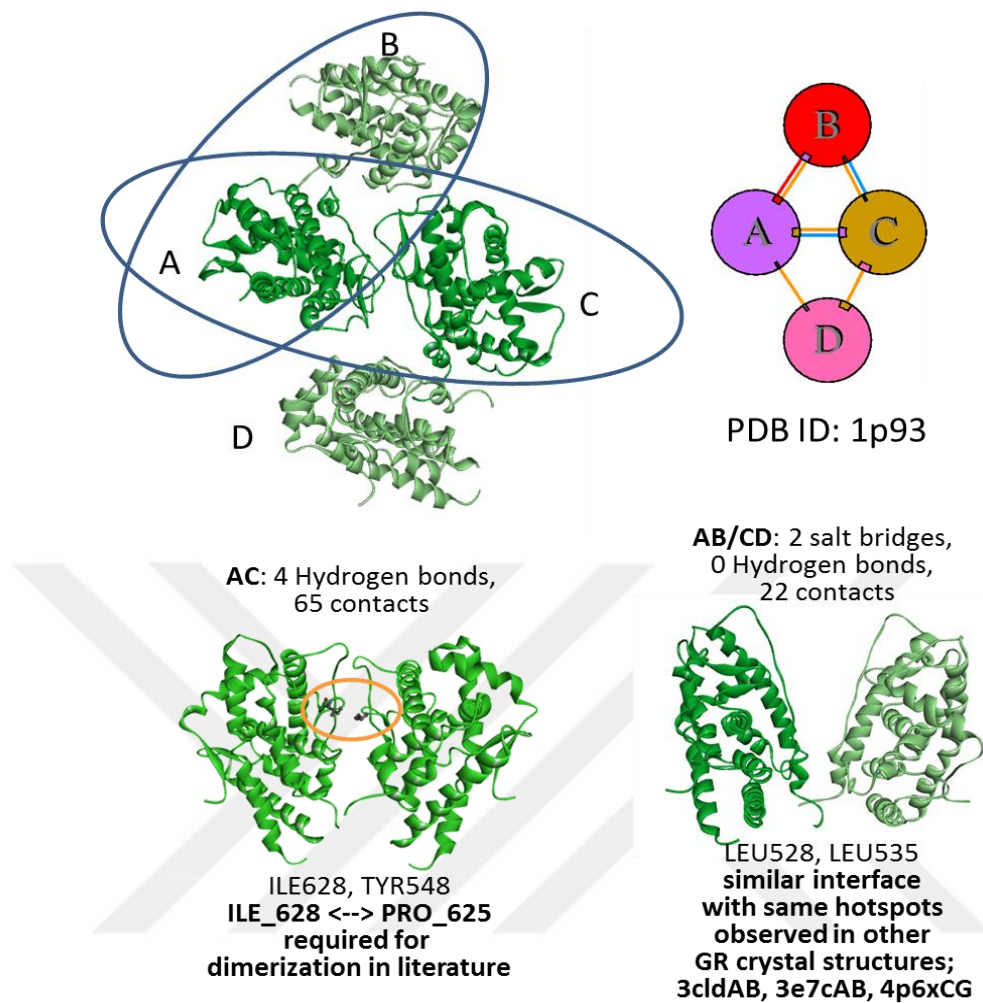


Figure 2.18: Tetrameric human GR LBD crystal structure with PDB ID 1p93 (Kauppi et al. 2003), **and the two dimer interfaces within.** Bond analyses from PDBsum (Laskowski et al. 1997).

Consistently, experimental data also suggests that GR has at least two alternative dimerization interfaces; one dependent on ILE628 like the 1p93AC or the 1m2zAD interface and the other not. Mutation of I628A on human GR was shown to decrease dimerization of LBD-LBD fragments in vitro by 10 fold, but did not completely disrupt all dimers (Bledsoe et al. 2002).

Mutation of the homologous amino acid in mouse as I634A causes loss of dimers in the nucleus, but not at the MMTV array on DNA, implying that the dimers on DNA may be formed through an alternative interface (Presman et al. 2016).

2.4.3 Oligomerization potentials of other steroid receptors

Interestingly, in the paper regarding tetramerization of GR upon DNA binding (Presman et al. 2016), they have also looked at the oligomerization status of other steroid receptors, and seen that the progesterone receptor (PR) and the androgen receptor (AR) are also found in higher oligomeric states, as shown in Figure 2.19 below. In fact, their oligomerization does not seem dependent on DNA binding.

Higher order oligomerization of AR is particularly interesting because hormone dependent aggregation of AR with a polyglutamine stretch amplification is implicated in spinal and bulbar muscular atrophy (SBMA) (Jochum et al. 2012).

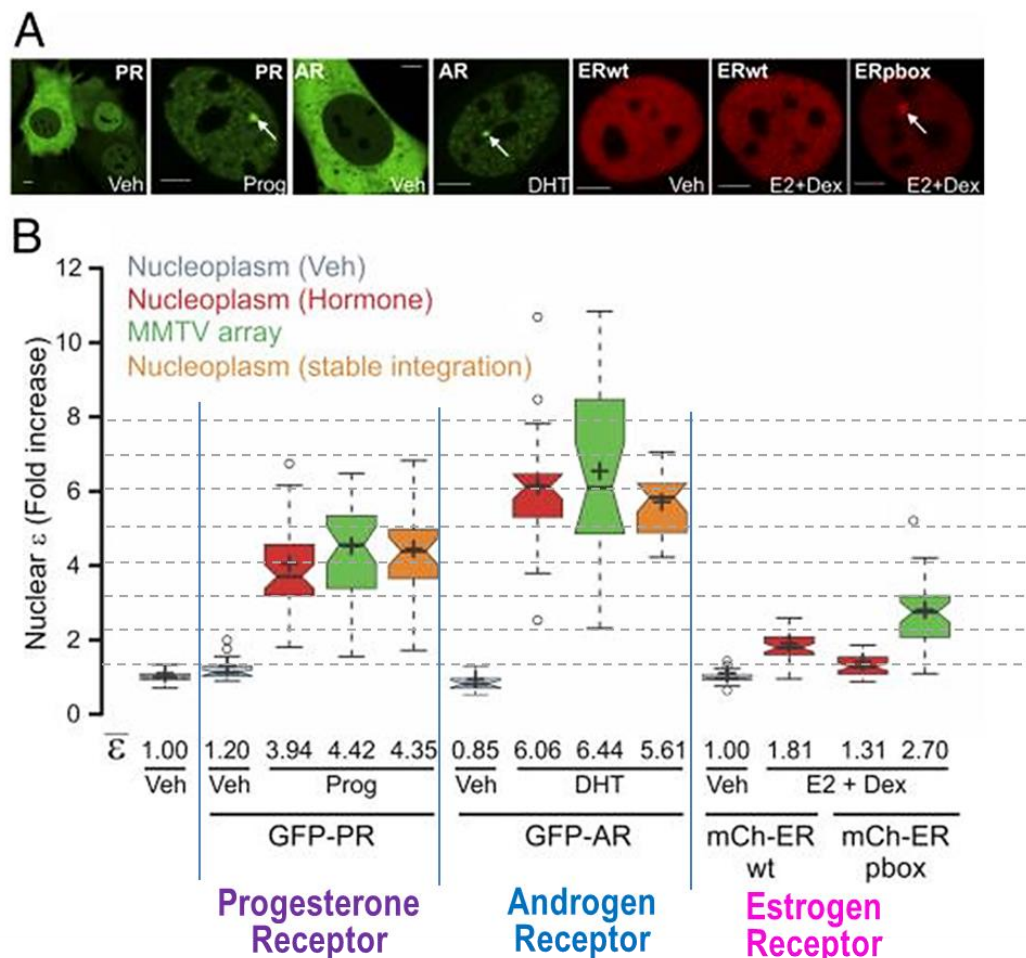


Figure 2.19: Oligomerization potentials of other steroid receptors (Presman et al. 2016).

2.4.4 *Glucocorticoid Receptor and Prostate Cancer*

Another interesting observation regarding the glucocorticoid receptor is that glucocorticoids may be associated with poor prognosis in metastatic prostate cancer patients during clinical administration of abiraterone and enzalutamide (Narayanan et al. 2016).

It was initially detected that the glucocorticoid receptor (GR), progesterone receptor (PR), and mineralocorticoid receptor (MR) can recognize the core sequence of the androgen response element (ARE) (Kim et al. 2006). It was then observed that the AGAACA consensus sequence organized as an inverted repeat with a 3n spacer could be bound both by AR and GR (Claessens et al. 2017).

But most interestingly, glucocorticoid receptor was observed to cause resistance to antiandrogen drugs by bypassing the blocking of the androgen receptor (Arora et al. 2013). Drug resistant preclinical and patient tumors showed elevated GR expression. GR was shown to replace AR to cause a resistant phenotype through the upregulation of similar target genes. A GR inhibitor was able to reconstitute sensitivity in enzalutamide resistant samples, while a synthetic GR agonist was able to confer resistance by itself. Upon inhibition of AR, GR expression was observed to be elevated in a subgroup of prostate cancer cells, which was associated with the loss of feedback repression of GR by AR (Arora et al. 2013).

2.5 *Circadian Clock and Cancer*

2.5.1 *Circadian Clock and the Transcription Factors Involved*

The circadian clock regulates the physiology of organisms with a ~24h oscillatory period. It coordinates many physiological variables including oscillations in hormone secretion, blood pressure, and cognitive function (Takahashi et al. 2008). At the molecular level, the cellular clockwork involves both positive and negative transcriptional feedback loops. The BMAL1/CLOCK heterodimer is central to circadian biology, binding E-box elements (CACGTG) in the *Period (Per)* and *Cryptochrome (Cry)* genes and exerting a positive effect on their transcription in mammals (Kume et al. 1999; Panda et al. 2002). Then, the transcriptional repressors PER and CRY repress BMAL1-CLOCK activity and their own transcription (Gekakis et al. 1998; Vitaterna et al. 1999; Oishi et al. 2000; Shearman 2000; Kavakli et al. 2017).

2.5.2 *Circadian Clock and Cancer*

Due to its widespread regulatory effects, it is expected that a relationship exists between the circadian clock and carcinogenesis, but the interplay is quite complex. Early research showed that arrhythmic mice with SCN-lesions and *Per2* mutant mice had faster tumor progression and higher rates of cancer incidence than control animals, respectively, which led to the generalization that disruption of the circadian clock would lead to increased propensity to cancer (Filipski et al. 2002; Fu et al. 2002). However, mice with deletions of *Clock* or *Cry* genes had similar rates of cancer incidence as wild type mice (Gauger and Sancar 2005; Antoch et al. 2008).

The subsequent case to examine was the effects of circadian alterations in a background of elevated propensity to cancer due to a mutation in a tumor suppressor or oncogene. The *p53*^{-/-} mice were used as model organism to investigate such relationship between them since p53 is a key tumor suppressor protein that is known to be mutated in nearly half of all human cancers types (Attardi and Jacks 1999). The *p53*^{-/-} mice were established to develop cancers such as lymphomas and lymphosarcomas, and therefore

have a median lifespan of about 20 weeks, (Donehower et al. 1992; Attardi and Donehower 2005) which is quite dramatic since wild-type mice of the same strain have a mean life-span of about 2 years (Rowlatt et al. 1976).

Intriguingly, the absence of the *Cry* genes in $p53^{-/-}$ background mice delayed the onset of cancer; $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ mice had about a 1.5 fold longer median lifespan compared to $p53^{-/-}$ mice (Ozturk et al. 2009). Several subsequent studies suggested that loss of *Cry* genes in $p53$ mutants amplifies pathways of apoptosis in the absence of a functional p53. The initial important observation was that $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ mice skin fibroblasts (MSF) did not have reduced oncogenic transformation compared to $p53^{-/-}$ mice, but were significantly more sensitive to UV-induced apoptosis (Ozturk et al. 2009). Detailed analysis of the augmented apoptotic behavior revealed that the levels of the p73 protein are elevated in $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ cells in response to DNA damage. In response to oxaliplatin treatment, *Cry* mutants were observed to have elevated levels of EGR1, which was transactivating p73 (Lee and Sancar 2011a) and result in DNA-damage induced intrinsic apoptosis mechanisms. Notably, p73 is a member of the p53 protein family which has tumor suppressive functions that may become more apparent in a $p53^{-/-}$ knockout setting, and is known to regulate some of the genes in the p53 pathway (Fontemaggi et al. 2002; Harms et al. 2004; Lin et al. 2009).

Additionally, *Cry* mutation was also shown to promote extrinsic apoptosis via two mechanisms. Firstly, levels of pro-apoptotic cytokines such as TNF α were shown to be elevated in *Cry* mutant mice, causing an elevated inflammatory state (Hashiramoto et al. 2010). TNF α is a death ligand released by immune cells that induces apoptosis in the target cells where it binds its receptor, making it a key player in the antitumor response of the immune system (Wajant et al. 2003). Therefore, it is quite interesting that the levels of this cytokine are elevated upon loss of CRY protein, since this may be playing an important role in the extended lifespans of $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ mice compared to $p53^{-/-}$ mice. A second and complementary extrinsic apoptosis mechanism observed after the deletion of *Cry* in $p53^{-/-}$ mutant mice involved the inactivation of GSK3 β , which fails to activate the anti-apoptotic activity of NF- κ B, which causes cells to be more sensitive to the induction of apoptosis by TNF α (Lee and Sancar 2011b).

2.5.3 Protein-Protein Interaction Network Analysis

Protein-protein interaction (PPI) networks can be constructed by extracting and visualizing protein-protein interaction data from databases, which can help determine key proteins within known networks. The degree of a node is defined as the number of edges linked to that node, therefore, in protein signaling networks, proteins with a very high degree number are interacting with many other proteins, which suggests a central regulatory role. Such proteins with high degree values are called “hub” proteins. (Tuncbag et al. 2009; Engin et al. 2012; Scardoni and Laudanna 2012; Cavga et al. 2015).

The betweenness centrality of a node is the number of times a node acts as a bridge along the shortest path between two other nodes. This reflects the amount of control that this node exerts over the interactions of other nodes in the network. In contrast to the measure of degree, the betweenness value favors nodes that join communities, rather than nodes that lie inside a community. Such proteins are called “bottleneck” proteins. Bottleneck proteins are thought to be biologically very significant, because even though they may not be making a lot of interactions, they make key interactions that play key roles in relaying a signal from one sub-network to another (Freeman 1977; Yu et al. 2007).

Chapter 3:

METHODS

In this work, RNA and protein effectors of transcription factor activity and gene expression have been studied combining experimental and computational methods.

Section 3.1 describes the RNA pulldown, mass spectrometry, and computational structural modeling methodologies used in the identification and analysis of protein interactors of the enhancer lncRNA KLK3e that influences androgen receptor mediated transcription.

Section 3.2 articulates the methods used in the computational analysis of data obtained from Rapid Immunoprecipitation Mass spectrometry of Endogenous proteins (RIME) and RNA sequencing assays performed in order to elucidate the effects of four novel androgen receptor inhibitors that disrupt the interactions, dimerization, or DNA binding of this transcription factor.

Section 3.3 describes the computational methods used in the structural modelling of the interactions and oligomerization of the glucocorticoid receptor.

Section 3.4 explains the methodologies followed during the analysis of RNA sequencing data of Cryptochrome deletion in a p53 background.

3.1 Identifying protein interactors of the lncRNA KLK3e

The methodology followed in order to identify the protein interactors of lncRNA KLK3e is described in the flowchart in Figure 3.1. Firstly, the lncRNA template DNA is amplified by polymerase chain reaction (PCR), then RNA is in vitro transcribed with biotin tags, attached onto streptavidin magnetic beads, and used to pulldown interacting proteins from a cell lysate. The proteins that interact specifically with KLK3e RNA compared to controls are then identified using mass spectrometry. Finally, computational modeling of the RNA-protein complex is formed to infer mechanism of action on how this enhancer RNA influences AR transcriptional activity.

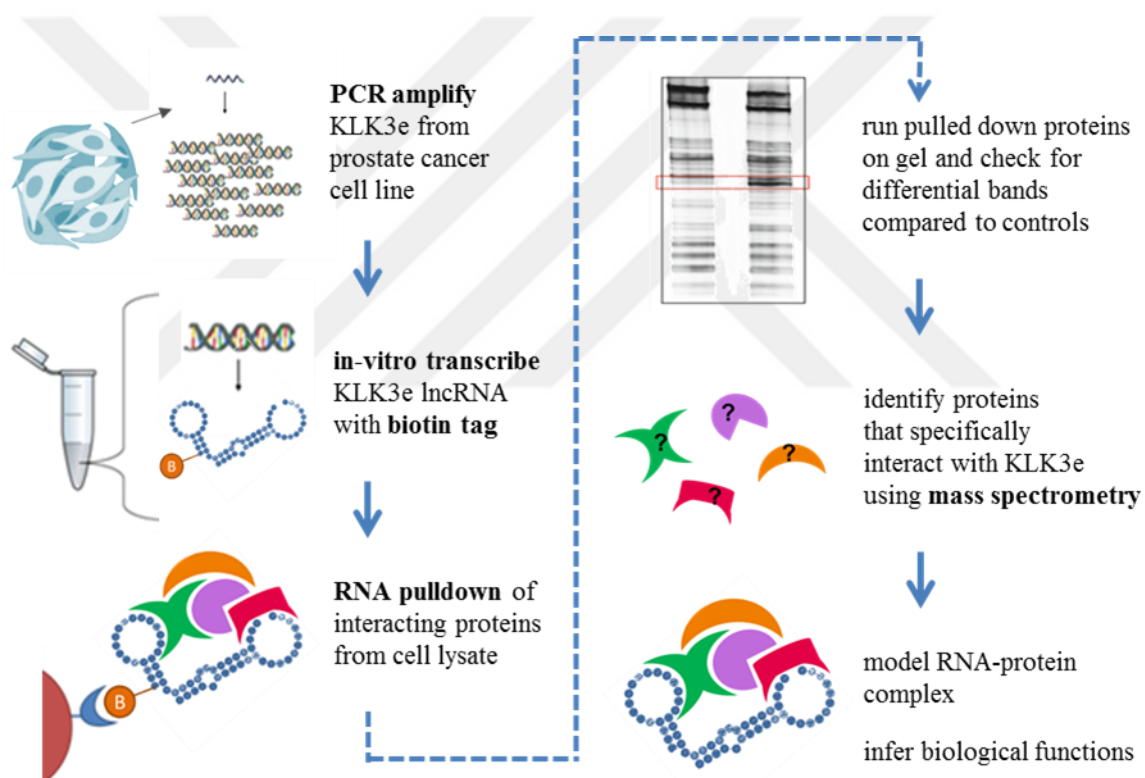


Figure 3.1: Project flowchart to identify protein interactors of lncRNA KLK3e.

Xist was used as a positive control as a well-established lncRNA with known protein interactors. mCherryGFP is used as a negative control to account for general affinity of RNA to proteins, as it is merely a transcript of the same size that should not have a lncRNA function.

The protocol followed was mainly as described in a protocol paper that describes detection of RNA-binding proteins by *in vitro* RNA pull-down in adipocyte cultures, with some adjustments, and will be described in detail below. (Bai et al. 2016)

3.1.1 Reverse-Transcription PCR amplification of *KLK3e lncRNA* and controls

Total RNA isolation from LNCaP and T293 cells were performed as described in the protocols provided in the Macherey-Nagel NucleoSpin RNA kits (Macherey-Nagel, Germany), including the DNase digestion step to eliminate genomic DNA. Total RNA were then reverse transcribed to make cDNA as described below.

The mixture below was prepared in a small PCR tube;

Total RNA	1 µg
Random Hexamer Mix	1 µl
2 mM dNTP Mix	2.5 µl
Nuclease-free H ₂ O	to a total volume of 13 µl

The mixture was then incubated at 65°C for 5 minutes for denaturation of RNA and primer, followed immediately by 4°C or ice. The following components were prepared as a master mix and 7.5 ul was added in each sample tube.

5X First-Strand Buffer	5 µL
0.1 M DTT	2 µL
Ribonuclease Inhibitor (40U/µL)	0.5 µL

The mixtures were then incubated at room temperature for 10 minutes.

Finally, 1µl of M-MLV Reverse Transcriptase (RT) was added in each tube except for RT- controls, in which same amount of nuclease-free water was added. The tubes were then incubated in 37°C for 1 hour, followed by 15 minutes of inactivation at 75°C, followed by 4°C or ice. The cDNA were then diluted with nuclease-free water, and used in 1:4 dilution for the Polymerase Chain Reactions (PCR) reactions below. PCRs were performed using the primers provided below. The T7 promoter extensions added to the Forward primers are marked in green, and optimized melting temperatures (T_m) are provided.

Table 3-1: PCR primers with T7 promoter extensions. These primers were used to amplify KLK3e lncRNA and controls to provide templates for in-vitro transcription of RNA with a T7 RNA Polymerase that utilizes the T7 promoter. KLK3e was amplified from total RNA reverse transcribed into cDNA from LNCaP cells, Xist from 293T cells, and mCherryGFP from pBABE-puro mCherry-EGFP-LC3B plasmid.

Primer Name	Primer Sequence	Tm
KLK3e_2k_T7_F	TAATACGACTCACTATAGGACCTGCTCAG CCTTTGTCTC	65 °C
KLK3e_2k_R_1984	CCGGTGTCCCCAAGCATATC	
Xist_T7_F	TAATACGACTCACTATAGGTCCCCCAACA CCCTTTATGG	68 °C
Xist_R_1820	ATGGAGGGAGGTTTCAGACCA	
CherGFP_T7_F	TAATACGACTCACTATAGGATCCAGCCCTC ACTCCTTCT	70 °C
CherGFP_R_1573	CCGACGGCATGAATTCGGTA	

The PCR reactions were prepared with the components listed below. After optimization, 8-10% DMSO addition was determined to provide the best results, due to the high GC content of the transcripts as well as the T7 promoter extensions.

Importantly, in the case of KLK3e, the final product concentration was too low. Therefore, an initial PCR was performed, and the cleanup was diluted 1:75 and used in a second PCR with the same conditions to further amplify the transcript.

Component	50 µl Reaction	Final Concentration
Nuclease-free water	to 50 µl (27µl)	
5X GC Buffer	10 µl	1X
DMSO	4 µl	8%
10 mM dNTPs	1 µl	200 µM
10 µM Forward Primer	2.5 µl	0.5 µM
10 µM Reverse Primer	2.5 µl	0.5 µM
Template DNA	2.5 µl	
Phusion DNA Polymerase	0.5 µl	1.0 units/50 µl

Thermocycling conditions:

Initial Denaturation	98°C	30 seconds	
Denaturation	98°C	10 seconds	} x 42 cycles
Annealing	65/68/70 °C	30 seconds	
Extension	72°C	50 seconds	
Final Extension	72°C	10 minutes	
Hold	4-10°C		

The PCRs were performed in 2x8 tubes for Xist and mCherryGFP, and 6x8 tubes for KLK3e using 1:75 PCR cleanup of the first PCR as a template. PCR reactions were then pooled 8X for PCR cleanup as described in the Macherey-Nagel Gel and PCR cleanup kits (Macherey-Nagel, Germany).

Importantly, since the KLK3e PCR products had some primer dimers, the NT1 buffer of the kit was diluted with water in a ratio of 2:1 water to NT1, to get rid of lower size bands, as described in the kit protocol. (Macherey-Nagel, Germany). This produced one clean band in the gel, which could be used as a template for the in-vitro transcription by T7 RNA Polymerase.

The PCR clean-up products had to be at least 160ng/ul in concentration to be used as template DNA in the in-vitro transcription reaction described below. Therefore, the elutions were made in low volumes with a maximum of 25ul NE Elution Buffer.

The PCR products that were cleaned-up were checked in a 1% Agarose gel to produce one specific band, as shown in the results section. Samples from the clean-ups were sent to sequencing to the Macrogen company (Macrogen Europe, Netherlands) via Sanger sequencing using the T7 promoter. Sequence confirmations were performed using the BLAST tool of the NCBI. The sequence comparisons are provided in the Results and Appendix sections.

3.1.2 *In vitro* Transcription of *KLK3e* lncRNA and controls, with Biotin tags

In vitro transcription was performed according to the protocol described below. Importantly, the Bio-UTP to UTP ratio was fixed at 1:10, which was determined to be the best ratio for these long transcripts, as more biotin could interfere with pulldown of interacting proteins.

The components were added in the specific order written, into a 1.5 ml sterile eppendorf. All components except for the last two enzymes were first warmed to room temperature, to prevent precipitation of spermidine in the transcription buffer.

In Vitro Transcription Reaction Components:

1)	DEPC- H ₂ O	18.5-x µl	
2)	Bio-16-UTP 2.5 µM	1.5 µl	(Jena Bioscience, Germany)
3)	10× NTP Mix*	4 µl	(Jena Bioscience, Germany)
4)	Template DNA	x µl (3 µg)	
5)	5X Transcription Buffer	8 µl	
6)	DTT 100mM	2 µl	
7)	RNAse inhibitor (Promega)	4 µl	
8)	T7 RNA Polymerase	2 µl (400 U)	(Jena Bioscience, Germany)
Total:		40 µl	

*NTP Mix: 10mM ATP, CTP, GTP and 9mM UTP

This mix was incubated for 4 hours at 37 °C on a heat block, with occasional mixing.

Biotinylated, in-vitro transcribed RNA products were then purified using a pre-chilled Macherey-Nagel NucleoSpin RNA kit (Macherey-Nagel, Germany), including the DNase digestion step to eliminate the template DNA. 2 in-vitro reaction tubes of 40ul were combined into one and completed to 100ul and bound on a single RNA spin column. Elution was made 2 times with 60ul RNAse free H₂O, which often produced RNA concentrations around 200-300ng/ul.

The RNA was then freshly used for subsequent pulldown reaction described below, to avoid RNA degradation during freezing and thawing. Alternatively, RNA could be stored at -80°C until pulldown.

For the visualization of RNA, Tris/Borate/EDTA (TBE) buffer and gel was used in 1X (from 10X stock, Thermo Scientific, US). The RNA samples supplemented with 2X RNA Dye and Low Range ssRNA ladder (NEB, US) were denatured by melting at 90°C for 10 minutes and were run on a 1% TBE gel at 50-60V. The gel was post-stained in EtBr for visualization to check for a single specific RNA band of expected size. Images are provided in the results section.

3.1.3 Testing biotinylated RNA binding by Streptavidin Magnetic Beads

In order to confirm that the RNA products from the IVT are biotinylated and that the Streptavidin magnetic beads can bind them efficiently and specifically, an RNA pulldown was tested without proteins, and by comparing the binding of biotinylated vs. unbiotylated RNA products from the in vitro transcription reaction. The materials and the protocol were as below.

In vitro transcription was performed as described, but by also preparing unbiotinylated controls using unbiotinylated UTP.

<u>Wash Buffer (1X)</u>	<u>Salt Buffer(1X)</u>	<u>Biotin Elution Buffer(1X)</u>
20mM Tris (pH 7.5)	20mM Tris (pH 7.5)	95% formamide
10mM NaCl	1M NaCl	10 mM EDTA pH 8.2
0.1% Tween-20	1mM EDTA	65°C for 2 minutes

Before use, the buffers were freshly supplemented with 0.5 mM DTT and 100U/ml RNasin, to help maintain RNA. A heat block was set to 95°C, Elution Buffer was prewarmed in 65°C, and Salt Buffer was placed on ice.

50 µl (500 µg) of Streptavidin Magnetic Beads were pipetted into a clean RNase-free tube. 50 µl of Wash Buffer was and beads were resuspended. Magnet was applied for

~30 seconds. Supernatant was removed and the wash was repeated. In a separate tube, 20 μ l of Wash Buffer was added, followed by 30 μ l (~30ug, 30pmol) of IVT RNA supplied with 0.1 μ g/ μ l yeast tRNA. This RNA mix was boiled at 95°C for 5 minutes for denaturation, then quickly put on ice for 3 minutes. RNA samples were pipetted onto previously prepared dry magnetic beads and resuspended. This mixture of beads and RNA was incubated at room temperature for 45min with occasional agitation by hand.

After the 1 hour incubation, magnet was applied and the supernatant was taken to a new tube. This was kept on ice for later comparison with the elute during Nanodrop measurement and RNA gel.

50 μ l of Wash Buffer was added onto the beads. Beads were resuspended and incubated in wash buffer at room temperature for 5 minutes. Magnet was applied and supernatant was carefully and completely removed. This wash was repeated one more time with Wash Buffer, and another time with cold Salt Buffer.

Then, 20 μ l of prewarmed Elution Buffer was added by resuspension, and beads were incubated 2 minutes at 65C to elute the RNA. Magnet was applied and the elute(supernatant) was transferred to an RNase-free microcentrifuge tube. This elution was repeated with 20 μ l of fresh prewarmed Elution Buffer, resuspended, then incubated 2 minutes at 65C. Magnet was applied and second elute was added onto the first. 10ul of elutes and supernatants were run on TBE RNA gel to compare the binding of biotinylated versus unbiotinylated RNA. Gel images are provided in the results section.

3.1.4 DHT treatment of cells

LNCaP cells were grown in standard RPMI 1640 medium (2.05 mM L-glutamine, with 25mM HEPES, Gibco, UK) supplemented with 10% Fetal Bovine Serum and Penicillin-Streptomycin. Cells were checked weekly for mycoplasma contamination using 2-3 days old media, and only mycoplasma-negative confirmed cells were used. After a few passages, upon reaching ~90% confluency, cells were treated with 10nM dihydrotestosterone (DHT) dissolved in 70% EtOH, in regular media for 4 hours. Cells were then collected either via Trypsin-EDTA (0.25%, Gibco, UK) from T75 flasks

(Sarstedt, Germany) or using cold PBS from 15cm plates, flash-frozen in liquid nitrogen, and stored in -80°C until lysis and protein extraction.

3.1.5 Nuclear protein extraction from LNCaP cells

During protein isolation, the cytoplasmic protein fraction was removed using the NE-PER Nuclear and Cytoplasmic Extraction Kit (Thermo Scientific, MA, USA), after which the nuclear protein fraction was isolated as described in a protocol paper regarding the detection of RNA-binding proteins by *in vitro* RNA pull-down in adipocyte cultures (Bai et al. 2016) This combination of methods was determined to produce the cleanest pulldown results after several optimization steps trying alternative buffers and methods such as sonication.

Eppendorf tubes and centrifuge were pre-chilled to prevent protein degradation. LNCaP cells collected from 1 15cm plate or 2 T75 flasks were dissolved in 1 ml ice-cold CER I cytoplasmic lysis buffer from the NE-PER Nuclear and Cytoplasmic Extraction Kit (Thermo Scientific, MA, USA), which was supplemented with 1X proteinase inhibitor from a 25X stock of Complete EDTA-free Protease Inhibitor Cocktail (Roche, Switzerland). The mixture was vortexed for 15 seconds, and kept on ice for 10 minutes, then vortexed again, and supplemented with 55ul ice-cold CER II solution. The mixture was vortexed again for 5 seconds, kept on ice for 1 minute, vortexed again, and then centrifuged in a pre-chilled 4°C centrifuge, at 16,000xg for 5 minutes. The supernatant was transferred to a new pre-chilled tube as cytoplasmic fraction, which was not used in the pulldown. The pellet remaining now contained the nuclear protein fraction.

The nuclear isolation buffer was prepared using as described in the protocol paper (Bai et al. 2016) with the following components; 25mM Tris-HCl Buffer (pH 7.5), 150mM KCl, 2mM MgCl₂, 0.5% NP-40, 1mM PMSF, 1X Protease Inhibitor and EDTA, 1mM DTT. All buffers were prepared in cold DEPC-treated double-distilled water, to avoid RNases. DTT was to be added after concentration measurement.

The nuclear fraction from the pellet of the previous step was moved into a pre-chilled 15ml, tight, glass Dounce homogenizer, and 1 ml nuclear isolation buffer was added per pellet. The nuclear fraction was sheared with 100 strokes in the Dounce homogenizer.

The homogenized lysate was transferred to pre-chilled eppendorfs, and centrifuged at 4°C for 17 minutes at 17000xg. The supernatant was then collected in another pre-chilled tube as the nuclear lysate. The pellet containing DNA, some chromatin and insoluble components were discarded.

Protein concentration of nuclear isolate was calculated using BCA Protein Assay Kit (Pierce, MA, USA), using an aliquot taken before addition of DTT. Just before combining with RNAs for pulldown, the protein lysate would be supplemented with 0.4 U/ul RNase Inhibitor (Promega, WI, USA).

3.1.6 RNA Pulldown of nuclear protein lysates with *KLK3e* lncRNA and controls

Preparation of RNA

30pmols of each biotinylated RNA sample cleaned-up from in-vitro transcription reactions as described above were all adjusted to a final equal volume with RNase-free or DEPC-treated water, usually 120ul, depending on their concentrations. The ng to pmol conversion of each RNA depended on its size, and the volume required of each were calculated according to these values;

	<u>~Size (nt)</u>	<u>1pmol in ng</u>	<u>30pmol in ng</u>
KLK3e	1984	671 ng	20,130 ng
Xist	1840	585 ng	17,550 ng
CherryGFP.	1573	506 ng	15,180 ng

The RNA were then denatured at 90°C for 2 minutes, or 70°C for 8 minutes, allowed to cool at room temperature with 2 minutes, then supplemented with 2X RNA Secondary Structure Buffer containing fresh DTT and RNasin (20 mM Tris-HCl pH 7.4, 200mM KCl, 20 mM MgCl₂, 2 mM DTT, 0.8 U/μl RNase inhibitor), and allowed to fold at room temperature for 30 minutes.

Preparation of Streptavidin Magnetic beads

40 μ l (400 μ g) Dynabeads M-280 Streptavidin magnetic beads (Invitrogen, CA, USA) per sample and one for protein clean-up (40 $\times$ 4=160 μ l) were washed and prepared as described in their manual, as summarized below. The beads were first washed 5 minutes in 1 ml Wash & Bind buffer (5 mM Tris-HCl pH7.4, 1 M NaCl) on rotation, and buffer was removed using DynaMag-2 Magnet magnetic rack (Invitrogen, CA, USA). The beads were then washed 2 times 2 minutes in 500 μ l Buffer A (0.1 M NaOH, 0.05 M NaCl), followed by 2 times 2 minutes in 500 μ l Buffer B (0.1 M NaCl), all followed by magnetization and buffer removal. The beads were finally resuspended in 320 μ l Wash & Bind buffer and distributed into 4 eppendorfs, 80 μ l in each. Buffer was removed before either RNA sample or proteins for pre-clearing were added onto each.

Binding RNA onto Streptavidin Magnetic beads & Protein pre-clearing

3 RNA samples that have finished folding at room temperature were supplemented with fresh DTT and RNAsin, and put into 3 eppendorfs containing Streptavidin magnetic beads. The biotinylated RNA was allowed to bind onto the beads with 1 hour incubation on slow rotation at room temperature.

In the meantime, the nuclear protein lysate (~450 μ l per sample) was pre-cleared with magnetic beads to remove any proteins that bind non-specifically to the beads. The lysate was added onto the beads and incubated at 4°C for 1 hour with slow rotation.

Combining protein with RNA bound on beads

After the 1 hour binding of the RNA onto the beads was complete, the beads were washed 2 times 5 minutes in Wash & Bind Buffer supplemented with DTT and RNAsin (5 mM Tris-HCl pH7.4, 1 M NaCl, 0.5mM DTT, 0.2U/ μ l RNAsin) on rotation.

Finally, the pre-cleared protein lysate was supplemented with fresh DTT and RNAsin (0.4 U/ μ l) and added onto the beads that should now have RNA bound.

The RNA-beads and protein mixture was incubated 3 hours at 4°C on slow rotation.

Washing & Elution of bound proteins

After the 3 hour incubation was finished, the beads, which should now have proteins interacting with the attached RNA, were washed 6 times for 2 minutes in cold wash buffer (25 mM Tris-HCl pH 7.4, 150 mM KCl, 2 mM MgCl₂, 0.25% NP-40, 1x Protease Inhibitor with EDTA, 1 mM DTT, 40U/ml RNase inhibitor) on fast rotation, and using magnet to remove buffer in between washes.

After washes were complete, the beads were magnetized, and buffer was removed. 25ul 4X Laemmli buffer (Bio-Rad, 50mM DTT) was added to the beads and they were boiled for 5 minutes at 95C. Magnetic rack was applied to separate the beads. The sample buffer containing the proteins from the beads at this stage could be stored at -20°C overnight.

Running & Staining Protein Gel

The protein samples were loaded and run in a 10 30ul well 4–15% Mini-PROTEAN TGX precast gel (Bio-Rad, US) at 25mA for ~55 minutes. After running, the gel was stained using the high-sensitivity Sypro Orange Protein Gel Stain (Bio-Rad, US) which was prepared 12ul in 40ml 7.5% Acetic Acid solution, for 1 hour at room temperature. the whole gel staining was then visualized using a UV transilluminator, and distinct bands and/or lanes were cut and submitted for in-gel digestion and mass spectrometry analyses at the Proteomics facility.

3.1.7 Mass Spectrometry Identification of Pulled-down Proteins

For mass spectrometry, the samples were submitted to the proteomics facility of Koç University, under the supervision of Dr. Nurhan Özlü. The proteins were trypsin digested in-gel, and cleaned with Stage-Tip method. Peptides were detected using the Thermo Scientific Q Exactive Quadrupole-Orbitrap Mass Spectrometer with nano Liquid Chromatography, UltiMate 3000 NCS-3500RS (n-LC MS/MS) system and matched to UniProt database. At least 2 technical replicates were performed per sample.

Initial mass spectrometry run was performed on 7 bands specific for the KLK3e and Xist samples excised from the gel shown in Figure 4.7 in the results section. A replicate

pulldown was performed, and this time the whole lane of proteins bound by KLK3e, Xist and CherGFP control were cut separately, and mass spectrometry was performed on the whole lane.

To select for proteins that have specificity towards KLK3e from the second mass spectrometry analysis with 3 lanes of samples as KLK3e, Xist and CherGFP, the proteins that met the following criteria were selected;

$$\Sigma\text{Coverage (KLK3e)} > \Sigma\text{Coverage(Xist)} + \Sigma\text{Coverage(CherGFP)} \text{ AND} \\ \Sigma\text{Coverage(CherGFP)} < 30$$

This yielded 143 proteins. These were then used for validation of proteins detected in the 4 specific bands cut from KLK3e in the initial pulldown. The proteins detected in the bands cut from KLK3e which were also validated with the criteria above (an intersection of the two replica pulldowns) were 41 proteins. A table of these proteins confirmed in two replicas, is provided in the Mass Spectrometry Results section in Chapter 4. Details and coverage values are provided in Appendix Table A.1.

3.1.8 Modeling of the KLK3e RNA-protein complex

In order to develop a structural model of the RNA-protein complex, initially the structural availability of the identified protein interactors of interest were investigated. Crystal structures of DDX5 were available in the PDB, covering the entire helicase domain. The other proteins of interest either had no available crystal structures, or the structures available did not belong to domains implicated to interact with RNA and/or AR, or the interacting domain on AR was unavailable.

First a template-based 3D model of the KLK3e RNA was obtained using the ModeRNA server (Rother et al. 2011), based on the top matching template PDB ID: 3CC20. AR (PDB ID:5JJM) and DDX5 (PDB ID:3FE2A) were docked using PRISM (Tuncbag et al. 2011; Baspinar et al. 2014). And finally, the RNA-protein docking of KLK3e RNA 3D model and the AR-DDX5 complex obtained by PRISM were docked using the PatchDock server (Duhovny et al. 2002; Schneidman-Duhovny et al. 2005). These steps are depicted in Figure 3.2 and are described below in further detail.

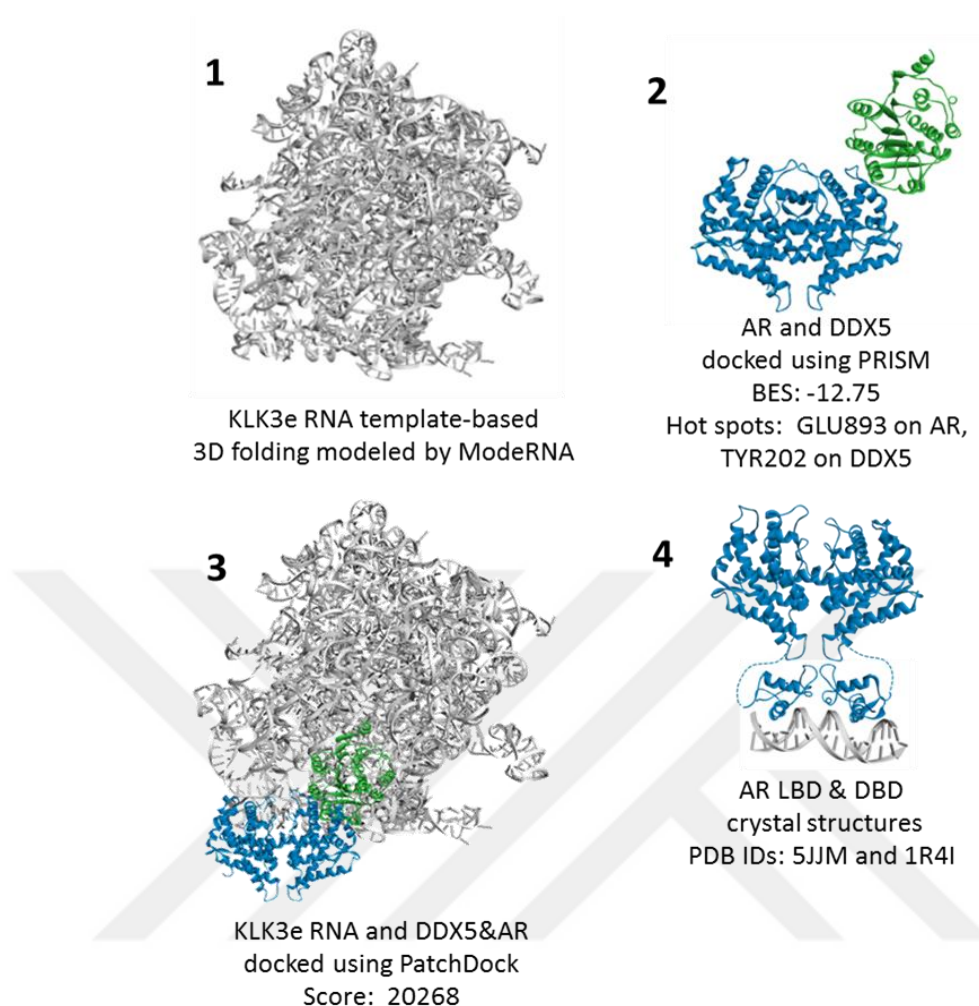


Figure 3.2: Modeling steps of the proposed RNA-protein complex. First, the 3D structure of the KLK3e RNA was modeled using ModeRNA (Rother et al. 2011) which is a template based tool that can handle large RNA. The template matched was a ribosomal RNA with PDB ID 3CC2-0. Secondly, the complex between AR and DDX5 was modeled using PRISM (Tuncbag et al. 2011; Baspinar et al. 2014) with the PDB structures 5JJM and 3FE2-A. The structure that allowed room for the AR DNA binding domain and that had a passing energy score was selected; a binding energy score of -12.75. The interacting hot spots are GLU893 on AR and TYR202 on DDX5, according to HotRegion (Cukuroglu et al. 2011). Third, the RNA-protein complex was modeled using the default parameters on PatchDock (Duhovny et al. 2002; Schneidman-Duhovny et al. 2005), as it was a tool that performed well in the CAPRI assessments and was able to handle large RNA as input. The best scored docking structure was selected, with a score of 20268. Finally, the DNA binding region of AR was incorporated. Details of these steps are explained in more detail in text.

Selection of Tools for the Modeling and Docking of KLK3e RNA

In order to model KLK3e, we needed an RNA 3D structure prediction tool that was able to handle the full size of KLK3e which is 2170 nucleotides (Hsieh et al. 2014). After investigation and testing, the ModeRNA tool published in *Nucleic Acids Research* (Rother et al. 2011) was the tool of choice because other tools available were not able to model the full size RNA. For example, the RNA Composer tool (Popenda et al. 2012), is only able to model a maximum of 500 nucleotides. As shown in the Appendix Figure A.2, modeling of the first 500 bases only results in a structure with several hairpin protrusions. As shown in Figure 3.2, the full size modeling with ModeRNA (Rother et al. 2011) results in a sponge-like structure with 3D motifs forming many cavities, which is expected to be formed by an RNA of this size. As shown in Appendix Figure A.3, a sample electron microscopy structure of an RNA-protein complex of an RNA of similar size with KLK3e shows a similar-sponge like structure. This structure with PDB ID 5FLX shows the 1869 nucleotides long hepatitis C viral RNA bound by human ribosomal proteins, the RNA acting like sponge and allowing docking of many proteins. For docking of the RNA and protein, we had to find a docking tool that accepts RNA input and can handle the large size of KLK3e. We referred to the docking tools that are top performers in the CAPRI assessment (Lensink et al. 2017) and after some testing we saw that only PatchDock is able to handle RNA input with the size of KLK3e. Another close tool was ClusPro (Kozakov et al. 2017), which also accepted the KLK3e RNA as an input but resulted in an error that the lines in output are not enough; the tool was able to work with an RNA of about 600 nucleotide size maximum.

Modeling of KLK3e RNA 3D Structure with ModeRNA

The ModeRNA tool was published in *Nucleic Acids Research* journal in 2011 (Rother et al. 2011). It is a template-based modeling tool, which was developed with the inspiration to create a comparative modeling tool for RNA similar to homology based modeling tools available for proteins, such as SwissModel (Waterhouse et al., 2018). The current webserver of ModeRNA is able to search for templates available in the PDB based on the RNA sequence provided, and generate an alignment file. Then, using this alignment and the template structure from the PDB, it can output a template-based

3D model of the RNA in pdb format. In summary, the tool initially copies coordinates from the regions that are identical with the template, performs base exchange operations on regions with substitutions, and tackles gap regions with the fragment insertion approach (Rother et al. 2011), similarly to SwissModel (Waterhouse et al., 2018). ModeBase can search matches from its own fragment libraries that contain RNA structure fragments up to 100 bases long, which were developed to cover all known types of local RNA structures. The fragment complementation stage is composed of an initial pre-filtering step that looks for a geometrical match of the flanking bases to the fragments of appropriate size from the library, followed by a fitting stage that inserts and tests the insertion of the 50 most promising fragment candidates. If a match cannot be found for a gap, the tool still generates the model with the warning of an unsealed gap. The tool also eliminates clashes and performs backbone repairs on imperfectly matched regions. The tool has been shown to perform well with modeling tRNAs of ~100 nucleotides long, as well as ribosomal RNAs of ~1500 nucleotides long. (Rother et al. 2011)

For the modeling of KLK3e, initially a template search was performed on the online submission section of the ModeRNA web server. This search identified the ribosomal RNA structure with PDB ID 3CC2-0 as the best matching template with a good resolution. The alignment file was then used to model the 3D structure of the KLK3e RNA. There were no warnings such as unsealed gaps. The RNA 3D structure was then minimized using the Yasara Energy Minimization Server (Krieger et al. 2009).

Modeling of the AR-DDX5 Protein-Protein Interaction using PRISM

The main working principles of PRISM (Tuncbag et al., 2011, Baspinar et al., 2014) are described in Section 3.3.1. In order to model the interaction between AR and DDX5, the PDB structures 5JJM and 3FE2 were used for AR ligand binding domain and DDX5 helicase domain, respectively. Two PDB structures were available for DDX5; 3FE2 had a better resolution of 2.6Å compared to 4A4D with a resolution of 2.7Å. PRISM generated several models, we selected the structure with DDX5 docked away from the side of AR that faces the DNA binding domain, and that had a binding energy score below -10. This was the model with the template interface 3nasAB and binding energy score -12.75, for targets 5JJM and 3FE2A, details of which are below;

PRISM Output;

Target1	Target2	Interface	Energy	Structure
3fe2A	5jjm	3nasAB	-12.75	View

Docking the KLK3e RNA with the AR-DDX5 Complex using PatchDock

As mentioned above, among top performing docking servers in the CAPRI assessment, PatchDock was the best option for our modeling purposes as it can handle large RNA input. The PatchDock tool was published in the Nucleic Acids Research journal (Duhovny et al. 2002; Schneidman-Duhovny et al. 2005), and was among the top 10 performers in the 2017 CAPRI assessment of docking servers (Lensink et al. 2017), as shown in Appendix Table A.4.

PatchDock is a geometry-based molecular docking algorithm that docks molecules based on shape-complementarity, similar to a jigsaw puzzle. Its high efficiency allows large-scale docking in short run times. The molecules are initially divided into concave, convex and flat surface patches. Then, candidate transformations are generated by matching complementary surface patches, i.e. concave to convex. The candidate match transformations are evaluated with a scoring function that assesses both geometric fit, eliminating clashes, as well as atomic desolvation energies. Finally, in order to eliminate redundant solutions, a root mean square deviation (RMSD) clustering is utilized. (Duhovny et al. 2002; Schneidman-Duhovny et al. 2005)

In order to obtain a structural model of the KLK3e with the AR-DDX5 complex, we have docked the KLK3e RNA 3D structure obtained using ModeRNA and minimized using Yasara, with the AR-DDX5 complex model we have obtained using PRISM. Patchdock was used with the default parameters of 4.0 clustering RMSD and default complex type. The top scoring model from PatchDock was used to propose the model shown in the Results section in Figure 4.10.

PatchDock Results;

Solution No	Score	Area	ACE	Transformation	PDB file of the complex
1	20268	4024.5	15.04	0.94 -0.99 1.66 82.59 28.63 167.90	result.1.pdb

3.2 Characterization of Novel Androgen Receptor Antagonists and their Role in Transcription

Treatments of cells with small molecules were performed at the Vancouver Prostate Centre by Flora Huang, followed by Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) and RNA sequencing. These experimental supplementary methods are provided in Appendix B. Two more biological replicas of the RIME analyses are being performed and will be added to the analysis.

3.2.1 Data Analysis of Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) results

RIME protocol (Mohammed et al. 2016) was applied using antibodies targeting androgen receptor (AR) in Vancouver Prostate Centre, as described in the Supplementary Methods in Appendix B.

Hierarchical Clustering

Tree diagrams in the form of dendograms that depict hierarchical clustering are commonly used for comparative understanding of omics scale biological data, and represent the rate of similarity across samples.

Perseus software is a tool published in Nature Methods in 2016 as a computational platform for comprehensive analysis of (prote)omics data (Tyanova et al. 2016).

The complete LFQ intensity data obtained from the RIME analysis was uploaded onto the Perseus software (Tyanova and Cox 2018), and log2 transformation and missing value imputation from normal distribution was applied on the LFQ intensity values, as advised for normalization of label-free mass spectrometry data (Karpievitch et al. 2012).

Then, hierarchical clustering of the normalized RIME mass spectrometry data was performed on Perseus (Tyanova and Cox 2018), using the default settings of; k-means pre-processing with maximum 10 iterations, Euclidean distance, average linkage and no constraints. The resulting dendogram represented overall similarity of protein profiles detected across samples.

Venn Diagrams

Venn Diagrams are also commonly used in biological sciences to compare sets of data such as genes or proteins in a simple visual representation.

The RIME data was analyzed to select for DHT regulated AR interactors that are disrupted by the drugs, in other words; proteins not present in EtOH and IgG control, present in DHT treated samples, and not present with one or more of the compounds. This corresponded to proteins that have a 0 LFQ intensity reading value in IgG negative control, a 0 LFQ intensity reading value in EtOH control, have a positive LFQ intensity reading value in DHT treated sample, and have a 0 LFQ intensity reading value in one or more of the compounds, based on readings from the mass spectrometry. Replicas are to be performed after which statistical analyses will be carried out.

For each small molecule sample, the sets of genes that have 0 LFQ intensity readings were uploaded onto the bioinformatics Venn diagram tool available online at <http://bioinformatics.psb.ugent.be/webtools/Venn/>. The intersecting regions show interactions that appear to be disrupted by more than one molecule in common. Interactions that appear to be disrupted uniquely by one molecule can also be seen.

3.2.2 RNA Sequencing Data Analysis

PolyA selected RNA were sequenced using the BGISEq500 platform at the BGI. The raw reads were pre-filtered by removing adaptor sequences, contamination and low-quality reads. The high quality of the filtered reads was confirmed using the FastQC (Andrews 2014) quality control tool. The reads were mapped and annotated using STAR Aligner (Dobin et al. 2013) (version 2.6.1c) using the latest release reference genome and annotations from the GENCODE (Frankish et al. 2018) project; release 29, GRCh38.p12. The reads were then counted using featureCounts (Smyth et al. 2013). Differential gene expression analysis for different treatments was then performed using the DESeq2 (Love et al. 2014) method.

3.2.3 Gene Set Enrichment Analysis (GSEA) of AR targeting molecules

Gene Set Enrichment Analysis was performed with the pre-ranked method.

Genes were ranked according to the $-\text{Log}_{10}(\text{P-value}) * \text{Sign}(\text{Log}_2\text{FoldChange})$ metric, which takes into account the significance and direction of expression changes, as used in various publications (Guryanova et al. 2016; Pandiyan et al. 2016; Weinberg et al. 2019). Non-coding genes were removed as they are generally not included in GSEA datasets and can alter the significance of coding genes. Genes with p-values of 1 were naturally eliminated as their ranking score equaled 0.

Pre-ranked genes were searched against the MSigDB Hallmark gene sets using the default 1000 permutations, 'weighted' enrichment statistic parameter and 'meandiv' normalization, including gene sets with more than 15 and less than 5000 members. Resulting gene sets were considered significantly enriched if they had an FDR q-value lower than 0.25, as advised in the GSEA User Guide and FAQ on interpretation of results. (Mootha et al. 2003; Subramanian et al. 2005).

3.3 Protein Interaction and Oligomerization Analyses of Glucocorticoid Receptor

3.3.1 PRISM computational modelling of protein-protein interactions

The PRISM: Protein Interactions by Structural Matching tool (Tuncbag et al., 2011, Baspinar et al., 2014) was used to model MR-GR interactions. The PRISM algorithm is a template-based protein interaction prediction tool that utilises interaction structures available in the Protein Data Bank (PDB) (Berman et al., 2000) as template to predict structural interaction between two query proteins, as depicted in Figure 3.3 below. PRISM generates multiple possible models using template interfaces that show a good match to any possible interaction interface between the two query proteins. Rigid docking results are then flexibly refined to resolve steric clashes, outputting a pdb file of the predicted interaction structure and a list of interacting amino acids in the predicted interface.

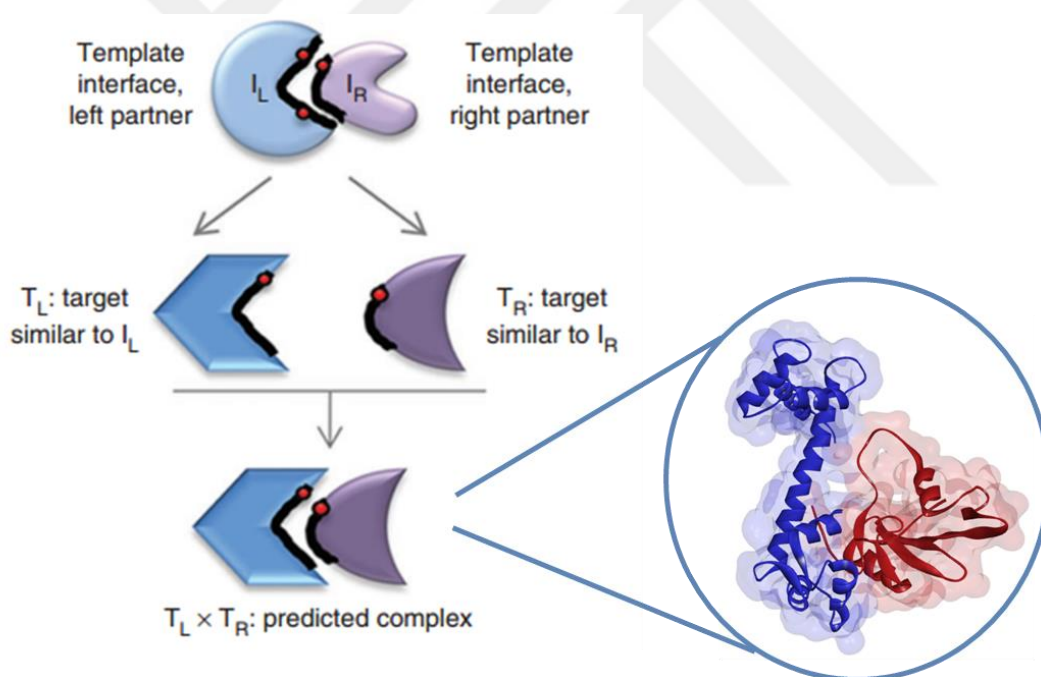


Figure 3.3: The PRISM: Protein Interactions by Structural Matching tool (Tuncbag et al., 2011, Baspinar et al., 2014) is a template-based protein interaction prediction tool that utilises interaction structures available in the Protein Data Bank (PDB) (Berman et al., 2000) as template to predict structural interaction between two query proteins.

3.3.2 *Energy calculations, hot spot and mutation analyses*

Binding energies of PRISM predictions were calculated using ZRANK (Pierce and Weng, 2007), an optimized energy function providing an arbitrary energy score for comparative analyses allowing ranking of multiple protein docking predictions. Energies were computed for all PRISM interaction predictions utilizing interfaces from the nuclear receptor protein family. For each possible interface per query pair of proteins, the prediction with the lowest energy score for that interface was used.

DNA binding and ligand binding domain structures of GR and MR were obtained from Protein Data Bank. Rat homology models were constructed using SwissModel (Waterhouse et al., 2018) where only the human or mouse structures were available. To ensure that homology modelling of structures would be reliable, a high sequence identity between rat and human or mouse homologs was confirmed using protein BLAST (Altschul et al., 1990).

Hot spot amino acid residues that contribute most of the binding energy to the protein-protein interaction interfaces (Bogan and Thorn, 1998, Cavga et al., 2015) predicted by PRISM were computed using HotRegion (Cukuroglu et al., 2012). HotRegion additionally identifies the regions of hot spot clusters important for the stability of protein complexes.

To determine mutations with deleterious effects on the interaction between the DNA binding domains of GR and MR, changes in binding affinity ($\Delta\Delta G$) were calculated using several tools; FoldX (Schymkowitz et al., 2005, Van Durme et al., 2011), BeatMusic (Dehouck et al., 2013), mCSM (Pires et al., 2014) and MutaBind (Li et al., 2016).

3.4 RNA-seq analysis of Cytochrome deletion in a p53 mutant background

Cell culturing, UV treatment, RNA extraction, RNA-seq library preparation, RNA-seq quality check, and qPCR validation procedures were performed by Dr. Mehmet Tardu from the laboratory of Prof. Dr. Halil İ. Kavaklı in Koç University. Validation of knock-outs and increase of apoptosis were performed by Tuba Korkmaz and Assoc. Prof. Nuri Öztürk in Gebze Technical University. These supplementary methods and results are provided in Appendix D.

3.4.1 RNA sequence Alignment & Differential Analysis

STAR aligner (v2.5.2) (Dobin et al. 2013) was used to align reads to the mouse reference genome GRCm38. Following alignment, a Pearson's correlation analysis was performed to obtain the transcript-level R^2 value between replicates of each condition. After mapping, Cuffdiff command in Cufflinks suite (Trapnell et al. 2012) was used with default parameters to identify differentially expressed genes between UV-treated and non-treated cells. Quartile library normalization and pooled dispersion estimation methods were performed for the differential expression analysis. Fragments per kilobases of exons per million mapped read (FPKM) and fold changes were calculated for each sample.

Differential expression analysis was proceeded with transcripts with a minimum alignment count of 10, a false discovery rate (FDR) <0.01 , and an 'OK' status from Cufflinks from the $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ and the $p53^{-/-}$ non-treated vs. UV-treated differential expression analyses, which allowed them to be inter-compared. Among the mapped and differentially analyzed transcripts, an expression change was considered significant only if it was significant according the cutoff by Cufflinks, and if it showed a fold change more than 2 ($\log_2|\text{foldchange}| \geq 1$).

3.4.2 Identification of differentially expressed genes (DEGs)

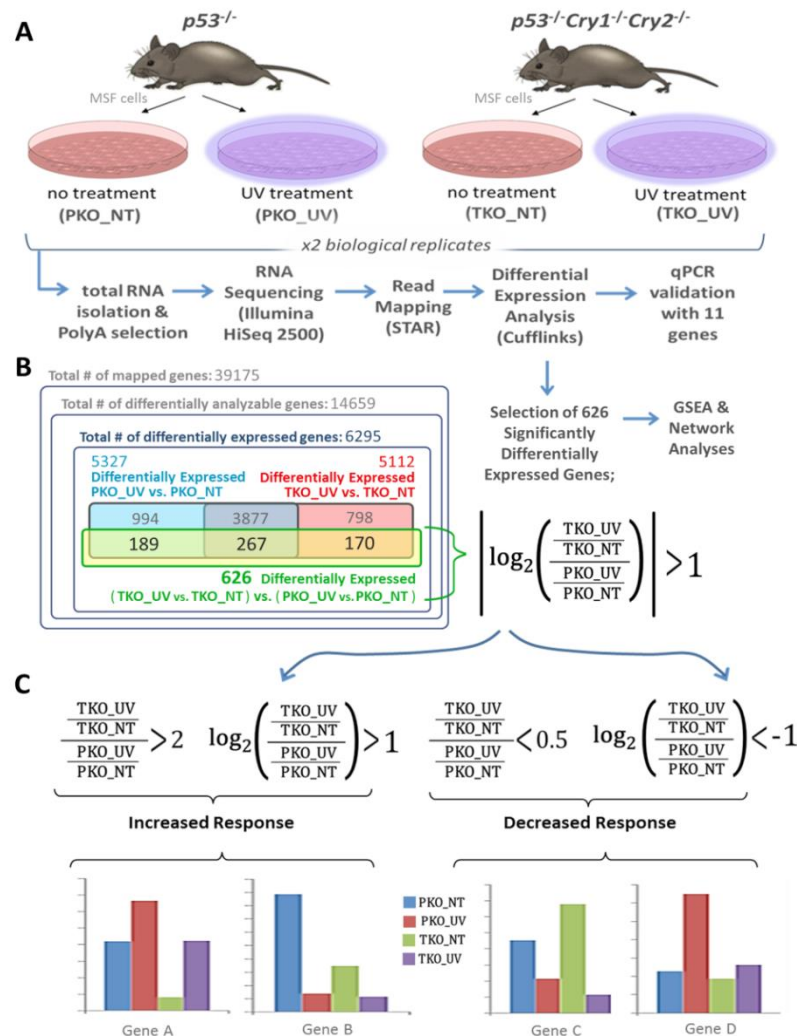


Figure 3.4: Experimental setup and RNA-Seq analysis methodology. **A)** Mouse skin fibroblast (MSF) cells from $p53^{-/-}$ (PKO) or $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ (TKO) mice either received no treatment or were treated with UV, then total RNA was isolated and RNA sequencing analysis was performed as described in the flowchart. **B)** The genes of interest were selected as those with a significant TKO_UV vs. TKO_NT to PKO_UV vs. PKO_NT ratio above 2, as formulized. This resulted in 626 genes of interest that are differentially regulated in response to UV treatment in $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ cells compared to $p53^{-/-}$ cells. **C)** The significantly differentially expressed genes were scored and grouped as “increased response” genes and “decreased response genes” according to the formulized values.

The focus of the present research was to identify the genes that were differentially expressed in response to UV treatment upon deletion of *Cry* genes in a $p53^{-/-}$ mutant background. Therefore, each gene was assigned a DEG Score according to the change in change they showed upon UV treatment in comparison of the $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ cells to the $p53^{-/-}$ cells, as shown in Equation 1;

$$\text{DEG Score} = \log_2 \left(\frac{\left(\frac{\text{TKO}_{UV}}{\text{TKO}_{NT}} \right)}{\left(\frac{\text{PKO}_{UV}}{\text{PKO}_{NT}} \right)} \right) \quad \text{Equation 3.1}$$

As described in **Figure 3.4.B**, significantly differentially expressed genes (DEGs) were identified as those with an absolute DEG Score above 1, indicating a change in change more than two fold. This threshold yielded 626 differentially expressed genes in response to UV-treatment in comparison of the $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ cells to the $p53^{-/-}$ cells. As described in **Figure 3.4.C**, a DEG Score above 1 indicated an increased response to UV in $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ cells compared to $p53^{-/-}$ cells. Correspondingly, a DEG Score below 1 indicated a decreased response to UV in $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ cells compared to $p53^{-/-}$ cells.

3.4.3 Gene Set Enrichment Analysis (GSEA) of DEGs in UV response in $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ cells compared to $p53^{-/-}$ cells

The Gene Set Enrichment Analysis (GSEA) (Subramanian et al. 2005) was performed using the Differentially Expressed Gene Score (DEG Score, Eqn. 1) for each gene among the 626 DEGs as a pre-ranked input. Enriched gene sets were searched from the MSigDB Hallmark data collection using the default ‘weighted’ enrichment statistic parameter and ‘meandiv’ normalization, including gene sets with more than 15 and less than 5000 members. Enriched gene sets with a nominal p-value < 0.05 were considered significant.

The MSigDB Hallmark gene sets are composed of genes that show consistent expression patterns in specific biological processes across many founder MSigDB gene sets based on publications. Since these gene sets have been carefully curated to include only genes that were consistently observed to be involved in a specific pathway, they

have high reliability. The “HALLMARK_TNFA_SIGNALING_VIA_NFKB” set is made up of 200 genes regulated by NF- κ B in response to TNF generated from 132 founder gene sets. The “HALLMARK_INTERFERON_GAMMA_RESPONSE” gene set is composed of 200 genes up-regulated in response to IFNG across 82 founder gene sets. The founder gene sets are obtained via expression profiling by array (Subramanian et al. 2005).

3.4.4 KEGG Pathway Enrichment Analysis

KEGG Pathway enrichment was analyzed with the ClueGO Extension Tool of Cytoscape (Bindea et al. 2009). The 626 significant DEGs list was loaded as a Marker List and *Mus musculus* was selected. In the ClueGO Settings Section, the Pathway was selected as KEGG, and a threshold of significance was set as “Show only pathways with $pV \leq 0.05$ ”. The remaining settings were left as default. The resulting gene list in the “Functional Groups with Genes” output document was included in the Venn diagram.

3.4.5 Protein-Protein Interaction Network Analysis

PPI network analysis in relation to the p53 protein was performed in order to further elaborate the biological mechanisms that may be involved in the delayed onset of cancer in *p53^{-/-}Cry1^{-/-}Cry2^{-/-}* compared to *p53^{-/-}* knockout mice. STRING database was used to obtain experimentally validated PPI information (Jensen et al. 2009).

Of the 626 DEGs, ribosomal, proteasomal or polymerase-related genes were eliminated because since their interactions are the studied extensively, they dominate the network when they are included and disable the cancer related pathways from appearing. This elimination, and the elimination of genes unrecognized by STRING resulted in 487 genes to be used in the input list (Appendix Table D.2).

The PPI network data was pulled from the STRING database (Jensen et al. 2009) with the following criteria; a maximum of 50 additional interactors per protein only in the first shell, based exclusively on experimental data and with a confidence score of 0.4 or above. The STRING output is provided in the references as a permanent link. (Jensen et

al. 2018) The PPI network data obtained from STRING database was visualized and analyzed using Cytoscape (Shannon et al. 2003). The proteins were clustered with the default perfused force directed layout algorithm of Cytoscape and only slightly modified for better visualization. Detached nodes and pathways were deleted, leaving only the largest pathway that included p53 in focus.



Chapter 4:

UNDERSTANDING THE ROLE OF LNCRNA IN ANDROGEN RECEPTOR MEDIATED TRANSCRIPTION

The aim of this project was to study how an enhancer RNA influences AR mediated transcription. In vitro RNA pulldown of KLK3e followed by mass spectrometry was used to identify the proteins this lncRNA interacts with. The RNA-protein complex was computationally modeled to propose how this lncRNA influences transcription by AR.

4.1 *PCR Amplification of KLK3e and controls*

The KLK3e lncRNA transcript was amplified using Reverse-Transcription Polymerase Chain Reaction (RT-PCR) as described in the Methods section, to provide a template for the in-vitro transcription of biotinylated KLK3e lncRNA.

The primers were designed based on the ~2000 bases long sequence described in the KLK3e paper (Hsieh et al. 2014). And this long KLK3e transcript was confirmed to be at the RNA level using and RT- control. The RT- controls may have any remaining genomic DNA contaminants, but have no complementary DNA of the total RNAs; if the amplification was from genomic DNA, the band would show up in RT- controls as well, which is not the case, confirming that an RNA is being transcribed matching the sequence provided in the study.

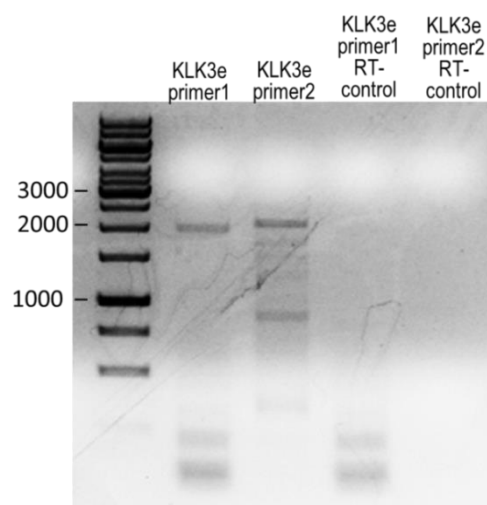


Figure 4.1: KLK3e transcript of ~2kb was confirmed to be at the RNA level.

Xist was selected as a positive control as a well-established lncRNA with known protein interactors. CherryGFP is used as a negative control to account for general affinity of RNA to proteins, as it is merely a transcript of the same size that should not have a function at the RNA level.

The gel image in Figure 4.2 below shows that the 3 transcripts, KLK3e, Xist and CherryGFP are all at similar, expected sizes, which would allow comparability in the upcoming RNA pulldown. Also, all transcript DNAs were obtained as a clean single band after PCR cleanup, as described in the Methods section, and these would provide good template DNAs for specific and efficient *in-vitro* transcription of RNA in the subsequent step.

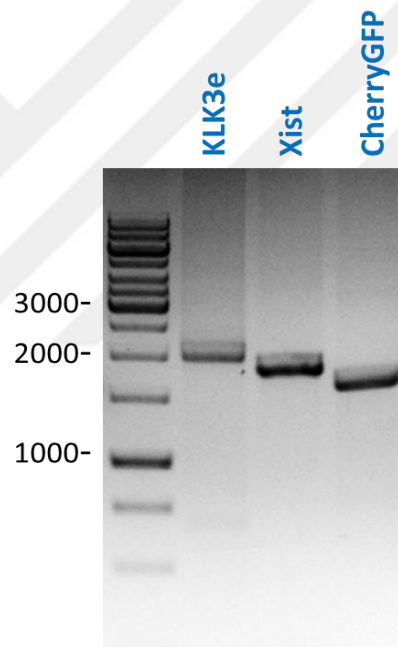


Figure 4.2: KLK3e transcript is of similar sizes with Xist and CherryGFP controls. All transcripts were obtained as a clean single band, without any non-specific bands, after PCR-cleanup of samples.

4.1.1 Sequence confirmation of amplified KLK3e and analysis of matches

The next step was to confirm the sequences of the KLK3e and Xist transcript templates that were amplified by RT-PCR. The cleaned up PCR products were sent to sequencing by Macrogen Europe. A BLAST search was made on the sequences that best match the sequence reads from the KLK3e and Xist transcripts, results of which are shown below.

The alignments of the sequences to the best matches, as well as to the sequence provided in the KLK3e publication, are provided in the Appendix section A.

Table 4-1: Sequences producing significant matches to the KLK3e transcript

Description	Max Score	Total Score	Query Cover	E value	Percent Identity	Accession
Homo sapiens KLK3 upstream enhancer/promoter region (LOC106128905) on chromosome 19	2126	2126	89%	0.0	95.08%	NG_042790.1
Homo sapiens kallikrein related peptidase 3 (KLK3), RefSeqGene on chromosome 19	2126	2126	89%	0.0	95.08%	NG_011653.1
Homo sapiens chromosome 19 clone LLNLF-214C7, complete sequence	2126	2126	89%	0.0	95.08%	AC011523.3
Homo sapiens serine protease gene cluster, complete sequence	2119	2119	89%	0.0	95.01%	AF243527.1
Homo sapiens prostate cancer antigen (KLK3) gene, promoter region	2115	2115	89%	0.0	94.93%	AF394907.1
Human prostate specific antigen gene, 5' flanking region	2069	2069	89%	0.0	94.42%	U37672.1
Homo sapiens LincPSA isoform lincPSA-3 lncRNA, complete sequence	1932	1932	81%	0.0	95.07%	MF521460.1
Homo sapiens LincPSA isoform lincPSA-2 lncRNA, complete sequence	1932	1932	81%	0.0	95.14%	MF521459.1
Homo sapiens uncharacterized LOC105372441 (LOC105372441), transcript variant 1, long non-coding RNA	433	433	15%	4e-118	100.00 %	NR_131205.1
Homo sapiens uncharacterized LOC105372441 (LOC105372441), transcript variant 2, long non-coding RNA	433	433	15%	4e-118	100.00 %	NR_131203.1

As can be seen from Table 4-1 above, the sequence read matches the enhancer/promoter region upstream of the KLK3 gene, which encodes the Prostate Specific Antigen (PSA), as intended. The sequence read matches a long-non coding RNA record with accession number NR_131205.1 with a 100% match score in 433 bases, and it matches the enhancer/promoter region of KLK3e, as well as another lncRNA record called lincPSA with match scores about 95% in about 2000 bases.

Since lncRNAs are a relatively new field, it can be expected that several database records exist that match the same sequence with slight variations. The short (433 bases) RNA match with accession number NR_131205.1 shows the KLK3e publication by Hsieh et al. as a reference, however, it is unclear why the database shows a sequence length of only 446 bp, while the publication provides a 2170 base long sequence in the Supporting Information (Hsieh et al. 2014).

The other lncRNA match, with a 95% score in 1932 bases, named “Homo sapiens LincPSA isoform lincPSA-3 lncRNA” is associated with a publication that describes androgen receptor-associated compared non-associated lncRNAs (daSilva et al. 2018). This publication lists linc-PSA-1 or linc-KLK3-1 among 893 lncRNAs detected as differentially expressed under androgen stimulation (daSilva et al. 2018, Supplementary Table 4) and 418 lncRNAs upregulated in prostate tumors compared to normal tissue (daSilva et al. 2018, Supplementary Table 6). The relevant sections are copied below; (daSilva et al. 2018)

893 lincRNAs were detected as differentially expressed under androgen stimulation;

<u>cufflinks_code</u>	<u>GeneID</u>	<u>Gene_name</u>	<u>log2ratio(Androgen/Vehicle) in LNCaP</u>	<u>FDR (0.05)</u>
XLOC_060509	TCONS_00027080	linc-KLK3-1	1.665973325	0.00380697

418 up-regulated and 6 down-regulated in tumor;

<u>cufflinks_code</u>	<u>GeneID</u>	<u>Gene_name</u>	<u>log2ratio(Tumor/Normal) reanalysis of data from Ren et al. 2012</u>	<u>FDR (0.05)</u>
XLOC_060509	TCONS_00027080	linc-KLK3-1	2.12565244	5.77E-08

In this publication, they have also performed an RNA immunoprecipitation (RIP) assay to identify RNA interactors of AR. When they immunoprecipitated AR, they performed a qPCR validation to check for previously published AR interacting lncRNAs KLK3e (Hsieh et al. 2014) and PCGEM1, against the negative control MALAT1 (Yang et al. 2013). KLK3e and PCGEM1 were confirmed to be among AR interactors when checked with qPCR (daSilva et al. 2018, Supplementary Figure S2). However, then they

performed RNA sequencing analysis, KLK3e and PCGEM1 were not among significantly AR associated lncRNAs (daSilva et al. 2018, Supplementary Table 8).

“A total of 619 lincRNAs were identified as significantly enriched ($FDR < 10\%$, DESeq2) in the anti-Androgen Receptor (antiAR) fraction in relation to the control fraction (non-specific IgG)” (daSilva et al. 2018, Supplementary Table 8)

Therefore, it seems that KLK3e turns up among RNAs that precipitate with AR when checked with qPCR, but the RNA level is not significant according to the thresholds used in the RNA-sequencing analysis.

In our project it was initially intended to use AR as a control during the RNA pulldown; Western Blotting would be performed to check for AR binding specifically to the KLK3e transcript. However, this approach was abandoned as it relied on replicating one publication, and it was decided to follow an unbiased approach where the aim would be to see different protein bands pulled down by the different RNA samples in the whole gel staining after the RNA pulldown. The gel lanes would then be subjected to mass spectrometry, and differential protein binding to the different RNA samples would be identified.

4.1.2 Sequence confirmation of amplified Xist transcript

The Xist transcript was also confirmed to have the expected sequence. As can be seen in the BLAST matches provided in Table 4-2 below, the sequence matched the database records for Xist lncRNA. The sequence read quality was low towards the end of the transcript (Appendix Figure A.1) which is common in Sanger sequencing due to loss of processivity at the ends of DNA strands. This may account for the shorter match of the transcript to the database, which is about 1037 bases long, while the Xist transcript obtained is actually around 1800 bases long, as can be seen in the PCR gel image previously shown in Figure 4.2. The sequence alignment of Xist transcript to the database sequence is provided in the Appendix section A.

Table 4-2: Sequences producing significant matches to the Xist transcript.

Description	Max Score	Total Score	Query Cover	E value	Percent Identity	Accession
Homo sapiens X inactive specific transcript (XIST), long non-coding RNA	1037	1037	40%	0.0	93.80%	NR_001564.2
Homo sapiens X inactive specific transcript (XIST), RefSeqGene on chromosome X	1037	1037	40%	0.0	93.80%	NG_016172.1
Human DNA sequence from clone RP13-216E22 on chromosome Xq13.1-21.1, complete sequence	1037	1037	40%	0.0	93.80%	AL353804.22
Homo sapiens X (inactive)-specific transcript (XIST) complete exon	1018	1018	40%	0.0	93.35%	M97168.1

4.2 *In vitro* Transcription of KLK3e lncRNA and controls

The *in vitro* transcription of KLK3e with biotinylated UTP had to be optimized on several levels, including reaction size, duration, DNA template amount, RNA Polymerase amount, buffer additives such as DTT and spermidine, as well as clean-up methods. The most significant observations made are provided in the figures below. The optimized reaction is described in the Methods section in detail.

4.2.1 DTT addition significantly improves *in-vitro* transcription

The first important observation was that addition of DTT significantly improved the *in-vitro* transcription reaction. In fact, without the addition of DTT, only a smear was being observed in the RNA gels, as can be seen in Figure 4.2. Even though the provided 5X T7 Transcription Buffer contained DTT at a final concentration of 10mM (Jena Bioscience, Germany), addition of DTT at an additional final concentration of 5mM significantly improved transcription. 1,4-Dithiothreitol (DTT) is a reducing agent that helps stabilize RNA and enzymes such as RNA polymerase.

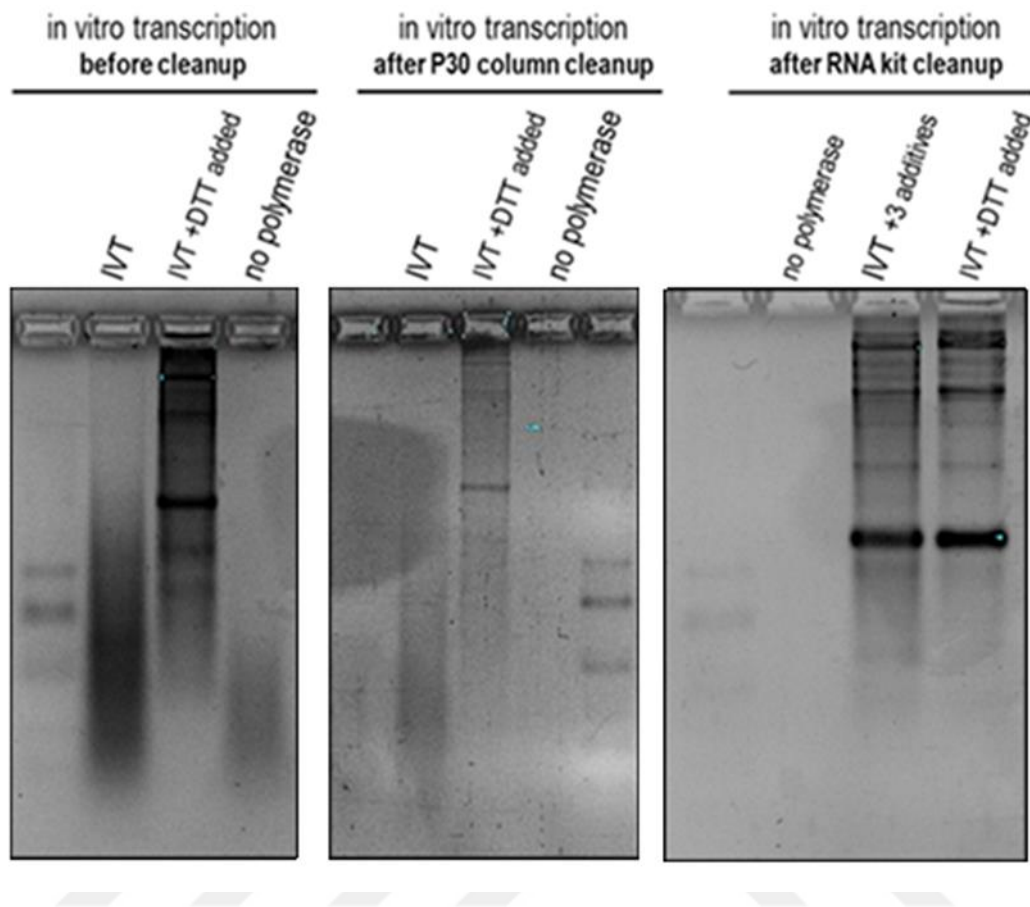


Figure 4.3: Optimization of in Vitro Transcription additives and clean-up. In vitro transcription of RNA was observed to work much more efficiently with the supplementation of DTT, producing a clear single band. The higher bands in the RNA gel was identified not to be caused by DNA remnants, by testing with a no RNA Polymerase control. The higher bands were probably due to RNA that was not denatured enough by boiling. P30 column cleanup appeared to cause loss of RNA yield, and RNA kit cleanup seemed to give better results.

4.2.2 Optimizing the amount of RNA Polymerase

The next step was to optimize the amount of RNA polymerase to be used. Too much could cause non-specific reactions, while too little would not produce enough RNA. As seen in Figure 4.4 below, the optimal range appears to be from 150 up to 400 units of T7 RNA Polymerase for a 40ul reaction. If the T7 RNA Polymerase is not stored in a stable -20°C or is subjected to multiple thawing and freezing cycles, higher side of this range may be required.

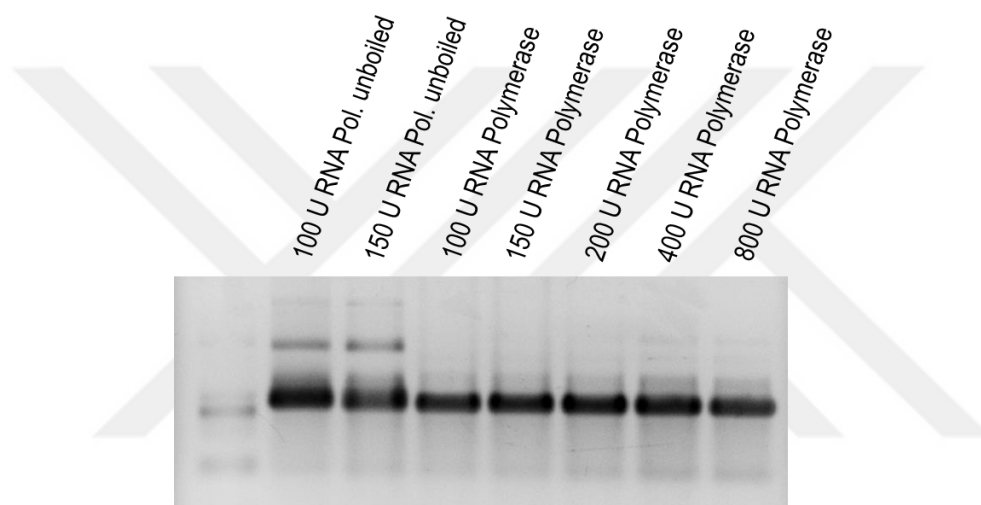


Figure 4.4: Optimization of the amount of RNA Polymerase. The range of 150 up to 400U RNA polymerase seemed optimal. Also, boiling was necessary for the denaturation of RNA in order to obtain a single band.

4.2.3 Confirming specific binding of biotinylated RNA to streptavidin beads

The next step was to ensure that the *in vitro* transcribed RNA was properly biotinylated with the use of Bio-16-UTP incorporation, and that the biotinylated RNA could be specifically bound on magnetic Streptavidin beads. As can be seen in Figure 4.5, the biotinylated lncRNA were binding on the beads as seen in the ‘elute’ sections, and the elutes did not contain any unbiotinylated RNAs, while the supernatants did.

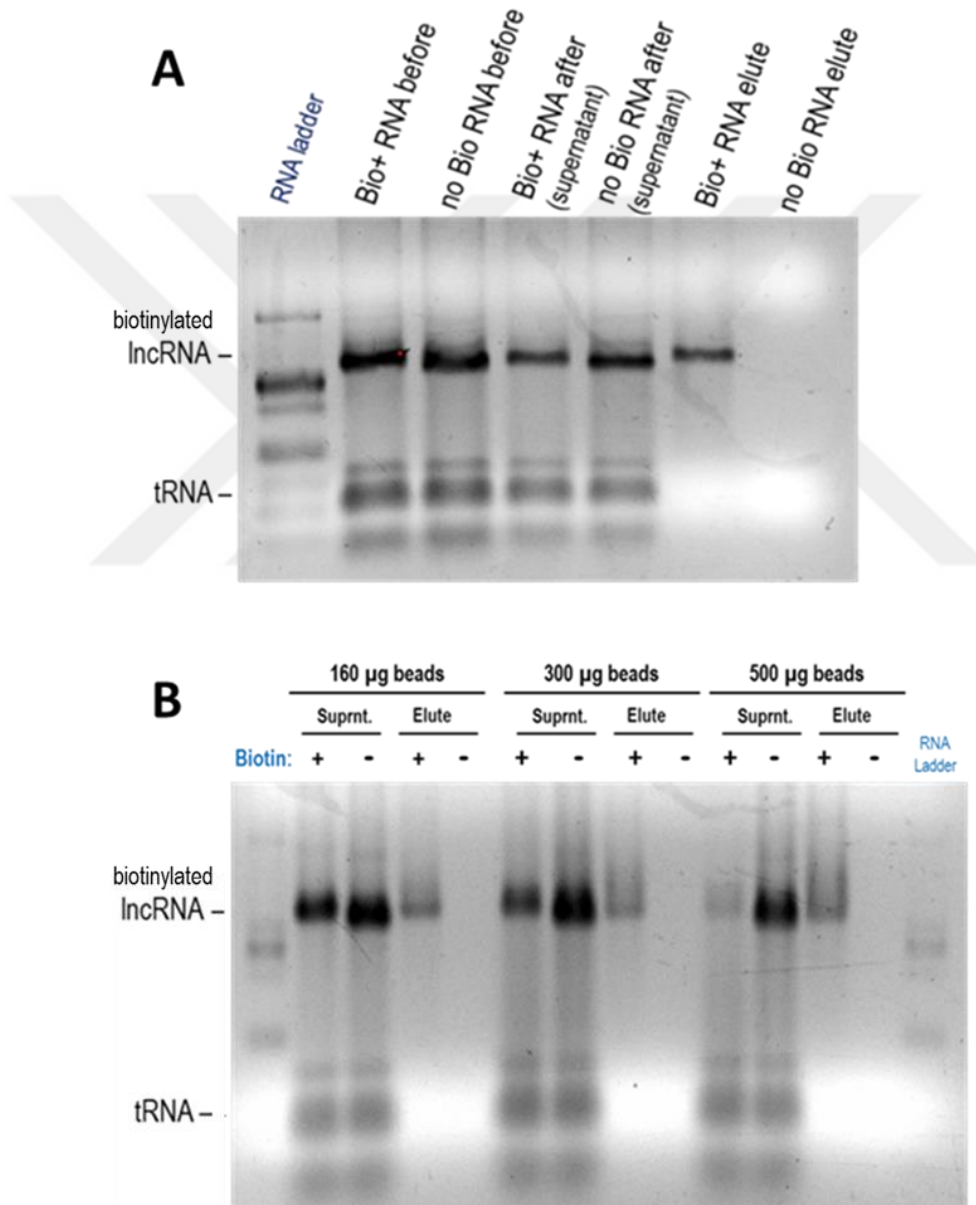


Figure 4.5: Confirmation of specific binding of biotinylated lncRNA on Streptavidin magnetic beads. Biotinylated lncRNA were bound to the beads as seen in the ‘elute’ sections. The elutes did not contain unbiotinylated RNA, while the supernatants did.

4.2.4 RNA gel showed specific *in vitro* transcription of KLK3e and control RNAs

The RNA *in vitro* transcribed from the PCR product clean-ups shown in Figure 4.2 were run on an RNA TBE gel after RNA clean-up. As shown in Figure 4.6 below, the *in vitro* transcription produced specific RNA bands of similar sizes for KLK3e, Xist and CherryGFP.

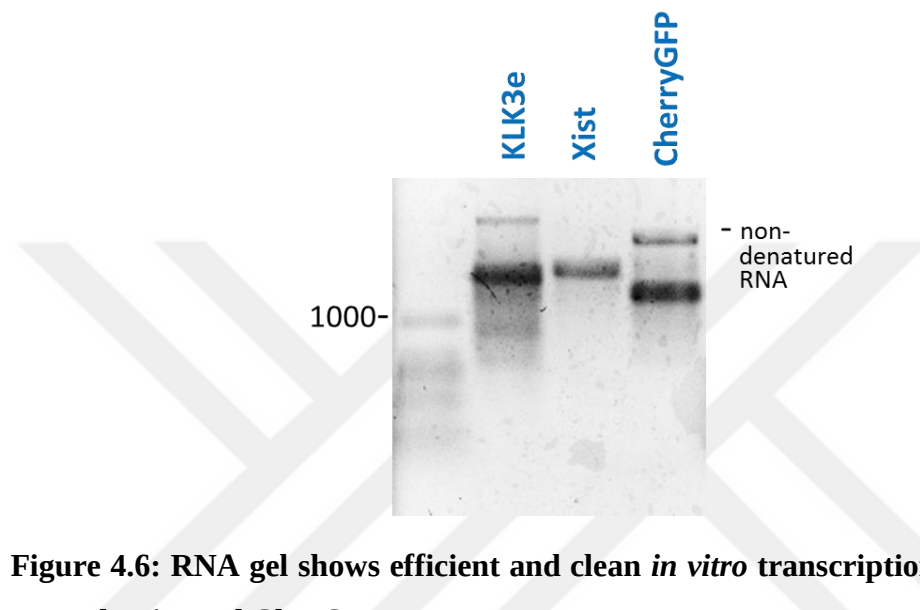


Figure 4.6: RNA gel shows efficient and clean *in vitro* transcription of KLK3e and controls Xist and CherGFP.

4.3 Optimization of RNA pulldown from cell lysates

Many pulldown methods were tried and proved unsuccessful. Combining RNA and protein first, and then adding the beads as described in several protocols failed to produce specific protein binding to RNA. Pierce and NEB Streptavidin beads had high background binding compared to M-280 Dynabeads (Invitrogen, USA). Separation of cytoplasmic fraction with a kit, followed by nuclear extraction with a home-made nuclear isolation buffer worked best. Sonication produced too much histone binding, Dounce homogenizer worked best. High salt concentrations such as 500mM NaCl or 300mM KCl during lysis and/or washes caused very little protein left. For long transcripts, a Bio-UTP to UTP ratio of 1:10 worked best. Additives such as BSA or tRNA to prevent non-specific protein or RNA binding did not work well. Western blotting yielded inconclusive results due to low levels of protein. The approach that worked best is described in the Methods section and results are shown below.

4.4 Optimized RNA pulldown from lysate showed protein bands specific for Xist and KLK3e

In vitro transcribed RNA with biotin tags were bound on Dynabeads M-280 Streptavidin magnetic beads, which were then mixed with nuclear LNCaP cell lysates to bind proteins interacting with the RNA bound on the beads. Use of a biotin-16-UTP to UTP ratio of 1:10 during *in vitro* transcription worked well in the pulldown, as suggested for long RNAs in one protocol paper (Panda et al. 2016). As shown in Figure 4.7 below, there were prominent bands unique to Xist and KLK3e. The bands marked in red were cut and submitted for mass spectrometry identification of the proteins. For validation of results, a replica of the pulldown was performed and each lane was submitted in whole.

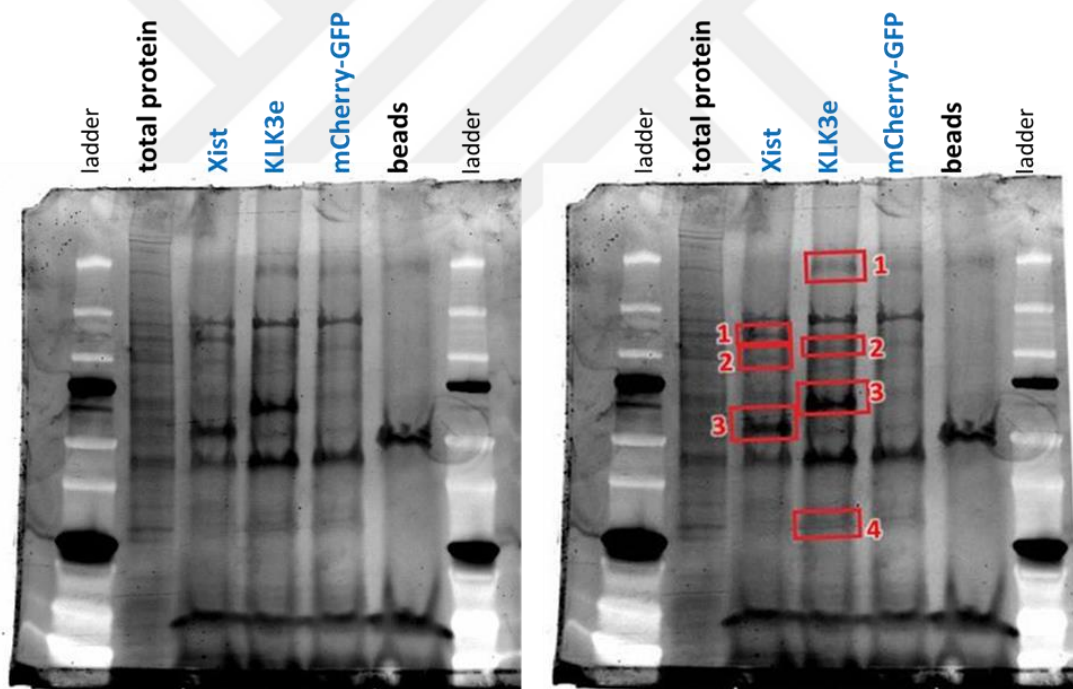


Figure 4.7: Optimized RNA pulldown showed protein bands specific to KLK3e and Xist. Bands marked red were cut and submitted for mass spectrometry identification of proteins.

4.5 *Mass Spectrometry shows prostate cancer and AR-related proteins bound to KLK3e lncRNA*

Mass spectrometry analysis was performed as described in the Methods section 3.1.7 in detail. The proteins shown in Table 4.3 on the next page were bound to KLK3e in two biological replicates, compared to Xist and CherGFP controls. Proteins shown in bold have Σ Coverage values larger than 30. Protein information was collected using UniProt (Bateman 2019) and Ensembl BioMart tool (Smedley et al. 2015).

As shown in Appendix Table A.2, hnRNP-K, hnRNP-U and Matrin-3 which are known interactors of Xist as described in the Literature Review, were among proteins highly detected in Xist pulldown, allowing confidence in the mass spectrometry data. PRC complex components such as Suz12 were not observed because female cells with X inactivation were not used; LNCaP lysates were used for Xist as well, for comparability.

Table 4-3: 41 proteins that specifically bind to KLK3e. Relationship of highlighted proteins with prostate cancer will be discussed in detail. Values in Appendix Table A.1.

Gene	UniProt ID	Description
RPS4Y1	P22090	ribosomal protein S4 Y-linked 1
RPL7A	P62424	ribosomal protein L7a
RPL19	P84098	ribosomal protein L19
H1FX	Q92522	H1 histone family member X
PRPF6	O94906	pre-mRNA processing factor 6
SART1	O43290	spliceosome associated factor 1, recruiter of U4/U6.U5 tri-snRNP
HIST1H4A	P62805	histone cluster 1 H4 family member a
HSPA8	P11142	heat shock protein family A (Hsp70) member 8
DDX5	P17844	DEAD-box helicase 5
CHTOP	Q9Y3Y2	chromatin target of PRMT1
ATAD3A	Q9NVI7	ATPase family AAA domain containing 3A
SYNCRIP	O60506	synaptotagmin binding cytoplasmic RNA interacting protein
RPL31	P62899	ribosomal protein L31
ATAD3B	Q5T9A4	ATPase family AAA domain containing 3B
DDX17	Q92841	DEAD-box helicase 17
RPL29	P47914	ribosomal protein L29
G3BP2	Q9UN86	G3BP stress granule assembly factor 2
HNRNPF	P52597	heterogeneous nuclear ribonucleoprotein F
EIF2AK2	P19525	eukaryotic translation initiation factor 2 alpha kinase 2
LIG3	P49916	DNA ligase 3
DDX54	Q8TDD1	DEAD-box helicase 54
HNRNPA0	Q13151	heterogeneous nuclear ribonucleoprotein A0
UTP23	Q9BRU9	UTP23 small subunit processome component
DDX3X	O00571	DEAD-box helicase 3 X-linked
HNRNPA1	P09651	heterogeneous nuclear ribonucleoprotein A1
PSPC1	Q8WXF1	paraspeckle component 1
DDX21	Q9NR30	DEAD-box helicase 21
KIF23	Q02241	Kinesin-like protein KIF23
RORA	P35398	RAR related orphan receptor A
DHX36	Q9H2U1	DEAD-box helicase 36
HNRNPUL2-BSCL2	H3BQZ7	HNRNPUL2-BSCL2 readthrough (NMD candidate)
RPL15	P61313	ribosomal protein L15
PARP1	P09874	poly(ADP-ribose) polymerase 1
FAM111A	Q96PZ2	family with sequence similarity 111 member A
RPS25	P62851	ribosomal protein S25
DGCR8	Q8WYQ5	DGCR8 microprocessor complex subunit
PRPF39	Q86UA1	pre-mRNA processing factor 39
CPSF1	Q10570	cleavage and polyadenylation specific factor 1
ZFR	Q96KR1	zinc finger RNA binding protein
ATM	Q13315	ATM serine/threonine kinase
DYNC1H1	Q14204	dynein cytoplasmic 1 heavy chain 1

DDX5 / p68

Perhaps the most interesting hit in the identified KLK3e interactors is DEAD-Box Helicase 5 (DDX5 or p68), which is an RNA helicase that was shown to be a co-activator of androgen receptor (AR) implicated in prostate cancer. (Clark et al. 2008) As shown in Figure 4.8, DDX5 appears to be a hub protein within the proteins interacting with KLK3e, therefore, it may help mediate interactions with the other proteins in the complex, as well as with AR.

In the 2008 publication by Clark et al., the interaction between AR and DDX5 was first caught in a yeast two hybrid screen, and then validated using immunofluorescence and co-immunoprecipitation in LNCaP cells. Using chromatin immunoprecipitation (ChIP) in LNCaP cells again, they were also able to show that DDX5 is recruited with AR to the promoter region of the key AR-regulated gene KLK3/PSA. RNAi silencing of DDX5 decreased PSA expression, and enhanced AR-mediated repression of CD44 splicing. Activation of DDX5 with phosphorylation enhanced AR-regulated transcription of target genes. Lastly, they demonstrated that prostate cancer tissues have increased expression of DDX5 compared to benign tissue. (Clark et al. 2008) In another publication in 2013, the same group demonstrated that DDX5 interacts with β -Catenin and RNA Polymerase II and contributes to AR regulated transcription in prostate cancer. Silencing of DDX5 caused a reduction of AR and β -Catenin recruitment to target promoters. (Clark et al. 2013)

DDX5 (p68) was also shown to interact with RNA polymerase II (Pol II) in vivo (Rossow and Janknecht 2003). These experimental observations that DDX5 interacts with AR and with RNA Polymerase II raises the possibility that this protein may be responsible for mediating the interaction observed between KLK3e RNA and AR and Pol II in the publication on KLK3e (Hsieh et al. 2014).

DDX5 (p68) and DDX17 (p72) were also observed to associate with histone deacetylase 1 (HDAC1) (Wilson et al. 2004). Interestingly, HDAC1 was shown to be required for androgen receptor function in prostate cancer, and HDAC inhibitors were observed to block AR-mediated transcription of key genes such as TMPRSS2 (Welsbie et al. 2009). This suggests another level at which DDX5 can influence AR function, through HDAC1 as well as the KLK3e RNA.

Nonsense-mediated mRNA decay (NMD)

Also of note, according to information obtained from BioMart (Smedley et al. 2015), 4 of the 41 proteins associated with KLK3e are implicated in nonsense mediated decay; DDX5, HNRNPUL2-BSCL2, PSPC1, and RPS25.

Non-sense mediated mRNA decay (NMD) is an RNA surveillance pathway first identified in human and yeast cells in 1979 (Chang et al. 1979). The function of NMD is to eliminate abnormal RNA with premature stop codons that encode partial polypeptides. These erroneous RNAs are either marked for degradation by ribonucleases, or their translation is repressed. (Baker and Parker 2004)

HSPA8 (Hsp70 member 8)

Heat Shock Protein Family A (Hsp70) member 8 (HSPA8) is an Hsp70 family protein. It is a chaperone with a protein binding domain at the C-terminus and an ATP-binding domain at the N-terminus (Ravagnan et al. 2001; Zhang et al. 2014).

Importantly, the Hsp70 proteins have recently been understood to interact with AR and play a role in prostate cancer, and are now targets for treatment. It was demonstrated that Hsp70 inhibitors not only reduce the levels of full-length AR, but also the truncated variant AR-V7, and consequently increase apoptosis in prostate cancer cell lines. (Kita et al. 2017) This is important because since the variant AR-V7 lacks the LBD, its emergence causes insensitivity to LBD-targeting drugs such as Enzalutamide, causing drug resistant phenotype.

Hence, Hsp70 is now a novel target for the treatment of castration-resistant prostate cancer (Moses et al. 2018). Hsp70 was also shown to bind the N-terminal domain of AR to modulate its function (Dong et al. 2019). Therefore, it seems that Hsp70 not only regulates the level of AR but also its function.

PARP1

Poly (ADP-ribose) polymerase (PARP) was observed to interact with AR and play key roles in DNA damage response and prostate cancer progression. PARP-1 was observed to modulate the association of AR with chromatin. Inhibition of PARP-1 enzymatic activity rendered prostate cancer cells sensitive to androgen deprivation and genotoxicity. Inhibition of PARP-1 in primary human prostate tumor cultures demonstrated antitumorigenic effects and reduction of AR activity. (Schiewer et al. 2012)

PARP-1 was also demonstrated to cause epithelial to mesenchymal transition (EMT) *in vivo* in mouse models (Pu et al. 2014). This implies PARP-1 can play a role in prostate cancer metastasis.

ATAD3A

ATPase family AAA domain containing 3A (ATAD3A) was observed to reduce PSA levels and cisplatin resistance *in vitro*. It was demonstrated to be an anti-apoptotic factor and a PSA secretion regulator in prostate cancer. (Huang et al. 2011)

RPS4Y1

Ribosomal Protein S4 Y-Linked 1 (RPS4Y1), as the name suggests, is encoded from the Y chromosome. This is the protein that is most highly specific to KLK3e compared to controls, as shown in Appendix Table A.1.

There is few research available on this protein, however, the data available in the Human Protein Atlas (Uhlen et al. 2010; Uhlén et al. 2015) show that its expression is high in prostate tissue, and in prostate cancer, as shown in Figure 4.9. Therefore, this protein is likely to be AR-regulated or AR-related.

It is quite interesting for a ribosomal protein to be encoded from the Y chromosome, and to be so sex-specific and tissue specific. And it is further intriguing that this male-specific ribosomal proteins seems to interact with the KLK3e RNA with very high specificity.

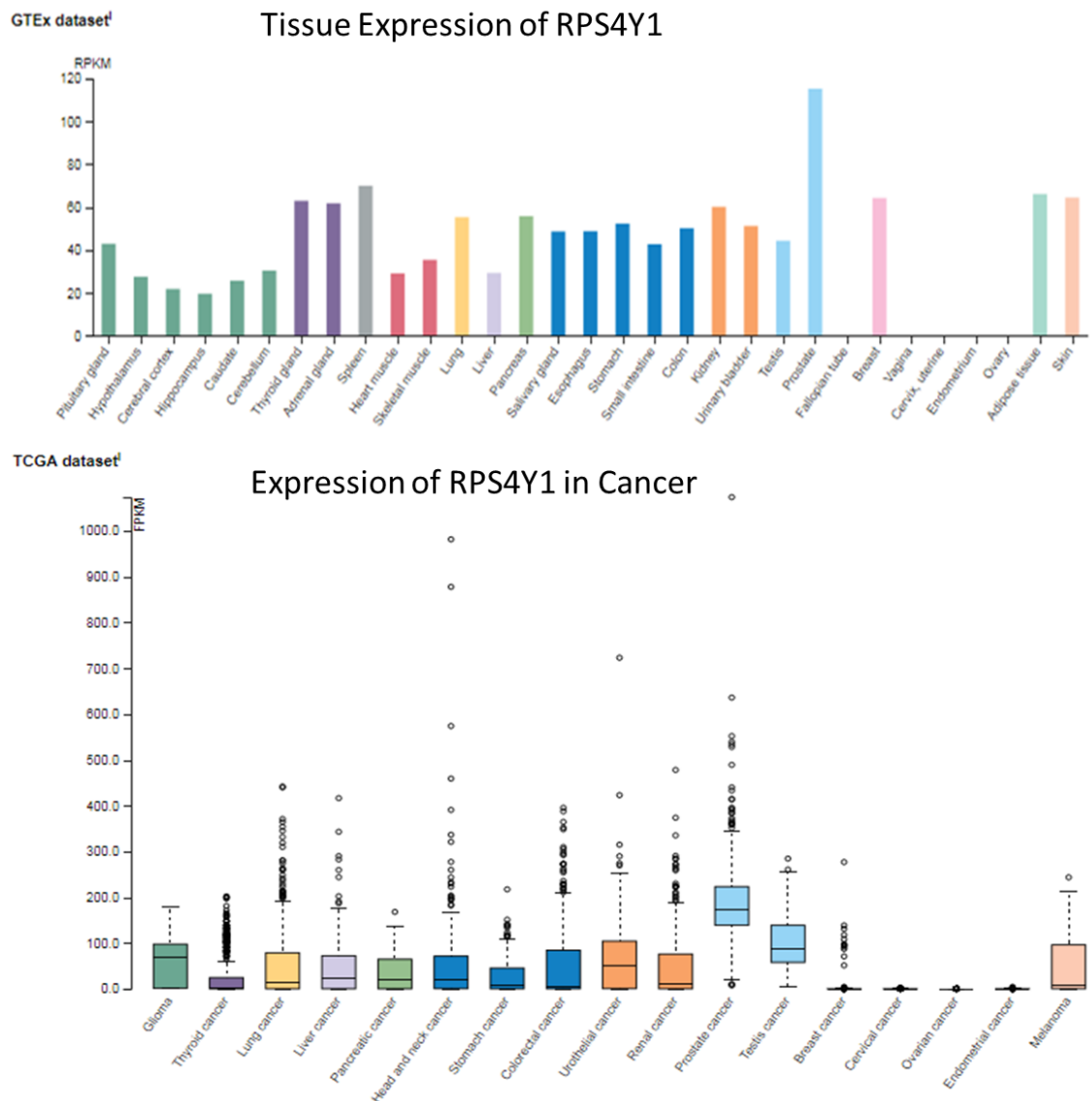


Figure 4.9: RPS4Y1 expression is high in prostate tissue and prostate cancer. Figures obtained from Human Protein Atlas (Uhlen et al. 2010; Uhlén et al. 2015), sourced from GTEx project (Carithers et al. 2015) and TCGA (Network et al. 2013). The results shown here are in whole or part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>

4.6 RNA- protein complex model and proposed mechanism of action

The next aim was to model the RNA-protein complex of KLK3e RNA with its interactor proteins. However, the availability of crystal structures, and the knowledge of interaction domains were limiting factors. Among the interactors identified to be of interest, DDX5 was the best candidate for modeling as crystal structures were available, and it is likely to interact with the crystallized ligand binding domain of AR since the LDB of VDR is known to be responsible for binding DDX5 (Wagner et al. 2012).

On the other hand, HSPA8, and its regulator BAG1L have been shown to interact with AR at the AF1 site at the N terminal domain to enhance AR function (Froesch et al. 1998; Luy et al. 2017). However, since the N-terminal domain of AR is not crystallized and not well conserved, the interaction of HSPA8 with AR could not be modeled and incorporated into the RNA-protein complex. Similarly, the interaction of PARP1 with AR is not well established in terms of the sites involved. There are no crystal structures available for ATAD3A. And in the case of RPS4Y1, since this is a ribosomal protein, it is probably in an alternative complex at the post-transcriptional stage.

In order to obtain a model of KLK3e, DDX5 and AR, first, a 3D model of KLK3e RNA was obtained using the template-based ModeRNA tool (Rother et al. 2011) which was able to model a large RNA like KLK3e. Next, AR (PDB ID:5JJM) and DDX5 (PDB ID:3FE2) were docked using PRISM (Tuncbag et al. 2011; Baspinar et al. 2014); the top candidate that allowed room for the DNA binding domain of AR was selected; it had a binding energy score of -12.75. Finally, KLK3e RNA and the AR-DDX5 complex were docked using PatchDock (Duhovny et al. 2002; Schneidman-Duhovny et al. 2005) using the default parameters, and the used model had the top score of 20268. The Methods Section 3.1.8 includes details on how the tools were compared to other available methods and selected, how each tool works, and how they were used.

The modeled RNA-protein complex is shown in Figure 4.10. DDX5 acts as an anchor to mediate the interaction between KLK3e and AR, as shown in the rotated figure. The folding of the KLK3e RNA into a tertiary structure forms a rather large, sponge-like complex with several cavities into which various proteins can bind and act as co-activators of AR mediated transcription. Since the KLK3e RNA is already transcribed from the enhancer region upstream of the KLK3/PSA gene loci, this RNA may be acting as a scaffold to bring AR co-activators to this transcription site, as it is being transcribed itself.

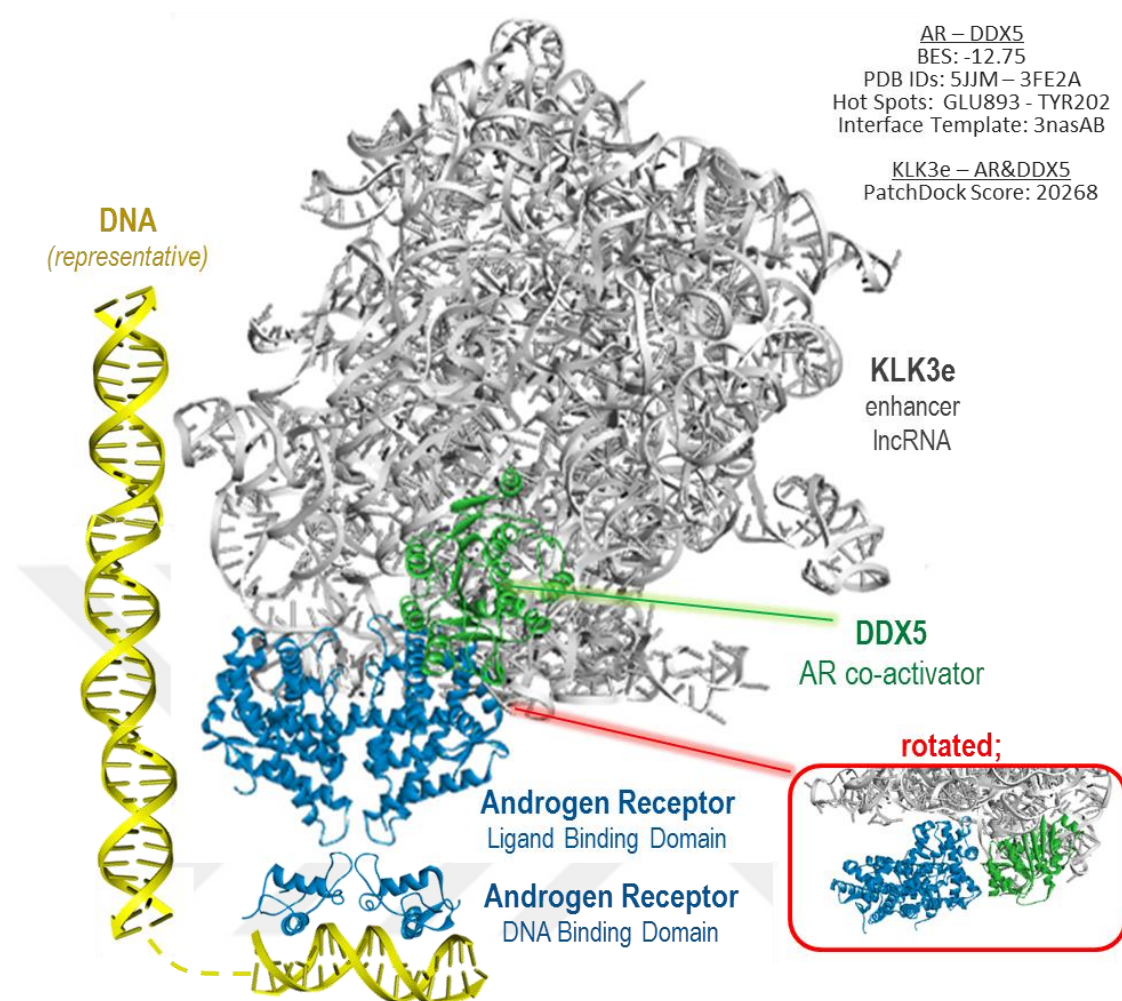


Figure 4.10: Proposed RNA-protein model; the KLK3e RNA being transcribed from the enhancer/promoter region upstream of the KLK3/PSA gene can act as a scaffold to bring co-activators of AR near this transcription site. 3D model of KLK3e RNA was obtained using ModeRNA (Rother et al. 2011) based on the top aligned template ribosomal RNA 3CC2-0. AR (PDB ID:5JJM) and DDX5 (PDB ID:3FE2) were docked using PRISM (Tuncbag et al. 2011; Baspinar et al. 2014) and had a binding energy score of -12.75. The interacting hot spots are GLU893 on AR and TYR202 on DDX5, according to HotRegion (Cukuroglu et al. 2011). KLK3e RNA and the AR-DDX5 complex were docked using PatchDock (Duhovny et al. 2002; Schneidman-Duhovny et al. 2005) using the default parameters; clustering RMSD 4.0 and default complex type. The PatchDock model used had the top score: 20268. AR DNA binding domain dimer PDB ID: 1R4I. DNA image is representative. Details of the modeling procedure are described in the Methods Section 3.1.8. The rotated structure shows that even though AR also has some geometric complementarity, the main anchor mediating the interaction between AR and KLK3e is the DDX5 protein.

4.7 Discussion

The important relationship between lncRNAs, transcription factors and co-activators is, of course, not limited to KLK3e, AR and DDX5.

In fact, a long non-coding RNA called steroid receptor RNA activator (SRA) has been shown to act as a coactivator of the androgen receptor (AR), glucocorticoid receptor (GR), estrogen receptor (ESR), and progesterone receptor (PR) (Liu et al. 2016). Another lncRNA called growth arrest-specific transcript 5 (GAS5) has been shown to interact with the glucocorticoid receptor and suppress its transcriptional activity. (Lucafo et al. 2015) Research has also identified that during lung development, lncRNAs are spatially associated with transcription factors (Herriges et al. 2014).

DDX5 has been demonstrated to be a co-activator of several steroid hormone receptors, interacting with estrogen receptor and vitamin D receptor (Wagner et al. 2012). In fact, DDX5 and DDX17 together, which have both been identified to interact with KLK3e in our analysis, are now considered keystones in steroid hormone regulation of cellular responses (Samaan et al. 2014). The DEAD-box RNA helicases p68 (DDX5) and p72 (DDX17) have been implicated in improved survival in ER α positive breast cancer (Wortham et al. 2009)

The KLK3e lncRNA appears to interact with several AR associated, prostate cancer related proteins, such as DDX5, HSPA8, ATAD3A, and PARP1.

The publication on KLK3e (Hsieh et al. 2014) indicated that this enhancer RNA interacts with AR, but perhaps this interaction was indirect, through another protein such as DDX5, as AR does not seem to have a direct affinity for KLK3e in our pulldown experiment followed by mass spectrometry.

Several DDX5 copies may also mediate other interactions in the complexes that is formed, as it appears to be a hub protein in the protein interaction network obtained from STRING database (Jensen et al. 2009), as shown in Figure 4.8.

GO and pathway enrichment analyses of the 41 KLK3e interacting proteins obtained from STRING database (Jensen et al. 2009) are shown in Appendix Table A.3. These results indicate that KLK3e may be involved in RNA processing. 6 of the identified

proteins are DEAD/DEAH box containing ATP dependent RNA helicases. 6 are ribosomal, 5 are associated with the spliceosome, and interestingly, 2 are involved in base excision repair.

The KLK3e RNA may even be involved in a few alternative complexes with different regulatory roles; perhaps one interacting with transcription factor co-activators near DNA, and one interacting with mRNA and ribosomal proteins.

4.8 *Future Perspectives*

The findings described here will be further validated and explored in the laboratory of Dr. Nathan Lack at the Vancouver Prostate Centre, by Flora Huang.

RNA immunoprecipitation (RIP) assay (Rinn et al. 2007) will be performed to immunoprecipitate DDX5, and the precipitated RNA will be reverse transcribed to confirm interaction with KLK3e by RT-qPCR.

The specificity of the interaction between KLK3e RNA and the candidate proteins will be analyzed using Electrophoretic Mobility Shift Assay (EMSA) with competitive non-specific RNA (Palusa and Reddy 2013). Also, deletion of the suspected RNA region responsible for an interaction, followed by EMSA can help determine the interaction region.

Immunoprecipitations (IPs) of DDX5 and AR will be performed without a cross-linker, in the presence of either RNase or RNase inhibitor, followed by Western Blotting to look for the presence of the other protein (AR or DDX5) to validate if RNA play a role in their interaction. If their interaction is indeed mediated by RNA, then RNase inhibitor should preserve their interaction, while RNase treatment will abolish it.

RNAi silencing of KLK3e, DDX5, HSPA8, PRS4Y1, and AR will be performed in LNCaP cells, followed by RT-qPCR of AR target genes, to compare their effect on the expression of AR target genes.

Finally, the possible involvement of KLK3e in ribosome maturation will be investigated. Ribosomes will be isolated and interaction of KLK3e with ribosomal

proteins as implicated by the results will be investigated using RT-qPCR of KLK3e. Treatment of ribosomal complexes with proteinase K will remove protein but leave RNA intact, allowing their identification.



Chapter 5:

**CHARACTERIZATION OF NOVEL AR ANTAGONISTS AND
THEIR ROLE IN TRANSCRIPTION**

The aim of this project was to study the effects of four novel small molecule inhibitors that target the Androgen Receptor at different sites. The effects of these drugs on the interactions of AR with other proteins such as co-activators were studied by RIME analysis, and on AR-mediated gene expression were studied with RNA-sequencing analysis.

The novel molecules were analyzed in parallel with the FDA-approved molecule Enzalutamide. These small molecules could then be further assessed for second-line therapy of prostate cancer after development of resistance to androgen depletion or other drugs.

Treatment of the cells and RNA preparation were performed at the Vancouver Prostate Centre in the laboratory of Dr. Nathan Lack, by Flora Huang. LNCaP cells were treated with DHT alongside either Enzalutamide which targets the ligand binding site of AR, or one of the small molecules 13789 which targets the BF-3 domain of AR, 17005 or 14449 which target the DNA binding domain, or 220010 which targets the N-terminal domain of AR. RNA was collected in three biological replicas.

For the RNA sequencing analysis, raw RNA sequence reads were mapped to the human genome using STAR Aligner (Dobin et al. 2013), and differential gene expression analysis was conducted using DESeq2 (Love et al. 2014), as described in detail in the Methods section. The sequencing reads were confirmed to have a high ratio of uniquely mapped reads of about 93% (Appendix Figure B.1). High correlation levels of the biological replicates were confirmed with assessment of the Pearson correlation coefficients (Appendix Figures B.2,3,4). Principle Component Analysis was performed to observe that the samples clustered in correlation with the treatments they received (Appendix Figure B.5).

5.1 Data analysis of rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) to detect small molecule inhibited protein interactions of AR

RIME methodology was applied as described in the Literature Review and Supplementary Methods sections. Data analysis aimed to identify proteins that are not present in IgG or EtOH controls, present in DHT, and not present with one or more of the compounds. Observed trends are summarized in Figure 5.1 below.

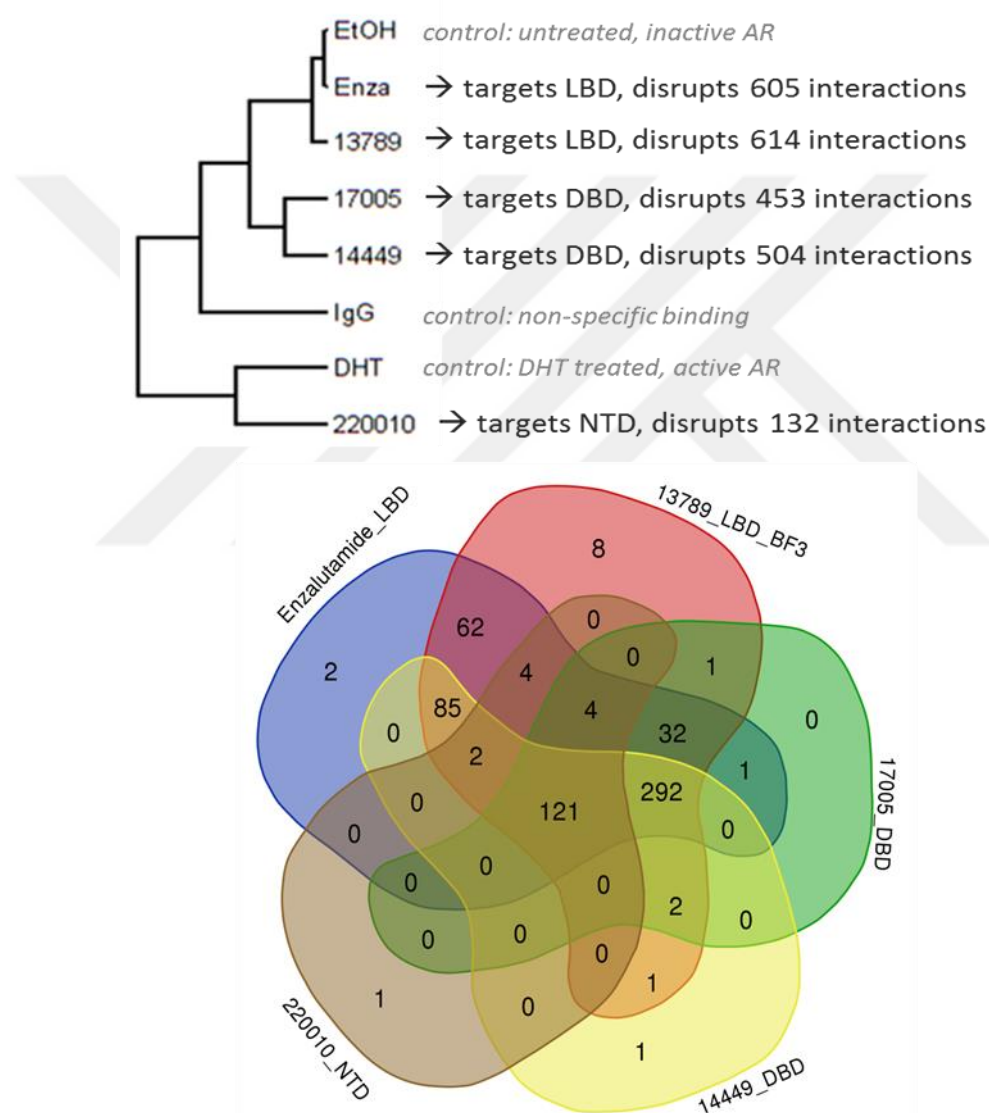


Figure 5.1: Similarities and differences of drug effects on AR interactions. LBD: Ligand Binding Domain. DBD: DNA Binding Domain. NTD: N Terminal Domain. Hierarchical Clustering of RIME mass spectrometry data performed using Perseus (Tyanova and Cox 2018). Venn diagram shows common and unique interactors that seem disrupted among the small molecules; a list representation of this Venn diagram is provided in Appendix Table B.2. Analysis described in detail in the methods section.

As seen in the clustering, Enzalutamide and molecule 13789 treated samples cluster closest to EtOH control, and farthest from DHT and IgG controls. These drugs seem to revert the AR interactome upon DHT treatment back towards the inactive state of AR.

Interestingly, even though these two drugs target different regions on the ligand binding domain, they appear to disrupt a very similar set of proteins, with only a few differences. Enzalutamide blocks the inner ligand binding site, while molecule 13789 binds at the outer BF3 site. The observation that they can impact such similar protein interactors may indicate that inhibition of the BF3 site has a very significant impact, that causes AR to behave like its inactive state without ligand binding.

Importantly, protein interactions that appear to be disrupted by the LBD inhibitors Enzalutamide and 13789 include known AR interactors FKBP5 (Stechschulte and Sanchez 2011) and FOXA1 (Yu et al. 2005), while these interactions are not disrupted by the other 3 drugs that target the DBD or NTD.

The DBD inhibitor molecules 17005 and 14449 cluster together next. They also seem to disrupt a lot of shared targets with each other, as well as with the LBD inhibitors, as shown in the Venn Diagram. Molecule 17005 is a DNA dimerization inhibitor, while 14449 is a DNA binding inhibitor. Interestingly, the DNA binding inhibitor 14449 affects more interactions, and more in common with the LBD inhibitors (86 vs 32) suggesting that DNA binding might be causing allosteric changes in the AR protein that affect its interactions through the LBD as well, which is expected.

The N terminal domain inhibitor 220010 appears to disrupt the least amount of interactions, and correspondingly clusters closer to DHT treated control. This is a bit curious since the upcoming transcriptomic profile after treatment with 20010 also showed effective downregulation of AR targets. 220010 may be affecting another behavior of AR than its protein interactions, such as its affinity for DNA or other interactors (RNA, ions etc.).

5.2 All small molecules tested downregulate key AR-regulated genes

An initial inquiry made during the start of the RNA-seq analysis was to see whether the drugs treatments caused downregulation of key AR regulated genes such as KLK2, KLK3, TMPRSS2 and FKBP5, which are normally upregulated by DHT treatment and AR activation, as described in detail in the Literature Review section. The levels of these genes in the drug + DHT treated samples were compared to the DMSO + DHT treated control sample. It was observed that all of the small molecules downregulated these key genes, generally at levels comparable to, if not better than, the FDA approved drug Enzalutamide, as shown in Figure 5.2 below. Therefore, it seemed that all of these drugs were counteracting the effect of DHT and inhibiting AR induced transcription of AR target genes.

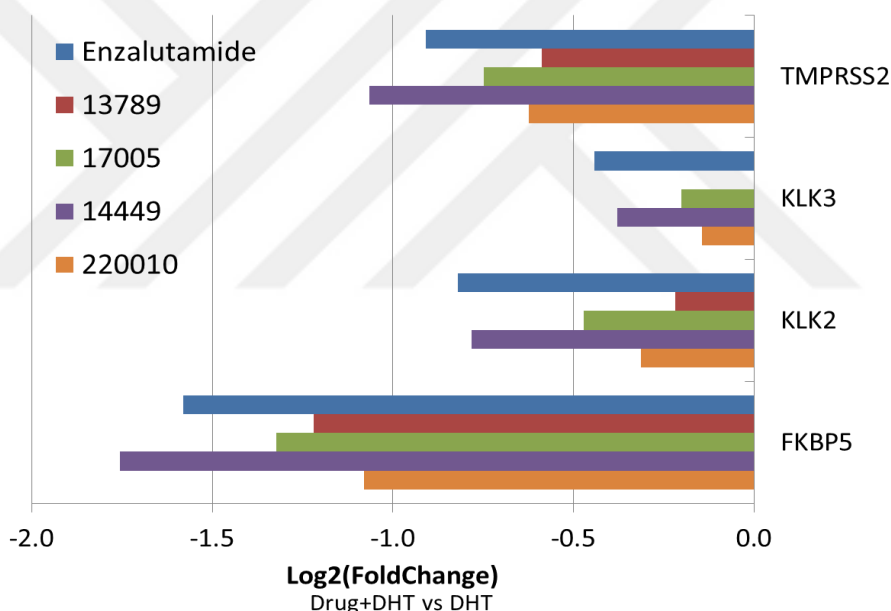


Figure 5.2: Small molecules tested downregulate key AR regulated genes at levels comparable to control drug Enzalutamide.

The next questions asked were how these small molecules affected genes that were significantly upregulated by DHT compared to EtOH control, and how they affected well-known androgen response hallmark genes from the MSigDB (Subramanian et al. 2005) HALLMARK_ANDROGEN_RESPONSE gene set.

5.3 Tested molecules generally counteract the effect of DHT on gene expression

DHT regulated genes were defined as genes with a $\text{padj} < 0.01$ in the DHT vs. EtOH differential gene expression analysis. When they were ranked in the heatmap from most upregulated to most downregulated, it could be observed that the small molecules generally counteract the effect of DHT on the genes, downregulating the DHT upregulated genes and upregulating the DHT downregulated genes.

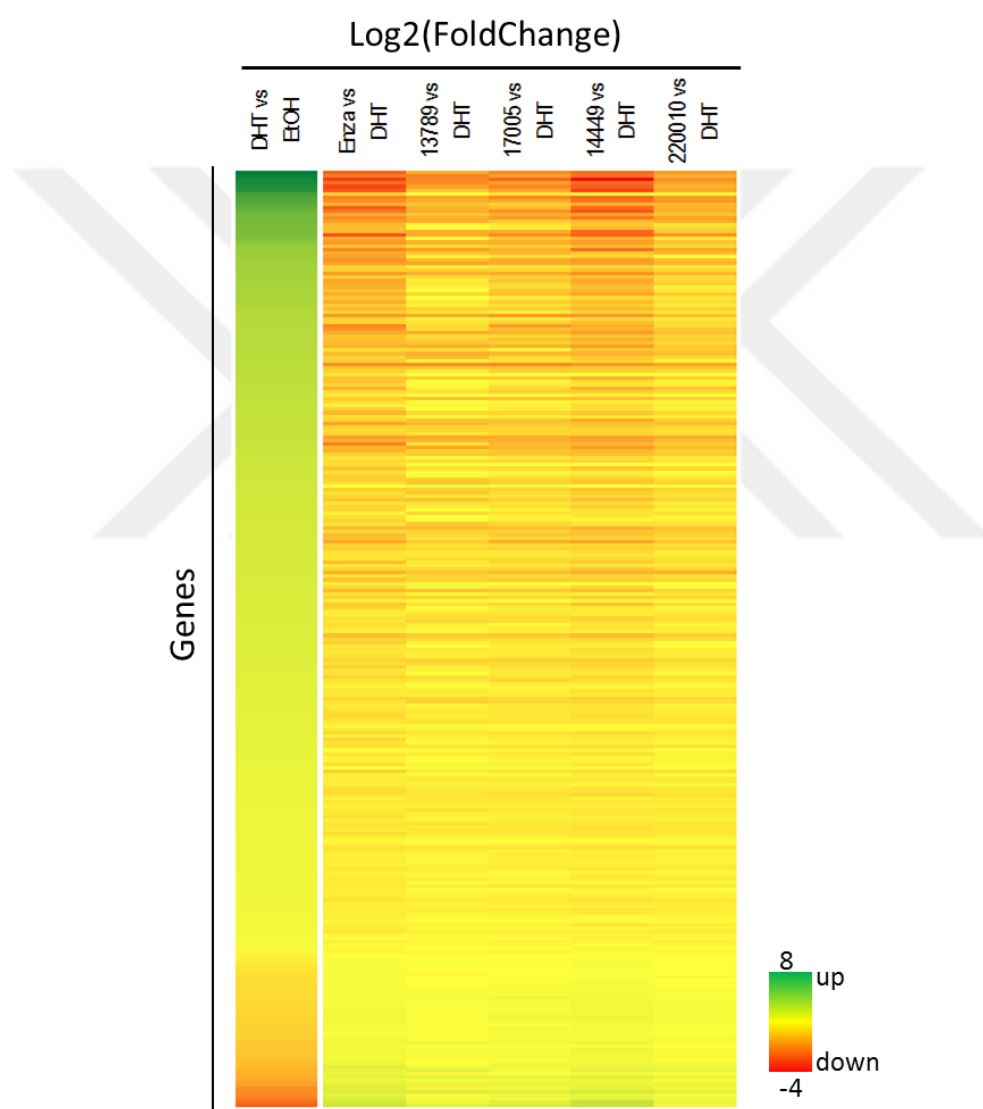


Figure 5.3: Tested small molecules generally counteract the effect of DHT in DHT regulated genes. Green represents upregulation, red represents downregulation.

5.4 Effect of small molecules on AR Hallmark genes

AR Hallmark genes were selected as those listed in the MSigDB (Subramanian et al. 2005) HALLMARK_ANDROGEN_RESPONSE gene set. This gene set is composed of 8 founder gene sets. These founder gene sets include not only genes regulated by androgen, but also genes regulated by forksolin, which is a protein kinase A activator that can activate the androgen receptor independent of androgens (Nazareth and Weigel 1996). This may account for the different trend observed in these genes compared to the DHT regulated genes shown in the previous section.

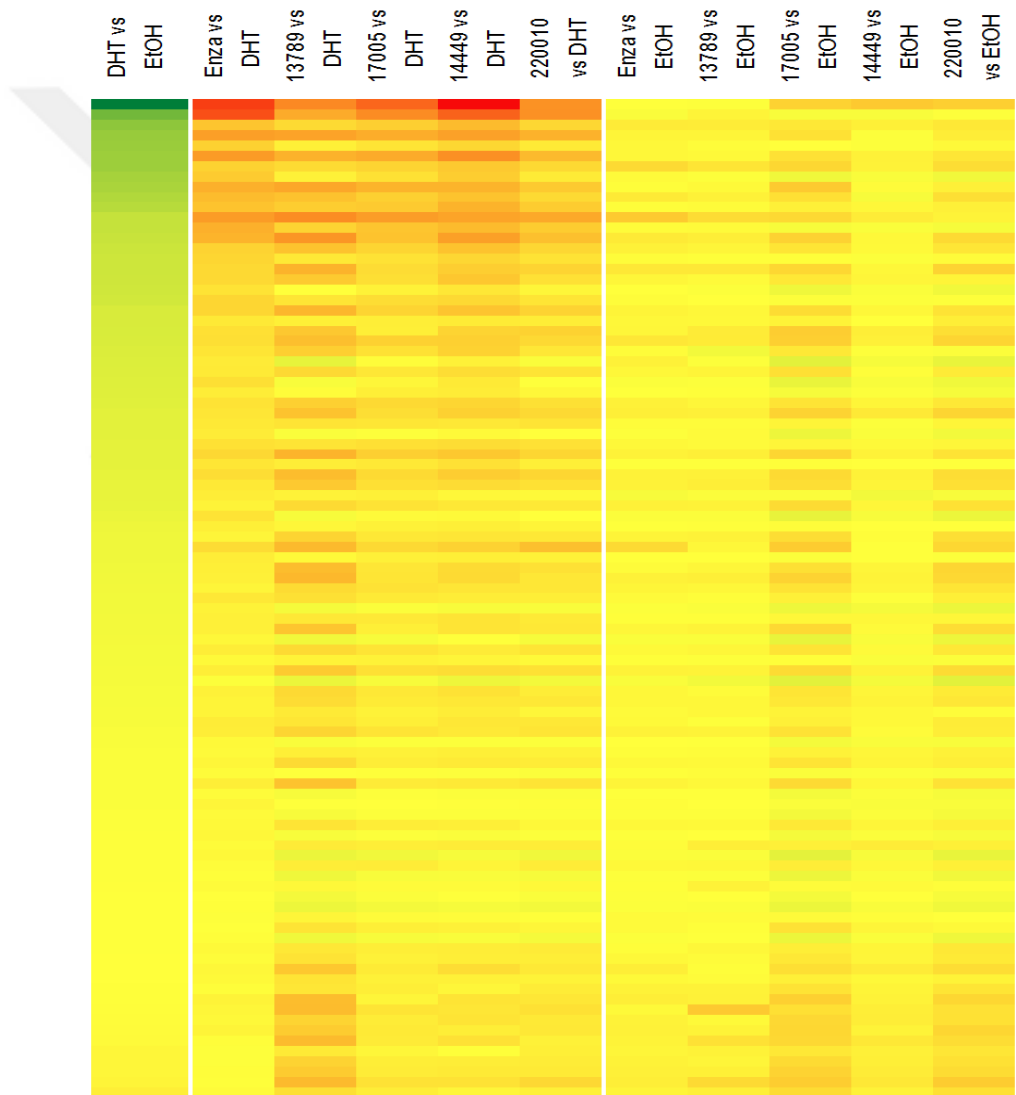


Figure 5.4: Effect of small molecules on AR Hallmark genes. Green represents upregulation, red represents downregulation.

5.5 A closer look at 22 genes common between DHT-upregulated and AR Hallmark genes

22 genes were common between the DHT regulated genes in our analysis and the AR Hallmark gene set. These included STEAP4, SGK1, TMPRSS2, FKBP5, KLK2 and KLK3, which are genes well known to be regulated by androgens, as described in detail in the Literature Review section.

Tested small molecules were also confirmed not to have a significant effect in the absence of DHT in the 22 representative DHT regulated AR Hallmark genes, with the exception of STEAP4, as shown in the figure.

Gene	Log2FC	DHT vs EIOH	Enza vs DHT	13789 vs DHT	17005 vs DHT	14449 vs DHT	220010 vs DHT	Enza vs EIOH	13789 vs EIOH	17005 vs EIOH	14449 vs EIOH	220010 vs EIOH
STEAP4	6.32	6.32	-3.09	-1.79	-2.40	-3.97	-2.01	-1.45	-1.42	-2.19	-2.33	-1.28
SGK1	4.03	4.03	-2.79	-1.33	-1.82	-2.49	-1.74	0.17	-0.17	0.20	0.21	0.07
TMPRSS2	3.23	3.23	-0.91	-0.59	-0.75	-1.07	-0.62	-0.30	-0.30	-0.35	-0.23	-0.35
MAF	2.96	2.96	-1.54	-1.45	-1.29	-1.49	-1.21	-0.13	-0.15	-0.46	0.12	-0.28
PMEPA1	2.88	2.88	-0.72	-0.24	-0.42	-0.63	-0.33	-0.12	-0.05	-0.01	0.09	-0.10
FKBP5	2.80	2.80	-1.58	-1.22	-1.32	-1.76	-1.08	-0.12	-0.09	-0.47	-0.21	-0.38
NKX3-1	2.79	2.79	-0.71	-0.57	-0.67	-0.85	-0.60	-0.58	-0.40	-0.64	-0.20	-0.53
KLK2	2.55	2.55	-0.82	-0.22	-0.47	-0.78	-0.31	-0.08	0.06	0.44	0.18	0.40
MAK	2.22	2.22	-1.06	-0.93	-0.68	-0.94	-0.59	-0.31	-0.24	-0.46	0.23	-0.43
CENPN	2.07	2.07	-0.95	-0.76	-0.84	-1.19	-0.76	-0.05	-0.06	-0.23	-0.12	-0.20
NDRG1	1.67	1.67	-1.25	-0.67	-0.91	-1.07	-0.80	-0.06	-0.08	0.23	0.16	0.26
CAMKK2	1.44	1.44	-0.64	-0.34	-0.45	-0.62	-0.44	-0.04	-0.01	-0.02	0.09	-0.07
KLK3	1.39	1.39	-0.44	0.00	-0.20	-0.38	-0.14	0.00	0.10	0.47	0.16	0.40
ANKH	1.32	1.32	-0.64	-0.41	-0.54	-0.60	-0.40	-0.05	-0.01	0.10	0.11	0.06
ZMIZ1	1.13	1.13	-0.35	-0.24	-0.26	-0.32	-0.27	-0.12	-0.11	-0.20	0.00	-0.27
HOMER2	0.93	0.93	-0.28	-0.04	-0.24	-0.30	-0.12	0.05	-0.01	0.26	0.08	0.25
STK39	0.89	0.89	-0.43	-0.74	-0.59	-0.64	-0.49	-0.21	-0.14	-0.42	-0.15	-0.37
ABHD2	0.82	0.82	-0.34	-0.43	-0.39	-0.35	-0.42	-0.05	-0.05	-0.16	0.09	-0.16
TSC22D1	0.77	0.77	-0.48	-0.43	-0.48	-0.53	-0.47	-0.10	-0.08	-0.16	-0.09	-0.15
RAB4A	0.75	0.75	-0.40	-0.30	-0.34	-0.47	-0.25	0.04	0.04	-0.02	0.00	0.04
MERTK	0.68	0.68	-0.28	-0.20	-0.24	-0.12	-0.08	0.22	0.12	0.12	0.38	0.20
B4GALT1	0.51	0.51	-0.25	-0.13	-0.21	-0.24	-0.16	0.05	0.06	-0.04	0.04	0.00

Figure 5.5: Effect of small molecules on AR Hallmark genes. Values shown are Log2 fold change values of differential analysis between molecule + DHT vs. DHT control. Green represents upregulation, red represents downregulation.

5.6 Gene Set Enrichment Analysis (GSEA) of genes affected by each drug shows greatest effect on androgen response genes

GSEA was performed to understand which cellular processes each drug was affecting, taking into account the significance of the changes, as described in the Methods section.

As expected, the gene set most highly affected by all drugs was the Hallmark Androgen Response gene set. Importantly, the 4 molecules tested seemed to impact the androgen response at higher levels than the FDA approved drug Enzalutamide.

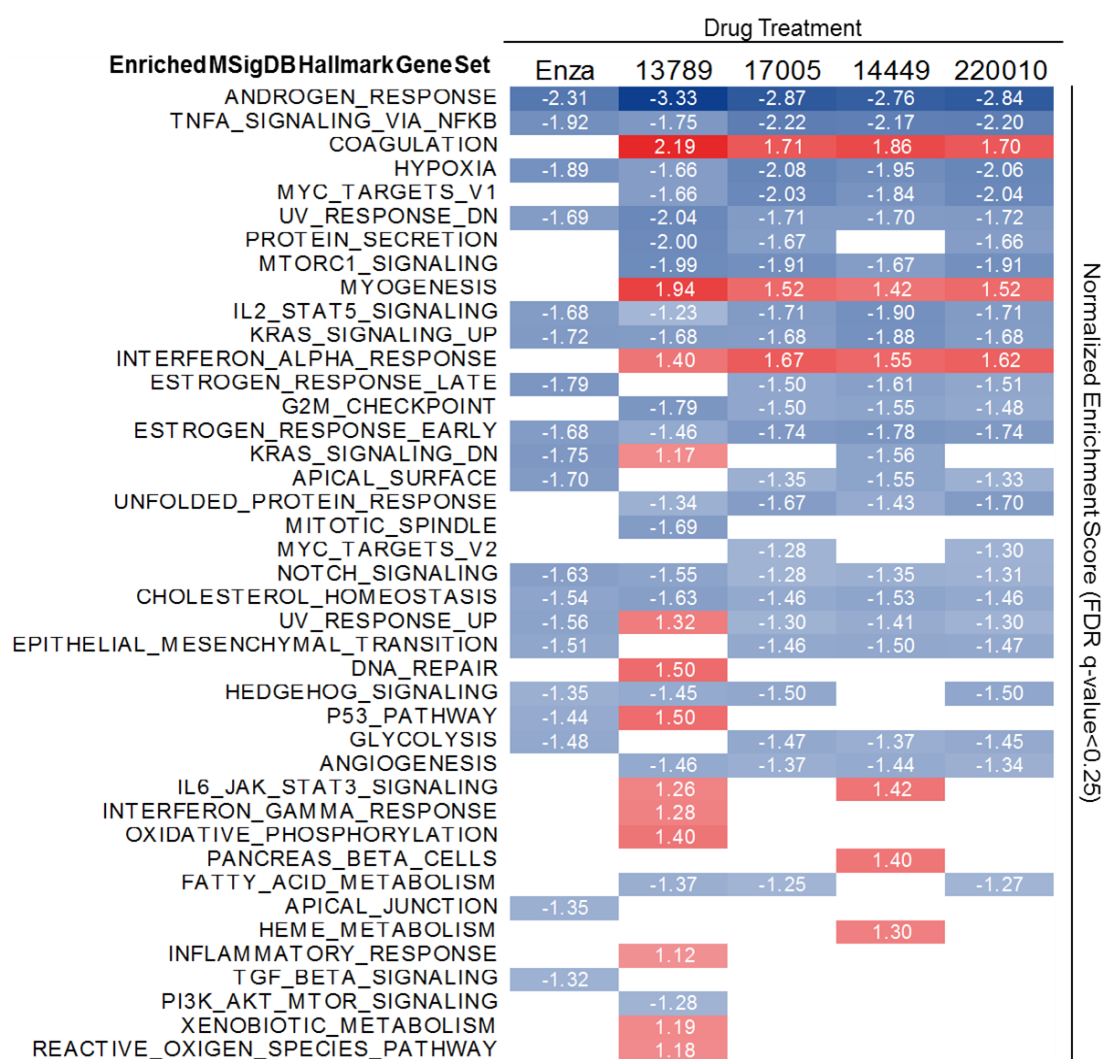


Figure 5.6: Gene Set Enrichment Analysis (GSEA) shows all drugs have the highest effect on androgen response, downregulating genes in this pathway, with normalized enrichment scores even more negative than the FDA approved drug Enzalutamide.

Also of note, Enzalutamide has a negative normalized enrichment score for all pathways or gene sets it affects, while the novel four molecules also positively affect some gene sets, such as coagulation and myogenesis.

The second most negatively affected gene set is the Hallmark TNF α signaling via NF- κ B set. This is particularly interesting because TNF- α was shown to upregulate Prostate-Specific Antigen (PSA) expression through the NF- κ B pathway in LNCaP cells. And NF- κ B was shown to be upregulated in androgen independent prostate cancer as well. (Chen and Sawyers 2002) This might mean that these small molecules can be effective in the treatment of castration resistant prostate cancer, as inhibition of TNF-mediated signaling has been implicated for the treatment of castration resistant prostate cancer (Srinivasan et al. 2010).

It is interesting but not unexpected to see coagulation among the enriched gene sets for the four novel molecules. As described in detail in the Literature Review section, many of androgen regulated genes in the prostate are enzymes with anti-coagulatory roles that help liquefy semen, or proteins that bind ions such as iron or zinc, since the general purpose of the prostate tissue is to secrete components of the semen. The Hallmark Coagulation gene set consists of genes involved in blood coagulation, which include serine peptidases, metalloproteinases, and ion binding proteins. (Subramanian et al. 2005) Some of these genes are also expressed in the prostate and often work in the direction of anti-coagulation and liquification, hence, it is not unexpected to see the coagulation gene set enriched by small molecules targeting prostate cancer.

Seeing hypoxia among negatively affected pathways is also expected, since hypoxia signaling has been observed to promote prostate cancer progression, and hypoxia in prostate tumors is associated with poor prognosis. (Marignol et al. 2008; Deep and Panigrahi 2016)

As MYC activation is considered among the hallmarks of cancer (Gabay et al. 2014), it is also pleasing to see it negatively regulated by the four novel molecules. As UV response involves DNA damage repair, it is also validating to observe that genes downregulated during UV response are also negatively regulated by these molecules.

These molecules also seem to work against K-Ras signaling, which is expected of anti-prostate cancer molecules.

Of note, unlike Enzalutamide, these molecules appear to have a positive effect on myogenesis. This is interesting because it may imply that these molecules can be effective against muscular disease that involve hyperactivity of the androgen receptor, such as Kennedy Disease/Spinal Bulbar Muscular Atrophy (KD/SBMA). (Monks et al. 2008)

It is also interesting that the four novel molecules similarly have a positive effect on the interferon-alpha response, because interferon- α was observed improve clinical outcome of chemotherapy in castration-resistant prostate cancer patients (Li et al. 2014b).

5.7 *Tested small molecules effect common & distinct genes*

The next approach was to take a closer look at the numbers of genes affected differently by each drug, to better understand how unique their effect was. As seen in table 5-1, all molecules cause decreased levels in genes compared to DHT treatment control. Molecule 13789 affects the most number of genes. All drugs have unique effects, but they largely affect genes common with other drugs.

Table 5-1: Numbers and compositions of genes affected by each drug, compared to DHT control. Genes with p-values<0.05 in drug vs DHT control comparison.

Differential Comparison	Total # of DEGs	# of DEGs unique to drug	Increased	Decreased	% Decreased	% non-protein coding
Enza vs DHT	499	140	163	336	67	24
13789 vs DHT	1247	725	425	822	66	16
17005 vs DHT	599	136	174	425	71	27
14449 vs DHT	657	166	222	435	66	22
220010 vs DHT	642	170	227	415	65	31

Looking at a narrower significance threshold, when only genes with p-adjusted values less than 0.1 are taken into account, it is seen that there are 28 common genes that all of the molecules effect significantly. And molecule 14449 has the most unique targets, and also many common with Enzalutamide, as seen in the Venn Diagram below in Figure 5.7.

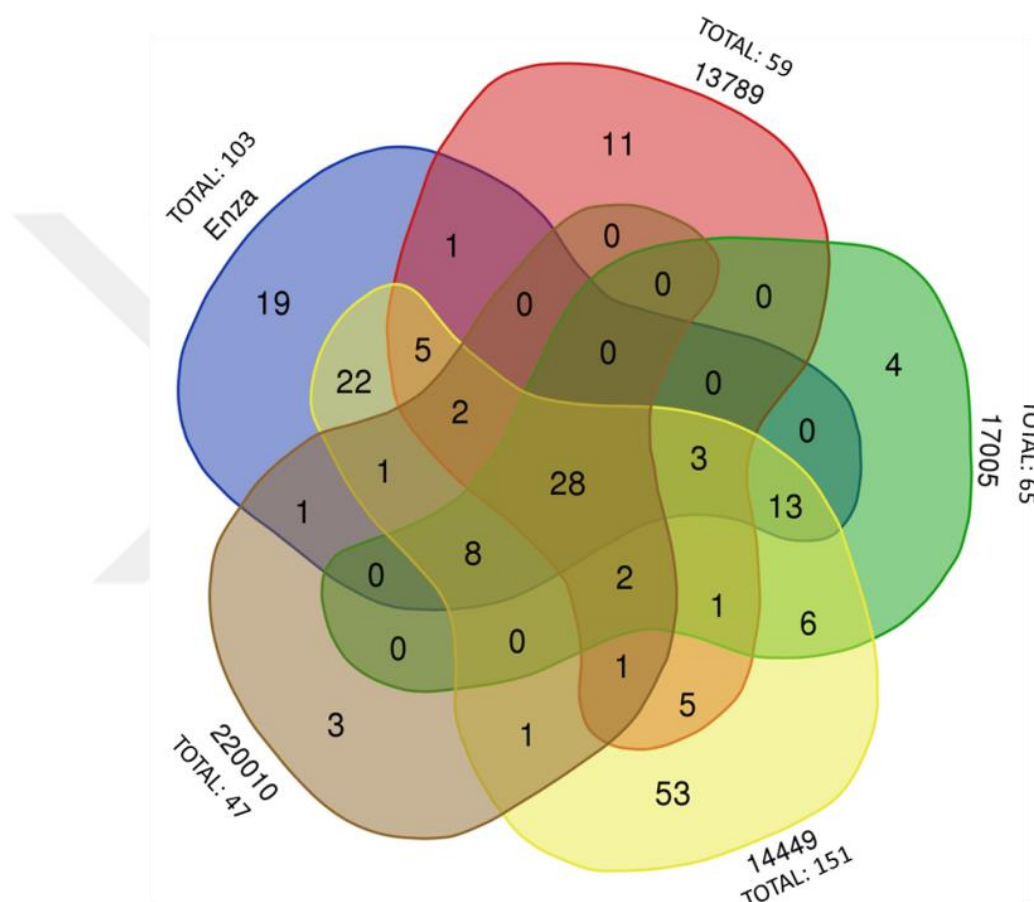


Figure 5.7: Number of genes significantly affected by each drug, compared to DHT control. Genes with p-adjusted values < 0.05 in drug vs DHT control comparison.

Venn diagram prepared using; <http://bioinformatics.psb.ugent.be/webtools/Venn/>

5.8 Discussion & Future Perspectives

One of the most interesting observations in this study was that enzalutamide and molecule 13789 appear to impact a very similar set of AR interactors. This implicates that inhibition of the BF3 site by 13789 has a significant impact that causes AR to behave like its inactive state without ligand binding.

The DBD inhibitor molecules 17005 and 14449 also appear to disrupt a lot of shared targets with each other, as well as with the LBD inhibitors, suggesting that DNA binding and dimerization might be causing allosteric changes in the AR protein that affect its interactions through the LBD as well, which is expected.

While the N terminal domain inhibitor 220010 was appeared to disrupt the least amount of interactions, it still showed effective downregulation of AR targets. Therefore, this molecule may be affecting another behavior of AR than its protein interactions, such as its affinity for DNA or other interactors (RNA, ions etc.).

It is important to note that replicas of the RIME analysis will be performed to enable statistical analysis of significantly disrupted interactions, and the observations shown here are preliminary trends.

The effects of these small molecules on the AR interactome and downstream transcriptome show that their effectiveness is at comparable, if not better, levels to Enzalutamide. Especially the DBD targeting molecules 17005 and 14449 are valuable drug candidates since they can target AR variants that lack the ligand binding domain, such as AR-V7 as well, and overcome the resistant phenotype.

In order to establish these molecules alternatives to Enzalutamide treatment in patients with resistance, in vivo studies would initially need to be conducted in animal models. They would then have to be improved in terms of pharmacokinetics to optimize their metabolization and tolerance. Pre-clinical and clinical assessment of small molecules that prove effective and safe can then be pursued.

Chapter 6:

OLIGOMERIZATION POTENTIALS OF THE GLUCOCORTICOID RECEPTOR

As explained in the Literature Review section, the glucocorticoid receptor is implicated to form oligomers with the mineralocorticoid receptor, and to form tetramers upon DNA binding. This project aimed to investigate how these interactions could occur at the structural level. The protein-protein interactions were structurally modeled using the PRISM: Protein Interactions by Structural Matching tool (Tuncbag et al., 2011, Baspinar et al., 2014).

6.1 *Interaction of Glucocorticoid and Mineralocorticoid Receptors*

One initial observation made in this project by our experimental collaborators was that GR and MR interact in rat hippocampus lysates and mouse cell lysates (Appendix Figure C.1). It was also observed in living cells that GR and MR interact in the nucleus and at DNA (Appendix Figure C.2), using the number and brightness assay described in the Literature Review for GR.

In light of the previous literature, our collaborators suspected that the D-loop interaction between the DNA binding domains (DBDs) of GR and MR, shown in Figure 6.1, may not be their only interaction interface. To investigate this, they mutated key residues in this interface and observed that the interaction still occurred. (data not shown) This strengthened the possibility of alternative interaction interfaces between GR and MR.

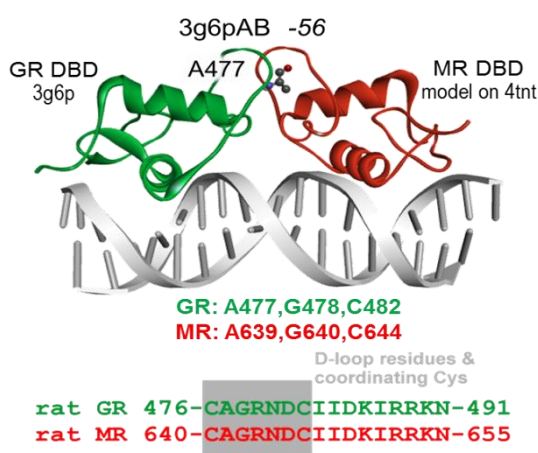


Figure 6.1: Structural prediction for MR and GR interaction through the classical dimerization domain in the D-loop of the DNA binding domain. Linear protein sequence highlighting D-loop amino acids. The PDB ID for the rat GR DBD is 3G6P. Rat MR DBD is a homology model based on human MR DBD (PDB ID 4TNT). The interaction interface template matched by PRISM is 3g6pAB, which is a rat GR DBD homodimer. The energy score predicted by ZRANK is -56. Computationally predicted interaction hot spots on GR and MR are provided below the structure. Mutations of Ala-477 (GR) to Lysine or Threonine are expected to disrupt this interface in computational predictions described in Methods. PRISM (Tuncbag et al., 2011, Baspinar et al., 2014), ZRANK (Pierce and Weng, 2007)

6.1.1 Structural predictions support multiple alternative interfaces for MR-GR interaction at the DNA-binding domain (DBD)

Experimental number and brightness assay results for MR-GR suggested a higher order complex bound at DNA compared to the nucleus. Moreover, GR tetramerization was recently reported to be triggered by DNA binding (Presman et al. 2016). It was therefore intriguing to investigate whether the DBDs could form alternative and higher-order oligomers, at the structural levels, using the PRISM tool (Tuncbag et al., 2011, Baspinar et al., 2014).

Results indicated that DNA looping could allow the formation of a tetramer of two GR-GR and MR-MR homodimers, as well as a tetramer of two GR-MR heterodimers, as shown in Figure 6.2 A and B respectively. Different conformations were possible, as well as a trimer that could be formed at a half-site.

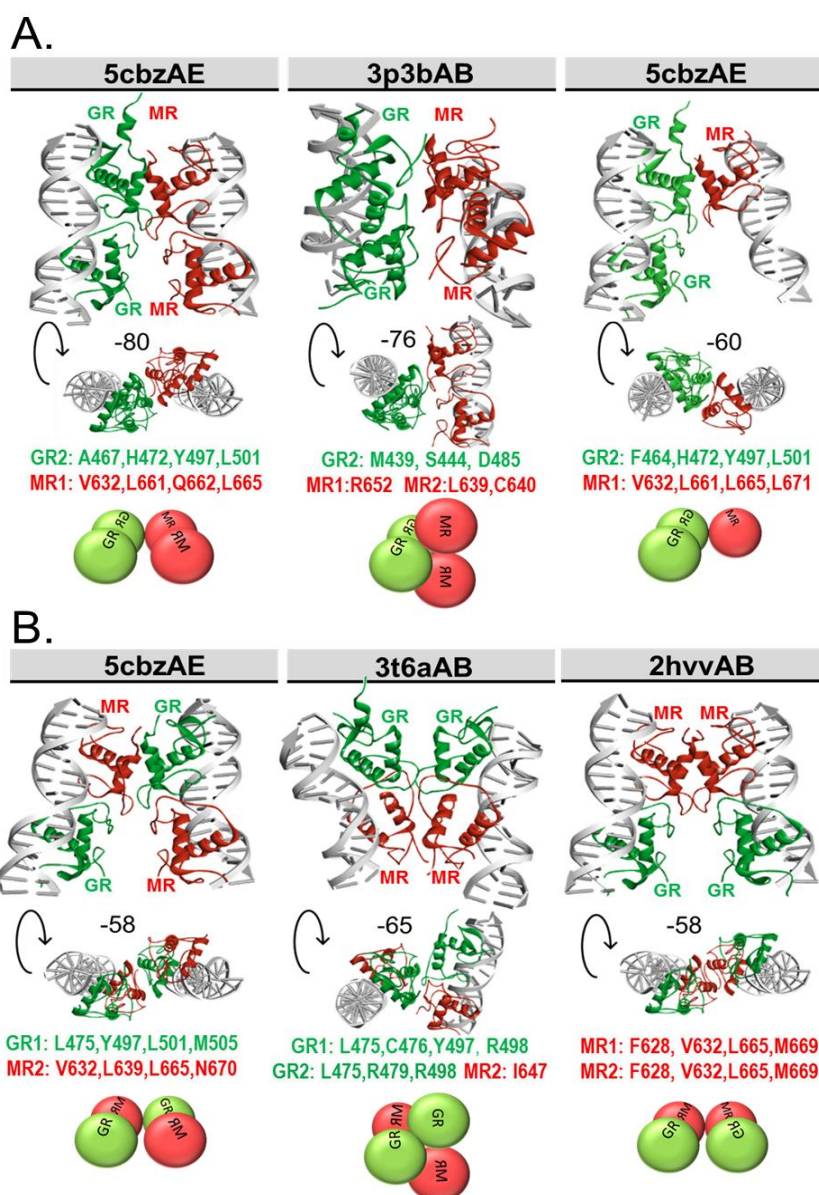


Figure 6.2: GR-MR DNA binding domain interactions structural predictions. The PRISM matched interaction interface template is indicated above each structure. Interacting hot spot residues are provided below each structure. The numbers near each interface indicate binding energy scores predicted by ZRANK. (A) Two possible modes for DBD-mediated interaction of a GR+GR homodimer with an MR+MR homodimer, and one for a GR+GR homodimer with an MR monomer, which may occur at a half site. (B) Three possible modes for DBD-mediated interaction of two true MR-GR heterodimers forming a tetramer at DNA. Bound DNA may loop between GRE binding sites. The DBD structures used from PDB are 3G6P for GR, and rat homology model of 4TNT for MR.

6.1.2 Structural predictions support multiple alternative interfaces for MR-GR interaction at the Ligand Binding Domain (LBD) as well

Research on GR has shown that the LBD is important for its dimerization and tetramerization (Kauppi et al. 2003; Presman et al. 2016). Since MR is highly homologous with GR at the LBD in addition to the DBD, the possible modes of GR-MR LBD interactions were also investigated. As shown in Figure 6.3 below, predictions were dominated by interfaces from GR, MR and AR crystal structures, increasing confidence in these models due to high homology between these steroid receptors.

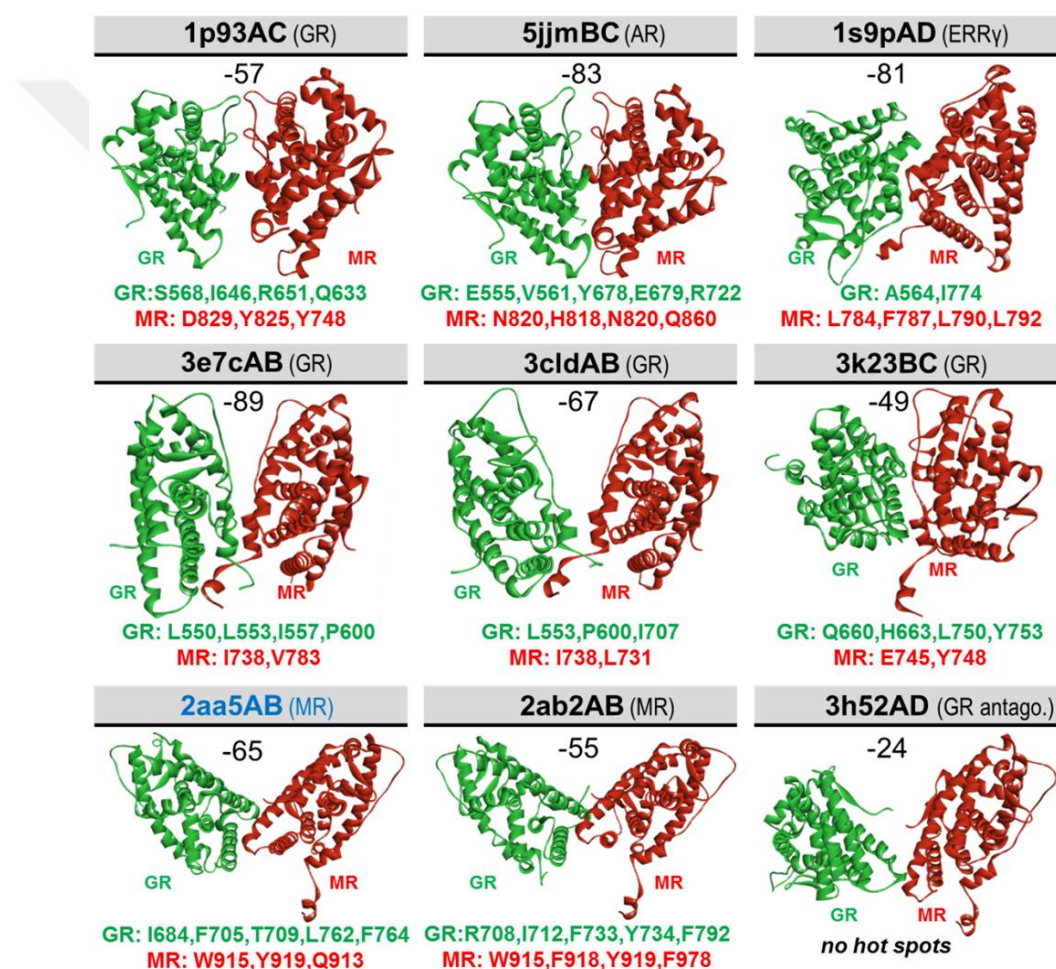


Figure 6.3: As indicated by experiments, GR-MR ligand binding domains can interact through alternative interfaces, which are from the same receptor family. MR and GR LBDs may associate through multiple alternative interfaces from steroid receptor family crystal structures indicated in parentheses. The LBD structures used are rat homology models of PDB structures 4P6X and 2AA2 for GR and MR respectively. Predicted interacting hot spot residues are provided below each structure.

Of note, the agonist bound structures used are not able to form a high affinity complex through the antagonist interface (3h52AD), and have no hot spots in this interface, as shown in Figure 6.3 that provides the computationally predicted hot spot below each model. This implied that agonist vs. antagonist binding may affect interface usage.

6.1.3 GR and MR LBDs can still interact when bound by different ligands, but their interface preference may be affected

The interaction abilities of full length GR and MR bound by different endogenous or synthetic ligands, or by antagonists were investigated. As shown in Appendix Figure C.3, alternative ligands or antagonist binding did not significantly affect the ability of GR and MR to interact, as shown in the number and brightness experiments conducted by our collaborators.

Computational predictions with PRISM (Tuncbag et al. 2011; Baspinar et al. 2014) indicated that the bound ligand or antagonist effects the interface preference of GR and MR. Different interface preference could help explain downstream differences upon drug treatments, as different parts of the protein would be left available for other interactors in each case, altering the interactome of the proteins.

Figure 6.4 shows the energy scores for different binding interfaces for GR-MR partners bound by alternative ligands. Especially the PDB structure 3h52 behaves quite differently; this RU486 antagonist bound GR LBD has an open arm towards the DBD and N-terminal side. This extension restricts its ability to use certain interfaces, as shown in Figure 6.4. This may cause inhibitor bound GR-GR or GR-MR complexes to ‘lock’ in a conformation that limits their interactions with other proteins and alters their downstream effects.

Figure 6.5 shows structures of some of the interactions preferred by different ligand binding situations.

GR type	GR ligand & model	MR ligand & model	interface						
			1s9pAD	5jmBC	2ab2AB	3cldAB	3e7cAB	1p93AC	3k23BC
agonist bound form	cortisol 4p6x	cortico. 2a3i	-75	-60	-76	-63	-54	-51	-45
	cortisol 4p6x	aldo. 2aa2	-81	-83	-55	-67	-89	-57	-49
	cortisol 4p6x	dexa. 4uda	-65	-86	-68	-	-	-46	-40
	cortisol 4p6x	spiro. 3vhu	-81	-107	-60	-78	-87	-50	-81
	dexa. 1p93	cortico. 2a3i	-76	-56	-77	-57	-68	-75	-46
	dexa. 1p93	aldo. 2aa2	-79	-74	-62	-82	-102	-45	-56
	dexa. 1p93	dexa. 4uda	-78	-64	-74	-	-	-46	-40
	dexa. 1p93	spiro. 3vhu	-64	-87	-61	-57	-68	-56	-43
antagonist closed form	RU486 1nhz	cortico. 2a3i	-82	-74	-84	-56	-49	-52	-
	RU486 1nhz	aldo. 2aa2	-76	-83	-53	-90	-82	-38	-
	RU486 1nhz	dexa. 4uda	-74	-106	-72	-	-	-42	-
	RU486 1nhz	spiro. 3vhu	-69	-82	-65	-70	-49	-41	-
antagonist open form	RU486 3h52	cortico. 2a3i	-64	-	-	-	-	-49	-74
	RU486 3h52	aldo. 2aa2	-56	-	-	-	-48	-28	-63
	RU486 3h52	dexa. 4uda	-41	-	-	-	-	-42	-70
	RU486 3h52	spiro. 3vhu	-78	-	-	-	-30	-40	-65

Figure 6.4: Binding energy scores of computational predictions of alternative interfaces between GR and MR with different agonist and antagonist combinations and conformations demonstrate interaction through the LBD regardless of the bound ligand, with varying interface preferences. Two conformations of the crystal structure of GR bound to antagonist RU486 were used; one closed, and one open form where the N-terminal end extending towards the DBD is outward as can be seen in the 3h52 modelled structures in Figure 6.5. The open antagonist bound form has limited interface usage, which can affect cofactor binding. “-” implies no predictions were found by PRISM for the indicated pair of proteins through the indicated interface template.

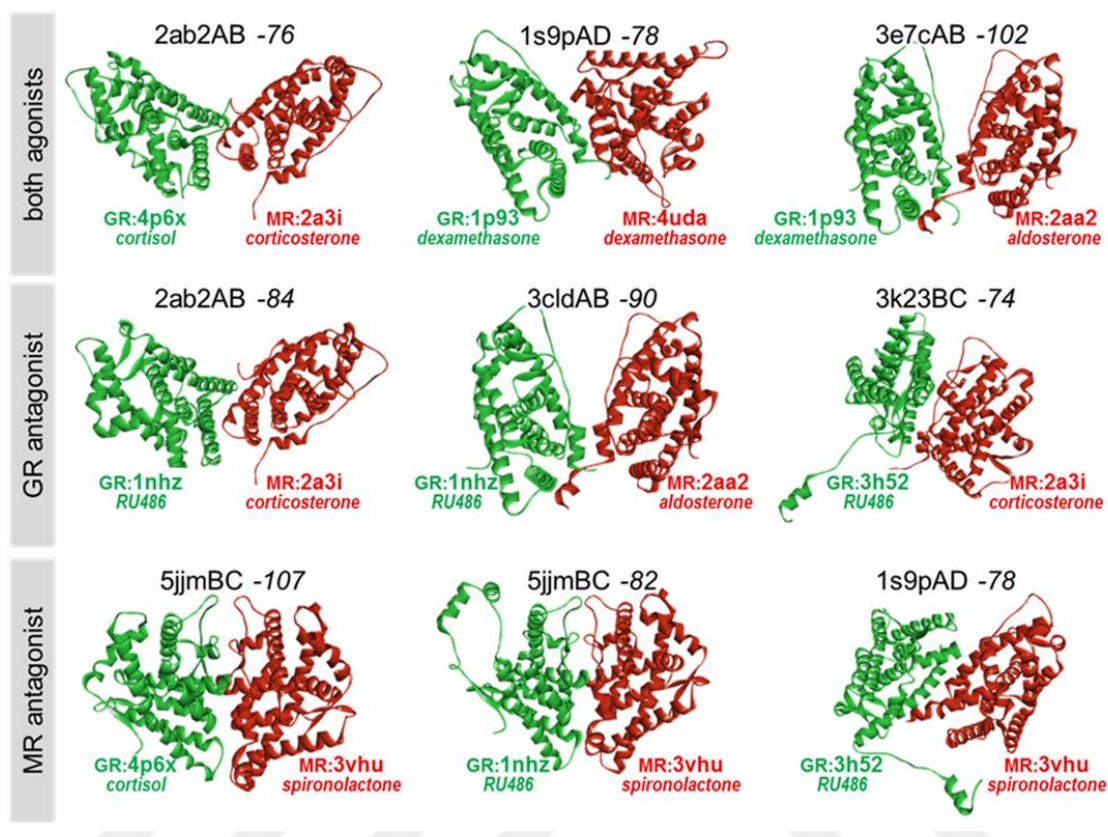


Figure 6.5: Example interaction structures of MR and GR LBDs when both are agonist bound, and when GR or MR is antagonist bound. For each pair of interactors, the most favorable interaction interface was displayed. The PDB IDs of the templates for rat homology modelling are provided below each structure, along with the bound ligand. The interaction interface template and predicted energy score is provided above each structure. cortico, corticosterone; aldo, aldosterone; dexamethasone; spiro., spironolactone.

6.2 Glucocorticoid Receptor Tetramerization

As summarized in the Literature Review section, GR was observed to tetramerize upon DNA binding, and the LBD was essential for this tetramerization. (Presman et al. 2016) The review section also inspected a GR LBD crystal structure that was in tetrameric form. Research has suggested that human GR LBD has more than one dimerization interfaces, as disruption of the ILE628 mediated interface does not completely abolish dimerization. (Bledsoe et al. 2002)

6.2.1 Glucocorticoid Receptor Tetramerization at the Ligand Binding Domain

Formation of a dimer of dimers at the LBD have also been investigated using PRISM protein-protein interaction prediction. (Tuncbag et al. 2011; Baspinar et al. 2014) Results suggest that alternative tetramerization modes of GR LBD may be possible, as shown in Figure 6.6 below.

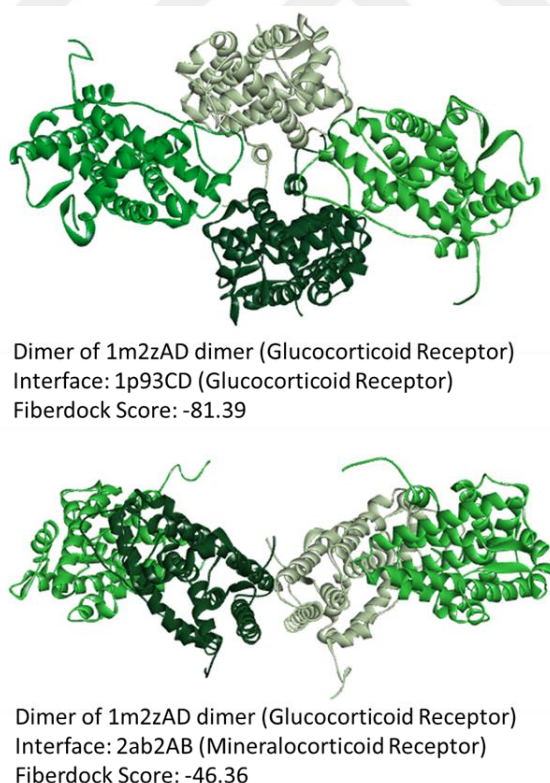


Figure 6.6: PRISM predictions suggest alternative tetramerization modes of GR LBD may be possible, in the form of a dimer of dimers, as indicated in the publication. (Presman et al. 2016) Interfaces are from GR or the highly homologous MR.

6.2.2 *Glucocorticoid Receptor tetramerization through DNA Binding Domain possible with chromatin looping*

The publication on GR tetramerization upon DNA binding suggested that there could be two possibilities; either both dimers bind the same DNA strand, or the dimers bind distant pieces of DNA which are brought close by chromosomal looping, and the dimers can form tetramers (Presman et al. 2016). The results we obtain with PRISM support the chromosomal looping model as shown below. Protein-protein interaction between two dimers of GR do not seem possible on the same DNA strand according to our predictions.

Furthermore, in accordance with the finding that the LBD is required for tetramerization, the interaction energy score of the DBD tetramer is much less favorable compared to the LBD predictions, as shown in the figure below.

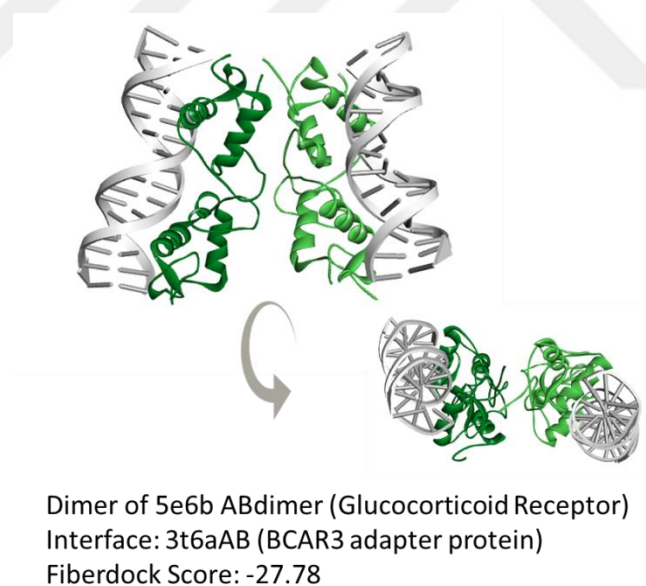


Figure 6.7: PRISM predictions suggest GR tetramerization can be supported through the DBD as well through chromosomal looping, but the energy is less favorable compared to the LBD tetramers, which is in accordance with the experimental observation that GR tetramerization is LBD dependent.

6.3 Discussion

Experimental disruption of a single interface between GR and MR was observed not to be sufficient for completely abolishing the interaction of GR and MR. Our predictions with PRISM demonstrated that the reason for this observation might be that GR and MR are able to interact through alternative interfaces shared by close relatives from the steroid receptor family of proteins.

Experiments also suggested that GR and MR are able to interact when bound by different ligands and antagonists. Our predictions demonstrated that different ligands or antagonist binding does not abolish the GR-MR interaction but instead changes their interface preference.

Also according to computational structural predictions, GR tetramerization affinity depends mostly on the ligand binding domain, but is also supported by the DNA binding domain, through chromosomal looping. Structure is depicted in Figure 6.8 below.

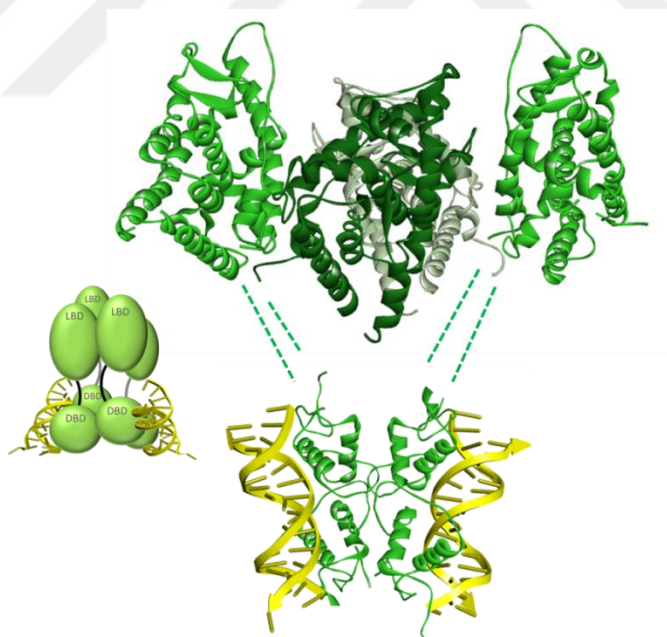


Figure 6.8: GR tetramerization structural model. According to computational structural predictions, GR tetramerization affinity depends mostly on the ligand binding domain, but is also supported by the DNA binding domain, through chromosomal looping, in accordance with the publication (Presman et al. 2016) where it was demonstrated that DNA binding and the presence of the LBD are necessary for tetramerization.

6.4 *Future Perspectives*

Similar interaction behaviors between different but closely related steroid receptors are likely to be observed between other members of this transcription factor family. As cited in the Literature Review section, for example, a recent publication in *Nature* has identified that interaction between the estrogen receptor (ESR) and progesterone receptor (PR) plays a role in breast cancer (Mohammed et al. 2015).

Such interactions between other steroid receptors are likely to have similar properties to the ones identified here. The glucocorticoid receptor (GR), mineralocorticoid receptor (MR), androgen receptor (AR) and progesterone receptor (PR) are structurally very similar. Therefore, they could form alternative hetero-interactions or homo-interactions, using similar interfaces. The interfaces currently available in the protein data bank in crystallized form for these receptors are likely to be utilized by other members of this family due to their structural similarity. It is quite probable that these receptors have similar behaviors such as ability to use alternative interfaces depending on cellular context. The observation that GR, PR and MR can recognize the core sequence of the androgen response element (ARE) is also an indicator of how similar these transcription factors can behave (Kim et al. 2006).

As shown in the literature section in Figure 2.19, steroid receptors other than GR, namely PR and AR, were also observed to form higher-order oligomers, such as tetramers, or perhaps even hexameric complexes. This is very interesting, and may pave the way for going beyond the traditional view that steroid receptors act in dimers.

Chapter 7:

**TRANSCRIPTOME ANALYSIS OF CYTOCHROME DELETION
IN P53 MUTANTS**

This work has been published in Functional & Integrative Genomics in May 2019. (Cavga et al. 2019) UV treatment of cells, RNA-seq library preparation and qRT-PCR experiments have been performed by Dr. Mehmet Tardu in the laboratory of Prof. Dr. Halil Kavaklı RNA-sequencing analysis, GSEA and PPI network analysis has been performed by Ayşe Derya Cavga. Supplementary experiments for the confirmation of apoptotic status were carried out by Tuba Korkmaz and Assoc. Prof. Nuri Öztürk in Gebze Technical University. (Cavga et al. 2019)

7.1 Confirmation of higher apoptosis in Cry deletion in a p53 mutant background

It has previously been shown that $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ MSF cells are more susceptible to ultraviolet (UV)-induced apoptosis than $p53^{-/-}$ MSF cells (Ozturk et al. 2009). To confirm this observation $p53^{-/-}$ and $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ MSF cells treated with oxaliplatin as UV-mimetic agent are probed for apoptosis and searched for the presences of PARP cleavage, which is an indicator of the apoptosis. As expected, Cry knockout cells in the p53-null background ($p53^{-/-}$, $Cry1^{-/-}$, $Cry2^{-/-}$) had more apoptosis compared to only p53-null ($p53^{-/-}$) cells in response to oxaliplatin in agreement with previous findings of $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ (**Appendix Figure D.1**) (Ozturk et al. 2009; Lee et al. 2013). Additionally, we also confirmed such phenomena are applicable in cancer cells, in the absence of Cry genes using the p53 mutant MDA-MB-231 breast cancer cell line. The results demonstrated that the absence of Crys resulted in enhanced PARP and Casp3 cleavage compared to p53 mutant cells (**Appendix Figure D.2.B,C**) which is consistent with a previous study (Korkmaz et al. 2018). Additionally, survival rate of the MDA-MB-231 cells were determined for 24 h and 48 h. They had comparable survival rate compared with wild type cells (**Appendix Figure D.2.D**).

These validations were carried out by Tuba Korkmaz and Assoc. Prof. Nuri Öztürk in Gebze Technical University.

7.2 RNA-seq data analysis showed high unique mapping and coverage

After all these confirmations, we, then, aimed to identify genes and pathways that contribute to UV-induced apoptosis at the transcriptome level in these cell lines. The cells were exposed to UV and total RNA was isolated and sequenced as described in the Methods section (**Figure 3.4**) and in the supplementary methods in Appendix D.

These experiments were performed by Dr. Mehmet Tardu from the laboratory of Prof. Dr. Halil İ. Kavaklı in Koç University.

After sequencing and de-multiplexing, RNA-seq data was aligned to the mouse reference genome GRCm38 using the STAR Aligner (Dobin et al. 2013).

Results indicated that the reads were uniquely mapped to gene loci (**Appendix Figure D.3.A**). To evaluate reproducibility among biological replicates, Pearson's correlation tests were performed on the expression values. There were strong correlations between biological replicates for each condition (Pearson's correlation coefficients, $R^2 > 0.95$) (**Appendix Figure D.3.B**).

All these quality control assessments allowed further analysis of the data to identify the differentially expressed genes (DEGs). For all analyses, only data from reads that uniquely mapped to the reference genome were used, and all multi-mappers were excluded. Differential expression analysis was performed using the Cuffdiff command in the Cufflinks suit (Trapnell et al. 2012) as described in Tardu et al (Tardu et al. 2016, 2017). The fold change in the number of mapped Fragments Per Kilobase of exon per Million mapped reads (FPKM) between UV-treated and non-treated samples were calculated using Cuffdiff.

Transcripts with a 0 (zero) FPKM read-out in any one of the 4 samples were eliminated due to causing division errors, however, the possible impact of including such transcripts in the analysis are discussed in detail in Section 7.7 . These cutoffs yielded a total of 14659 differentially analyzable genes (**Figure 3.4.B**).

7.3 Differentially Expressed Genes were confirmed via qRT-PCR

The validity and reproducibility of the RNA-Seq data was confirmed via qRT-PCR analysis on 11 genes. Total RNA was isolated from both $p53^{-/-}$ and $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ MSF cells were exposed to UV. After conversion of the total RNA into cDNA, real-time PCR (qRT-PCR) was performed using appropriate primers (Appendix Table D.1). A comparison of qRT-PCR results and RNA-seq results revealed similar expression patterns for each gene, indicating that the RNA-seq and qRT-PCR results were correlated, allowing confidence in the data and further analysis (**Figure 7.1.A,B**). After the identification of key genes, the expression level changes in 11 key genes were also confirmed via qRT-PCR (**Figure 7.1.C,D**).

These qPCR validations were carried out by Dr. Mehmet Tardu from the laboratory of Prof. Dr. Halil İ. Kavaklı in Koç University, and by Tuba Korkmaz and Assoc. Prof. Nuri Öztürk in Gebze Technical University. Primers and methodology are provided in Appendix D.

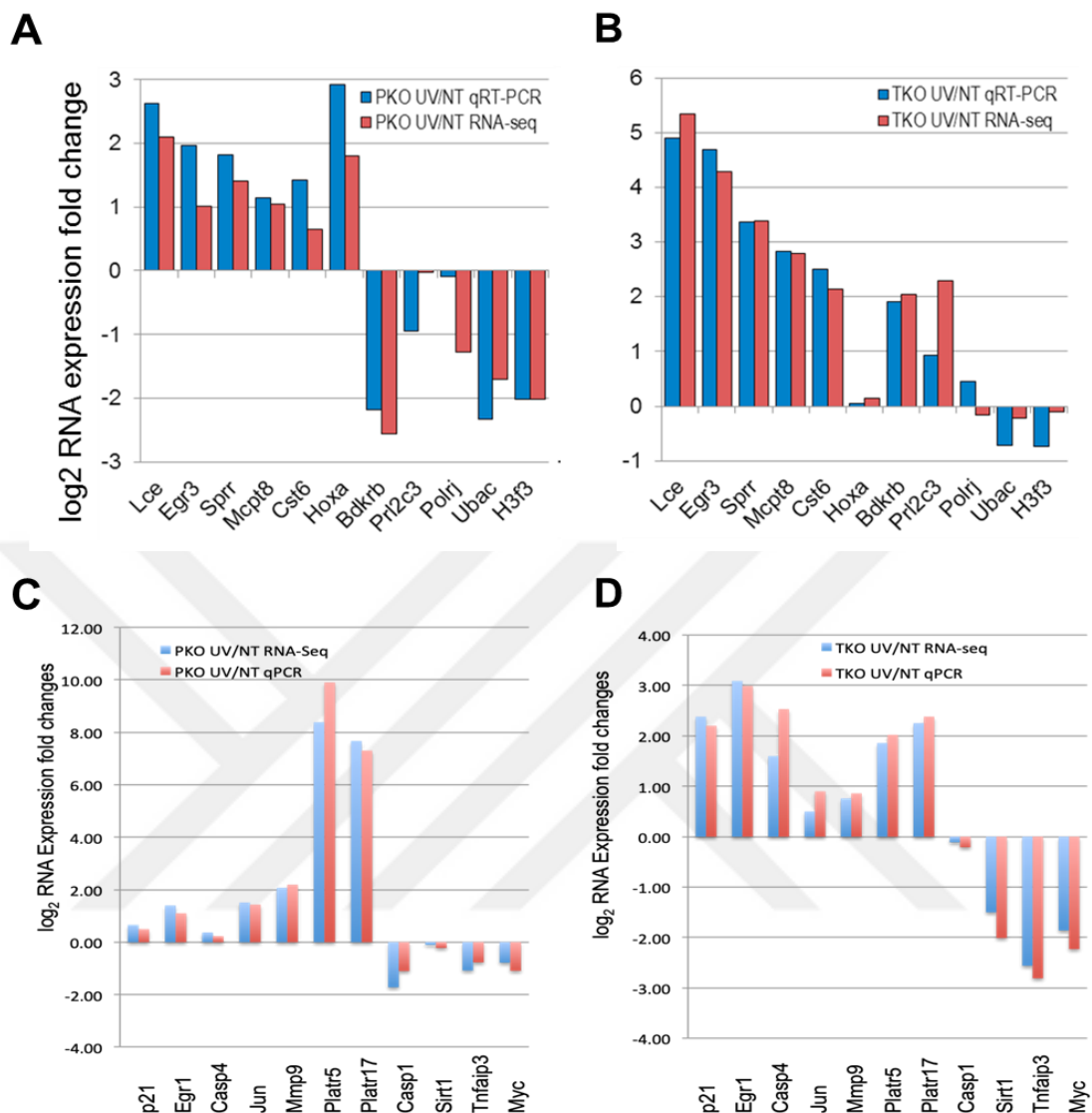


Figure 7.1: Validation of RNA-seq data with qRT-PCR. RNA-seq results were compared with qRT-PCR results initially in randomly selected 11 DEGs (A&B) and later in identified 11 key DEGs (C&D). Values represent $\log_2[(\text{PKO_UV}/\text{PKO_NT})]$ and $\log_2[(\text{TKO_UV}/\text{TKO_NT})]$ ratios.

These qPCR validations were carried out by Dr. Mehmet Tardu from the laboratory of Prof. Dr. Halil İ. Kavaklı in Koç University, and by Tuba Korkmaz and Assoc. Prof. Nuri Öztürk in Gebze Technical University. Primers and methodology are provided in Appendix D.

7.4 Identification of differentially expressed genes (DEGs) in $p53^{-/-}$ Cry1 $^{-/-}$ Cry2 $^{-/-}$ cells compared to $p53^{-/-}$ cells in response to UV

Analysis of the DEG results indicated that 5327 and 5112 genes were differentially regulated in response to UV in $p53^{-/-}$ and $p53^{-/-}$ Cry1 $^{-/-}$ Cry2 $^{-/-}$ cells, respectively. Among these, 2339 DEGs were upregulated, and 2988 DEGs were downregulated in $p53^{-/-}$ cells while 2216 DEGs were upregulated and 2896 DEGs were downregulated in $p53^{-/-}$ Cry1 $^{-/-}$ Cry2 $^{-/-}$ cells (**Figure 3.4.B, Appendix Figure D.4.A**).

Since the aim of this study was to identify the genes that were differentially expressed in response to UV treatment upon deletion of *Cry* genes in a $p53^{-/-}$ mutant background, affected genes of interest were identified by comparing the change in FPKM value for each transcript in $p53^{-/-}$ Cry1 $^{-/-}$ Cry2 $^{-/-}$ UV-treated (TKO_UV) versus non-treated (TKO_NT) to the change in $p53^{-/-}$ UV-treated (PKO_UV) versus non-treated (PKO_NT). A gene was considered significant in terms of altered UV damage response in $p53^{-/-}$ Cry1 $^{-/-}$ Cry2 $^{-/-}$ versus $p53^{-/-}$ cells only if it showed a change in change more than two folds; that is, if the absolute value of the DEG Score in Equation 1 was more than 1 (**Figure 3.4.B**).

This analysis resulted in a total of 626 transcripts. 267 of these transcripts were significantly differentially expressed in response to UV in both cell lines, and their fold changes were more than two-fold between the two cell lines. 189 genes were significantly differentially expressed in response to UV in $p53^{-/-}$ cells but not in $p53^{-/-}$ Cry1 $^{-/-}$ Cry2 $^{-/-}$ cells, while 170 genes were significantly differentially expressed in response to UV in $p53^{-/-}$ Cry1 $^{-/-}$ Cry2 $^{-/-}$ cells but not in $p53^{-/-}$ cells (**Figure 3.4.B**). The DEG Score for each gene was calculated with Equation 1, where genes that received a score above 1 were considered as “increased response” genes, and genes receiving a score below -1 were considered as “decreased response” genes ((**Figure 3.4.C, Appendix Figure D.4.B**)). The up- or down-regulation statuses of these genes in response to UV within $p53^{-/-}$ cells or $p53^{-/-}$ Cry1 $^{-/-}$ Cry2 $^{-/-}$ cells are also provided (**Appendix Figure D.4.C**).

7.5 Gene Set (GSEA) and KEGG Enrichment Analyses Identified Pathways that may Enhance Apoptosis in $p53^{-/-}$ $Cry1^{-/-}$ $Cry2^{-/-}$ MSF Cells upon Exposure to UV

To functionally characterize the 626 DEGs and identify pathways, GSEA and KEGG Enrichment Analyses were performed (Subramanian et al. 2005). The GSEA results revealed that increased response genes were enriched in the Interferon-gamma ($IFN\gamma$) response and tumor necrosis factor alpha ($TNF\alpha$) signaling via nuclear factor kappa-B ($NF-\kappa B$) related genes (**Figure 7.2.A**). Our results were consistent with a previous study, which had shown that the absence of the *Cry* genes sensitize $p53^{-/-}$ mutant cells to tumor necrosis factor α ($TNF\alpha$)-initiated apoptosis through $NF-\kappa B$ signaling, through the inactivation of $GSK3\beta$ (Lee and Sancar 2011b). Another previous study also discussed that enhanced immune-surveillance of cancer cells may also help delay the onset (Ozturk et al. 2009). The present study showed that the increased response genes were enriched in genes upregulated during $IFN\gamma$ response, which is known to be essential for immunity against pathogens and for tumor surveillance (Schoenborn and Wilson 2007). Therefore, an increased $IFN\gamma$ response indicates an increased immune surveillance of tumor cells, which is likely to contribute to the delayed onset of cancer.

GSEA also enabled identification of additional genes in these pathways (**Figure 7.2.B**) which possibly contribute to enhanced apoptosis in $p53^{-/-}$ mice upon deletion of *Cry* genes. Some of these genes act in $IFN\gamma$ response, such as *Casp1*, *Casp4*, and some act in $TNF\alpha$ signaling, such as *Egr3* and *Jun*. Interestingly, some genes are involved in both pathways, such as *Tnfaip3* and *Cdkn1a*. The KEGG pathway analysis allowed identification of genes such as *Mmp9* and histones, which were enriched in the KEGG pathway “Transcriptional Misregulation in Cancer”. We further confirmed that the expression levels of the identified genes were significantly changed in PKO and TKO cell lines with qPCR. Results indicated that the transcriptional levels were similar to those identified by RNA-seq (**Figure 7.1**).

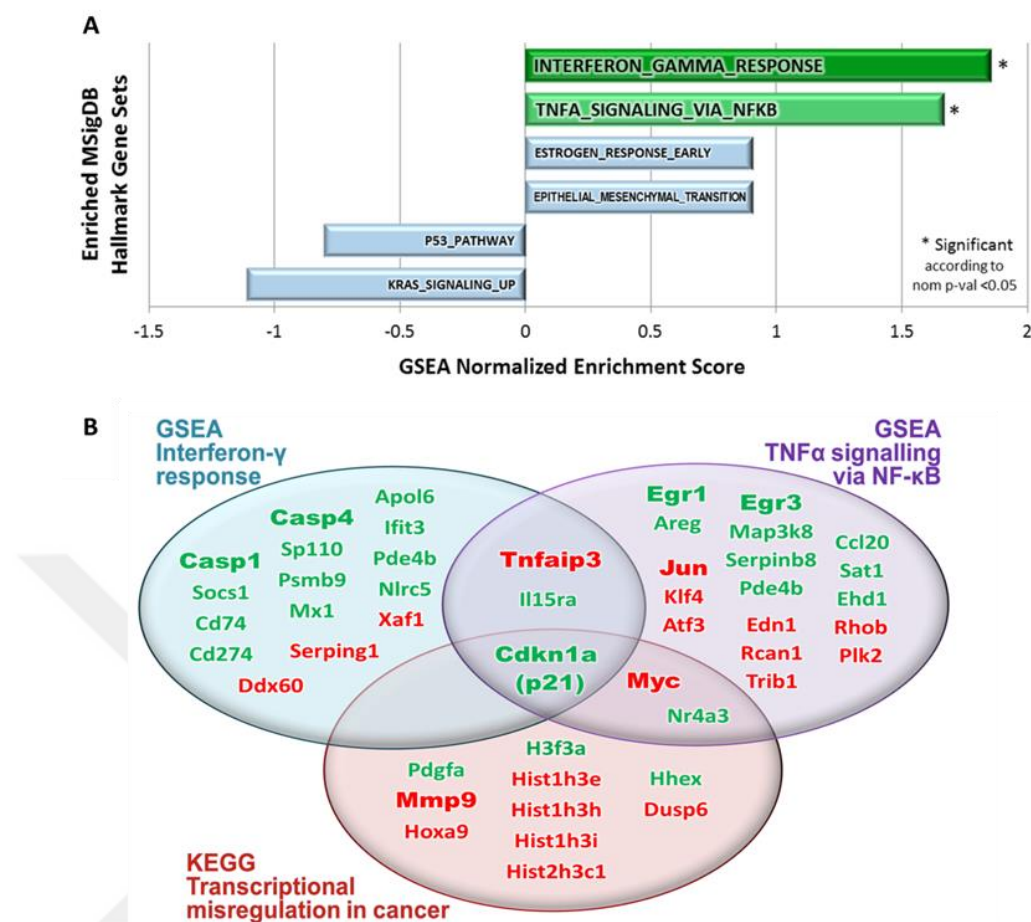


Figure 7.2: GSEA and KEGG Pathway analysis of 626 DEGs demonstrate that *Cryptochrome* mutations in *p53* mutant cells cause increased responses in IFN γ response and TNF α signaling related genes, as well as transcriptional alterations significant in cancer. **A)** GSEA revealed that the 626 genes that were significantly differentiated in response to UV in *p53*^{-/-}*Cry1*^{-/-}*Cry2*^{-/-} cells compared to *p53*^{-/-} cells are enriched significantly in 2 MSigDB Hallmark gene sets. The Hallmark INTERFERON_GAMMA_RESPONSE gene set is composed of 200 genes up-regulated in response to IFNG across 82 founder gene sets. The Hallmark TNFA_SIGNALING_VIA_NFKB set is made up of 200 genes regulated by NF- κ B in response to TNF α generated from 132 founder gene sets (Subramanian et al. 2005). **B)** Venn diagram depicting genes from the 626 selected genes that belong to 2 MSigDB Hallmark gene sets and the KEGG Pathway “Transcriptional misregulation in cancer”. KEGG Pathway enrichment was analyzed with the ClueGO Extension Tool of Cytoscape (Bindea et al. 2009).

Tumor necrosis factor alpha-induced protein 3 (*Tnfaip3*, A20) is at the intersection of the IFN γ response and the TNF α signaling gene sets (**Figure 7.2.B**) and shows a decreased response to UV in *p53^{-/-}Cry1^{-/-}Cry2^{-/-}* compared to *p53^{-/-}* knockouts (**Figure 7.3**). This is particularly interesting because *Tnfaip3* normally inhibits NF- κ B activity and TNF-mediated programmed cell death (Ma and Malynn 2012) and decreased response can therefore contribute to increased apoptosis. Reduced *Tnfaip3* response may also contribute to immune activation since *Tnfaip3* is an important gatekeeper in immunity and *Tnfaip3* deficient mice were observed to show severe inflammation (Lee et al. 2000; Coornaert et al. 2009).

Cyclin-Dependent Kinase Inhibitor 1A (*Cdkn1a*, *p21*) is also at the intersection of DEGs belonging to both the IFN γ response and the TNF α signaling gene sets and it is among increased response genes (**Figure 7.3**). *p21* is known to be a key apoptotic regulator in downstream of both *p53* and *p73* (Fontemaggi et al. 2002) and the *p73* tumor suppressor gene was previously shown to be specifically upregulated in the *p53^{-/-}Cry1^{-/-}Cry2^{-/-}* cells (Lee and Sancar 2011a). The cyclin-dependent kinase inhibitor CDKN1A binds to and inhibits the activity of cyclin-CDK2 or -CDK4 complexes, and therefore acts as a regulator of cell cycle progression at G1 (Wade Harper et al. 1993).

Another notable outcome of the GSEA is to identify up-regulation of *Egr3* and *Egr1* in the TNF α signaling gene set. Up-regulation of the *Egr1* was previously shown to be involved in enhancing apoptosis in *p53^{-/-}Cry1^{-/-}Cry2^{-/-}* mice by transactivating *p73* (Lee and Sancar 2011a). Our results suggest that the increased response in *Egr3* may have a cooperative effect with *Egr1* in this signaling pathway (**Figure 7.3**).

Caspase-1 (*Casp1*) and Caspase-4 (*Casp4*) were observed to be among the genes that show an increased response to UV in *p53^{-/-}Cry1^{-/-}Cry2^{-/-}* knockout cells compared to *p53^{-/-}* knockout cells (**Figure 7.3**). CASP1 is involved in the IFN γ response where it cleaves and activates the pro-inflammatory cytokines proIL-1 β and proIL-18. In fact, a CASP1-dependent programmed cell death named pyroptosis has recently been identified, and was observed to trigger an inflammatory response (Sollberger et al. 2014). Therefore, increased response of this gene may be triggering immune response in *p53^{-/-}Cry1^{-/-}Cry2^{-/-}* knockout mice and contribute to delayed onset of cancer compared to *p53^{-/-}* knockout mice.

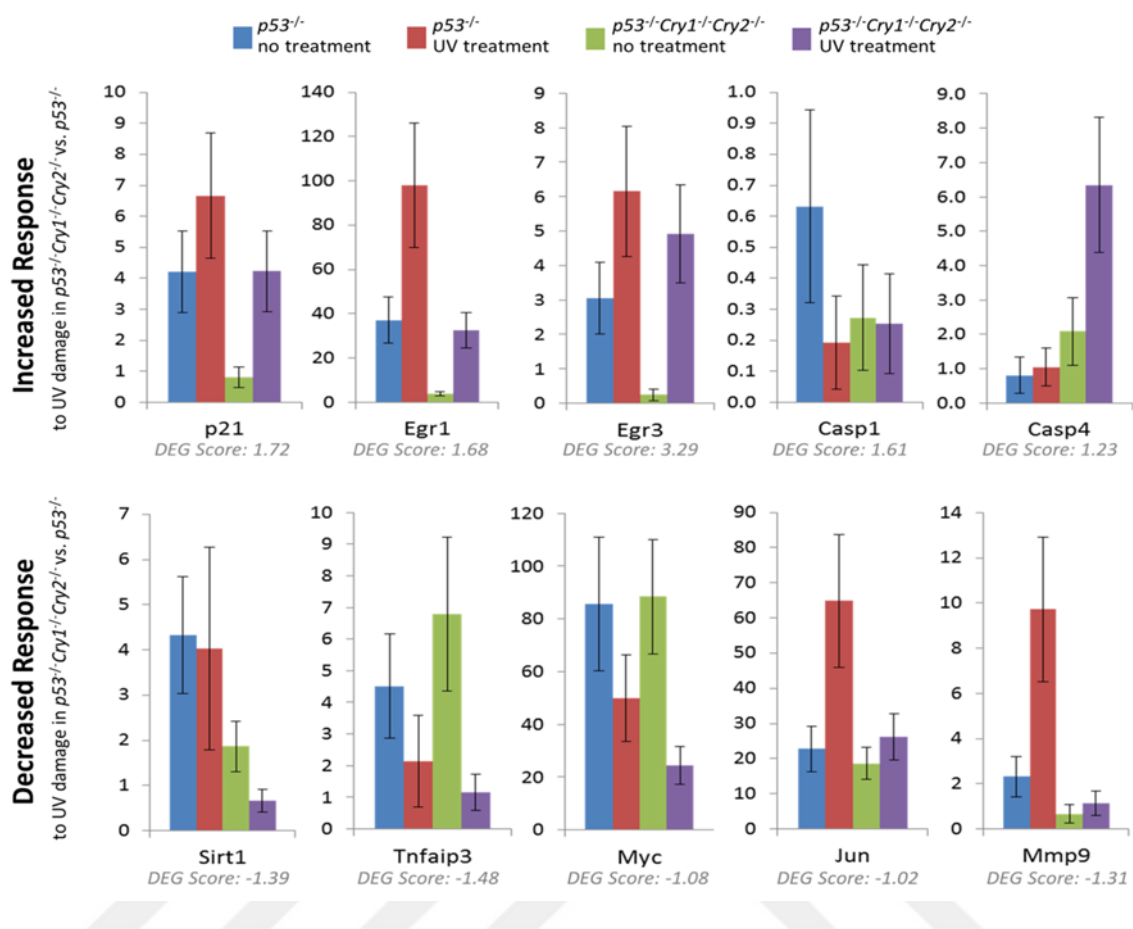


Figure 7.3: RNA FPKM levels of key genes from gene set enrichment analysis MSigDB hallmark gene sets and PPI network analysis. Numbers in y axes represent FPKM values from RNA-Seq analyses, with different scales for each gene. Error bars represent confidence interval for the estimation of the FPKM, which is output by Cufflinks. The DEG Score = $\log_2[(TKO_UV/TKO_NT)/(PKO_UV/PKO_NT)]$ value from Equation 1 is provided for each gene.

Myc is another particularly interesting DEGs identified in the TNF α signaling via NF- κ B and transcriptional misregulation in cancer gene sets (**Figure 7.2.B**). *Myc* is a proto-oncogene that acts as a transcription factor in cell cycle progression, apoptosis and cellular transformation. Mutations, overexpression, rearrangement and translocation of *Myc* have been associated various hematopoietic tumors, leukemias and lymphomas in humans (Hofmann et al. 2015). Interestingly, *Myc* is more downregulated (**Figure 7.3**) by UV in the *p53*^{-/-}*Cry1*^{-/-}*Cry2*^{-/-} knockout setting, which may help explain the slower development of cancers such as lymphomas compared to the *p53*^{-/-} knockout mice.

KEGG pathway analysis showed enrichment in genes involved in transcriptional misregulation in cancer, and drew attention to a particular gene *Mmp9*, encodes matrix metalloproteinase 9 (**Figure 7.2.B**). Matrix metalloproteinases (MMPs) are involved in the breakdown of extracellular matrix, which makes them a suspect in cancer metastasis and angiogenesis. MMP9, is shown to be involved in breast cancer metastasis (van Kempen and Coussens 2016). Therefore, it may be important that *Mmp9* has a decreased response to UV in the *p53*^{-/-}*Cry1*^{-/-}*Cry2*^{-/-} MSF compared to the *p53*^{-/-} MSF cells.

All these results suggested that additional genes are required in the important pathways that are likely to cause delayed onset of cancer upon *Cryptochrome* deletion in a *p53*^{-/-} mutant background.

7.6 Protein-protein interaction network (PPIN) analysis helps identify key protein interactions responsible for delayed onset of cancer

In addition to the KEGG and the GSEA, PPIN analysis of the 626 DEGs was also performed to identify protein-protein interactions that might play key roles related to cancer upon deletion of *Cry* genes in a *p53* mutant background. *P53* (*Trp53*) was included as a seed gene to reveal proteins that interact directly or indirectly with p53 when it is present, in order to understand which proteins might be interacting with p73 or similar proteins from this gene family that act when p53 is absent (Harms et al. 2004; Lin et al. 2009). The topological analysis of the PPI revealed the top hub and bottleneck proteins that are at key pathway intersections, as described in the methods section.

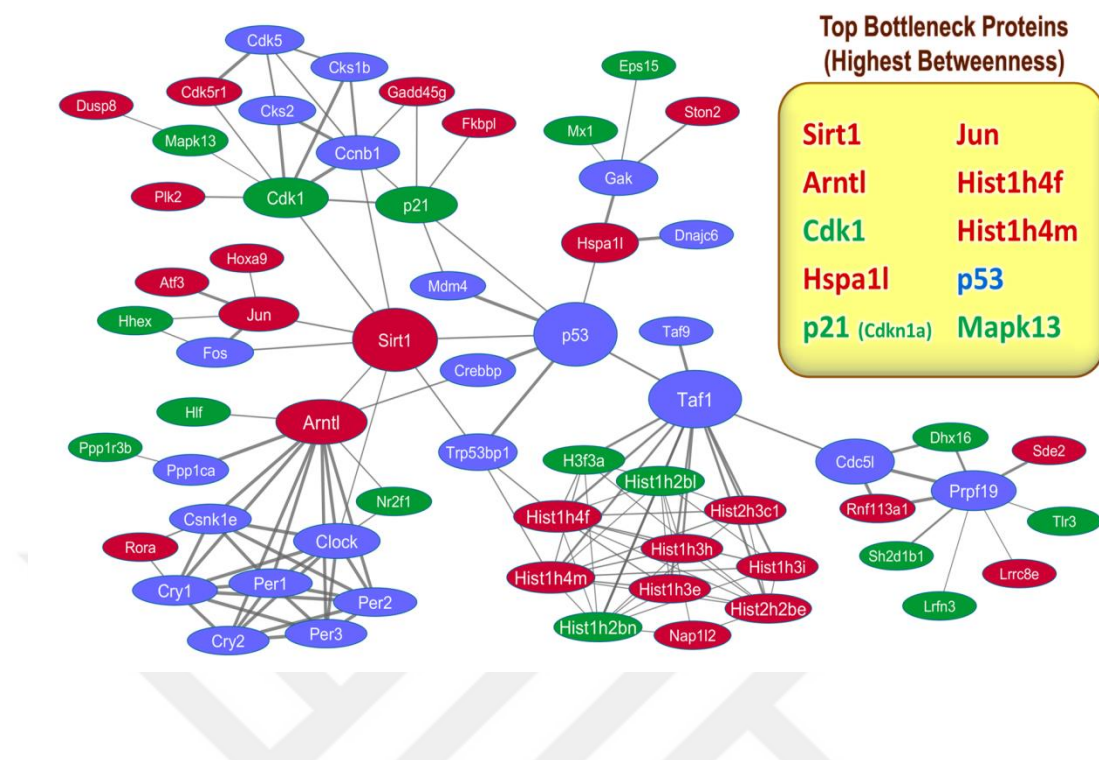


Figure 7.4: Protein-protein interaction network analysis reveals key genes that interact with p53 associated genes. PPI network was constructed using STRING database for protein-protein interaction information, and Cytoscape for visualization and analysis (Shannon et al. 2003; Jensen et al. 2009). Interaction information is based exclusively on experiments and with a confidence score of 0.4 or above. A maximum of 50 additional interactors per protein were included. Of the 626 DEGs, 487 genes were used as seed genes as described in the Methods section in detail. The STRING output is provided in the references as a permanent link (Jensen et al. 2018). Ellipses represent nodes, which are proteins, and the lines between them represent protein-protein interactions, are called edges. The nodes are color coded according to their change in response to UV in *p53^{-/-}Cry1^{-/-}Cry2^{-/-}* MSFs compared to *p53^{-/-}* MSFs; green indicates increased response genes, red indicates decreased response genes. The blue nodes are additional nodes added by STRING database based on experimental interaction data with high confidence score. The node width represents the degree of each node, and the node height represents the betweenness centrality of each node; therefore, proteins with higher significance in the network appear bigger. The edge thickness and opaqueness increases with higher confidence score for the interaction from the STRING database. Top 10 seed genes with highest degrees and betweenness centralities are represented in the Venn diagram, and colored green if they are from the increased response genes, or red if they are from the decreased response genes.

The PPI pathway analysis draws attention to Sirtuin 1 (SIRT1) as a protein that connects p53 related proteins to other key proteins in the network (**Figure 7.4**), and it is among the significantly decreased response genes (**Figure 7.3**). SIRT1 is a NAD-dependent protein deacetylase that can regulate the activity of key proteins involved in cancer (Luo et al. 2001). Most interestingly, *Sirt1* mutant cells were shown to have increased apoptosis in response to stress, and SIRT1 was shown to interact with p73 and suppress its transcriptional activity (Luo et al. 2001; Jin et al. 2007). Therefore, it is particularly significant that the levels of this protein are decreased in $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ cells compared to $p53^{-/-}$ cells (**Figure 7.3**). This reduction in *Sirt1* transcriptional levels may help explain the elevation of p73 target genes such as *p21*, and delayed onset of cancer due to increased apoptosis.

GSEA and PPI Network analyses suggested *Jun* as a key gene that was shown to have decreased response to UV damage in $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ MSF cells compared to $p53^{-/-}$ MSF cells. This observation is in agreement with the previous finding that Cryptochrome mutations sensitize $p53^{-/-}$ mutant cells to TNF α -induced apoptosis through NF- κ B signaling (Lee and Sancar 2011b). Interestingly, *c-Jun* was previously shown to have a protective function from UV-induced cell death (Wisdom et al. 1999). The protection yielded by JUN may be lifted by *Cryptochrome* deletion, making the cells more prone to apoptosis.

7.7 Correction of zero division errors allows gain of information on non-coding genes

There is no consensus “best practice” on how to handle 0 (zero) FPKM readings during RNA-seq analysis; one published approach is to add 0.1 to the raw counts (Li et al. 2017), and another is to set a pseudo count of 0.001 (Rapaport et al. 2013) to all zero entries. Both approaches have been tried and evaluated. The first approach of adding 0.1 to the raw counts resulted in a reduced DEGs list of 487 genes. This approach resulted in loss of information because a large fraction of the differentially expressed genes have FPKM values between 0 and 1, and an addition of 0.1 alters differential analyses. Appendix Table D.4)

The second approach of adding 0.001 to the raw counts resulted in a significant DEGs list of 696 genes, yielding 70 additional significant transcripts to the 626 that had been identified without changing the raw counts. (Appendix Table D.3) This list of 696 genes was evaluated as an input gene list to the GSEA, which resulted in the same list of enriched Hallmark gene sets, but in this case, all of the gene sets were outside the significance threshold of a maximum nominal P value of 0.05. This was probably affected by the fact that the majority of the additional 70 genes have more significant DEG scores than the genes in the initial list since near-zero divisions produce large numbers, and therefore mask the significance ranking of other genes. (Appendix Table D.3)

The MSigDB data collections of the GSEA tool are based on expression array data, which has much less sensitivity compared to RNA-seq data, and would not have been able to detect transcripts with near-zero expression levels. Since the gene set data of the GSEA were curated from expression arrays, it seemed unrealistic and unfitting to include the near-zero, sensitively obtained RNA-seq data in the GSEA analysis. Therefore, the GSEA analysis was carried out with the 626 DEGs that did not include the genes with zero FPKM values.

Importantly however, the majority of these ‘zero cases’ are non-coding transcripts, and some of these changes are at quite significant levels. For example, for the Gm26168 non-coding transcript in *p53^{-/-}Cry1^{-/-}Cry2^{-/-}* cells, the FPKM reading in the non-treated sample is 0, while it moves up to an astonishing 706.18 FPKM reading upon UV treatment. (Appendix Table D.3)

Of these additional 70 ‘zero case’ genes, 46 are non-coding transcripts comprised of pseudogenes, long noncoding RNA (lncRNA) genes, or unclassified non-coding RNA genes according to the updated Mouse Genome Database, as shown in **Figure 7.5** (Appendix Table D.3) (Smith et al. 2018).

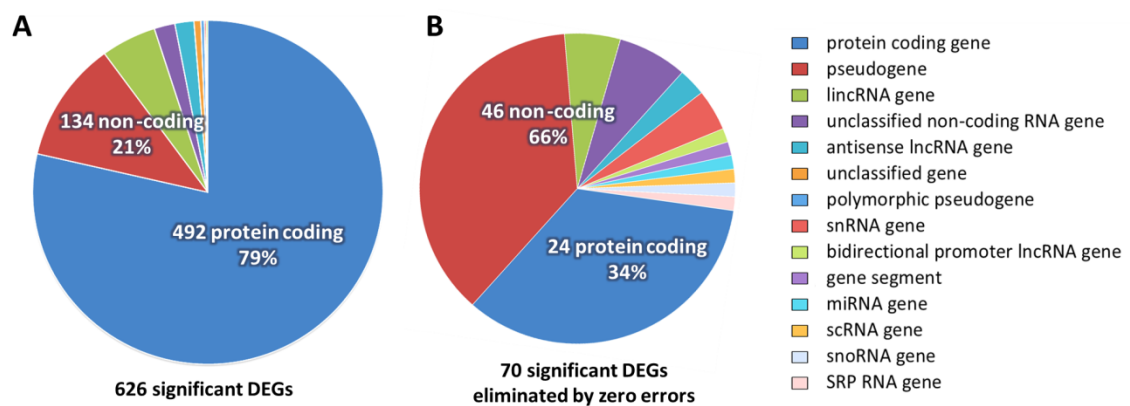


Figure 7.5: Elimination of transcripts with zero FPKM causes loss of information, especially on non-coding transcripts. A) The feature types of the 626 significantly differentially expressed genes (DEGs) in response to UV in $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ cells compared to $p53^{-/-}$ cells. B) The feature types of 70 significant DEGs that were eliminated due to a lack of a consensus approach on how to handle 0 (zero) FPKM readings that cause division errors. 46 out of 70 such DEGs were non-coding transcripts, with examples such as 0 to 706.18 FPKM changes between TKO_NT to TKNO_UV, which are actually quite significant (Appendix Table D.3). Feature type information obtained from the MGI database (Smith et al. 2018).

Although the transcriptions of these non-coding RNAs were highly regulated by UV exposure in $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ cells, only a few of them are annotated and this make hard to extrapolate their function in delaying the onset of cancer. Two long non-coding RNA genes that have been annotated are pluripotency associated transcript 5 (Platr5) and pluripotency associated transcript 17 (Platr17), which have DEG scores of -9.84 and -8.71, respectively (Appendix Table D.3). Interestingly, using weighted gene coexpression network analysis, these transcripts have previously been identified as lincRNAs associated with the pluripotent state of ESCs, clustering tightly among known pluripotency factors such as *Nanog* (Bergmann et al. 2015). Whether these noncoding RNAs play roles in apoptosis depending on the absence of *Cry* genes in $p53$ mutant background are yet to be shown.

Interestingly, the transcriptional level of a noncoding transcript, which is not annotated Gm26168, sharply increased from 0 to 706.18 FPKM reading upon UV treatment of *p53^{-/-}Cry1^{-/-}Cry2^{-/-}* cells, while such increase for the Gm26168 in response to UV is not observed in *p53^{-/-}* cells (Appendix Table D.3). Our data suggests that such novel noncoding RNAs may play roles in apoptosis in a Cry dependent manner in the absence of p53.

The comprehensive roles of these non-coding transcripts in the context of circadian clock disruption and *p53* mutation will be better understood as RNA-seq analyses become more widely used to detect alterations in the levels of these transcripts in similar contexts, and as more non-coding transcripts are functionally characterized.

7.8 Discussion

The GSEA and PPI network analyses of the differentially expressed transcripts identified by RNA sequencing demonstrated that deletion of both *Cry1* and *Cry2* genes in the *p53^{-/-}* background causes increases in pro-apoptotic and anti-tumorigenic pathways such as TNF α via NF- κ B and immune-surveillance through IFN γ , in response to UV damage. These results are in agreement with previous findings that implicated involvement of NF- κ B regulated cell autonomous mechanisms, and draw attention to the roles of enhanced immune surveillance and the possible involvement of key genes such as *p21*, *Egr3*, *Sirt1*, *Jun*, *Mmp9* and *Casp1*.

The results obtained allow the proposition of the pathway model described in **Figure 7.6**. This model represents how the newly identified genes may be involved in the delayed onset of cancer upon deletion of *Cry* genes in *p53* mutant background in light of previous knowledge. According to the results obtained, CASP1 and CASP4 may work under IFN γ during immune response (Sollberger et al. 2014). TNFAIP3 and JUN, working under TNF α , can contribute to the regulation of NF- κ B and enhanced apoptosis (Wisdom et al. 1999; Ma and Malynn 2012). p21 is possibly involved in the increase of apoptosis mediated by p73, which is consistent with a previous report (Lee and Sancar 2011a) and the decrease in SIRT1 may result in activation of p73 and enhance apoptosis

in $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ cells (Jin et al. 2007). This study and the presented model opens new avenues for further investigation in explaining the delayed onset of cancer brought about by a Cryptochrome deletion in a $p53$ knockout background.

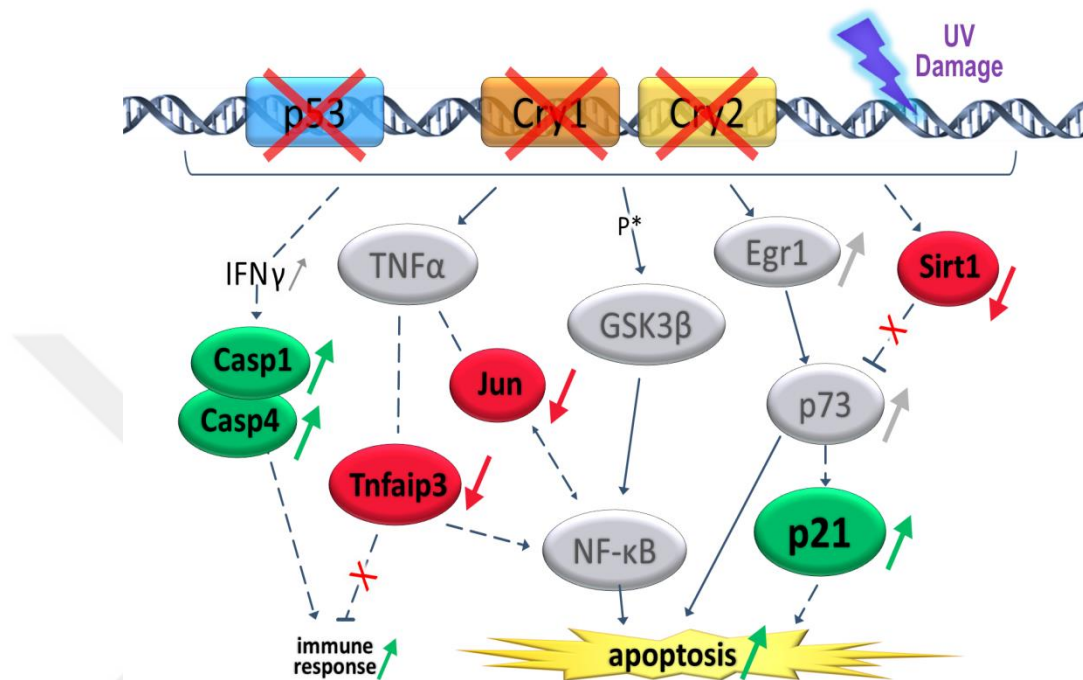


Figure 7.6: Proposed pathway model for involvement of novel candidate genes in the delayed onset of cancer in $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ mice compared to $p53^{-/-}$ mice. The proteins in grey have been identified in previous research to be involved in increased apoptotic levels after Cry deletions in $p53^{-/-}$ cells. The current RNA-seq analysis has drawn attention to the possible involvement of the genes in red and green, after UV treatment in $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ cells compared to $p53^{-/-}$ cells. These genes fit well into apoptotic pathways from previous knowledge. Of the dark blue lines between proteins, arrow-headed lines indicate induction, while bar-headed lines indicate inhibition. Undashed lines indicate findings from previous studies, dashed lines indicate interactions proposed by present analyses that are supported by known interactions from the literature (Fontemaggi et al. 2002; Jin et al. 2007; Lin et al. 2009; Ozturk et al. 2009; Hashiramoto et al. 2010; Lee and Sancar 2011a).

Chapter 8:

CONCLUSION

The presented research analyzed the effect of lncRNA, co-activators and oligomerization on transcription factors, and investigated the effects of deletion of a transcriptional repressor on cancer.

Through lncRNA pulldown and mass spectrometry, it was observed that a lncRNA can act as a scaffolding sponge to bring several co-activators of androgen receptor near the transcription site of target genes. The KLK3e lncRNA was observed to bind known Androgen Receptor (AR) co-activators such as DDX5, DDX17 and HSPA8, which can explain its role in upregulating prostate cancer associated genes. Such interactions between lncRNA and other transcription factors, and steroid receptors in particular have been previously implicated. Also, DDX5 has been shown to influence gene expression regulated by the estrogen receptor. These suggest that lncRNAs such as KLK3e that are transcribed from upstream enhancer/promoter regions relevant to transcription factor activity can bring together complexes of co-activators and steroid receptors, playing critical roles in steroid hormone-regulated healthy or pathogenic gene expression.

Inhibition of different sites on the classical steroid receptor AR, followed by RIME and RNA sequencing analyses demonstrated the impact each site has on interactions with co-activator proteins and on regulation of gene expression. Interestingly, inhibition of a protein interaction site appeared to have comparable effects to inhibition of ligand binding. Prevention of DNA binding and DBD dimerization also seemed to have a comparable impact on the protein interactions and transcriptional activity of the receptor, suggesting that allosteric changes upon DNA binding and dimerization may play key roles in mediating transcription factor activation.

As implicated by experiments, computational modeling of the interactions of glucocorticoid receptor (GR) with the mineralocorticoid receptor (MR) elucidated that GR and MR can interact through alternative interfaces. Also in accordance with experimental observations, GR and MR were predicted to interact when bound by

different ligands or even antagonists. Furthermore, computational predictions demonstrated that the ligand or antagonist bound alters interface preference, which can leave different sites available for binding of co-activators, explaining why different ligands can produce different cellular outcomes. Computational modeling of tetramerization of GR demonstrated alternative tetramerization interfaces, and showed that the DNA binding domain could not tetramerize on one strand, but rather, chromosomal looping could occur to form a dimer of dimers binding at two DNA loci.

Finally, deletion of the circadian transcriptional repressor cryptochrome in a p53 mutant background caused increases in apoptotic and anti-tumorigenic responses to DNA damage at the transcriptome level, explaining the observation that cryptochrome deletion in p53 mutant mice delays the onset of cancer. Protein interaction network analysis of differentially expressed genes pinpointed genes such as Sirt1, p21 and Jun that play key roles in alleviating tumorigenicity. Interestingly, differentially expressed transcripts also contained a high ratio of non-coding RNA, demonstrating they might be important in damage response and apoptosis regulation.

Collectively, these results demonstrated the effects of lncRNA, co-activator binding and oligomerization on transcription factor activity, and the impact that deletion of one transcription factor can have on cancer.

BIBLIOGRAPHY

- Altschul SF, Gish W, Miller W, et al (1990) Basic local alignment search tool. *J Mol Biol* 215:403–10. doi: 10.1016/S0022-2836(05)80360-2
- Andrews S (2014) FastQC A Quality Control tool for High Throughput Sequence Data
- Antoch MP, Gorbacheva VY, Vykhovanets O, et al (2008) Disruption of the circadian clock due to the Clock mutation has discrete effects on aging and carcinogenesis. *Cell Cycle* 7:1197–1204. doi: 10.4161/cc.7.9.5886
- Antonarakis ES, Chandhasin C, Osbourne E, et al (2016) Targeting the N-Terminal Domain of the Androgen Receptor: A New Approach for the Treatment of Advanced Prostate Cancer. *Oncologist* 21:1427–1435. doi: 10.1634/theoncologist.2016-0161
- Armanini D, Endres S, Kuhnle U, Weber PC (1988) Parallel determination of mineralocorticoid and glucocorticoid receptors in {T}- and {B}-lymphocytes of human spleen. *Acta Endocrinol* 118:479–482
- Arora VK, Schenkein E, Murali R, et al (2013) Glucocorticoid Receptor Confers Resistance to Antiandrogens by Bypassing Androgen Receptor Blockade. *Cell* 155:1309–1322. doi: <https://doi.org/10.1016/j.cell.2013.11.012>
- Arriza JL, Weinberger C, Cerelli G, et al (1987) Cloning of human mineralocorticoid receptor complementary DNA: Structural and functional kinship with the glucocorticoid receptor. *Science* (80-) 237:268–275. doi: 10.1126/science.3037703
- Attardi LD, Donehower LA (2005) Probing p53 biological functions through the use of genetically engineered mouse models. *Mutat Res Mol Mech Mutagen* 576:4–21. doi: <https://doi.org/10.1016/j.mrfmmm.2004.08.022>
- Attardi LD, Jacks T (1999) The role of p53 in tumour suppression: Lessons from mouse models. *Cell. Mol. Life Sci.* 55:48–63
- Bai Q, Bai Z, Sun L (2016) Detection of RNA-binding Proteins by In Vitro RNA Pull-down in Adipocyte Culture. *J Vis Exp* 1–7. doi: 10.3791/54207
- Baker KE, Parker R (2004) Nonsense-mediated mRNA decay: Terminating erroneous gene expression. *Curr. Opin. Cell Biol.* 16:293–299
- Baspinar A, Cukuroglu E, Nussinov R, et al (2014) PRISM: A web server and repository for prediction of protein-protein interactions and modeling their 3D complexes. *Nucleic Acids Res* 42.: doi: 10.1093/nar/gku397
- Bateman A (2019) UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res* 47:D506–D515. doi: 10.1093/nar/gky1049
- Beavan S, Horner A, Bord S, et al (2001) Colocalization of glucocorticoid and mineralocorticoid receptors in human bone. *J Bone Miner Res* 16:1496–1504. doi: 10.1359/jbmr.2001.16.8.1496
- Beer TM, Armstrong AJ, Rathkopf DE, et al (2014) Enzalutamide in Metastatic Prostate Cancer before Chemotherapy. *N Engl J Med* 371:424–433. doi: 10.1056/NEJMoa1405095
- Bergmann JH, Li J, Eckersley-Maslin MA, et al (2015) Regulation of the ESC transcriptome by nuclear long noncoding RNAs. *Genome Res* 25:1336–1346. doi: 10.1101/gr.189027.114
- Berman HM, Westbrook J, Feng Z, et al (2000) The Protein Data Bank. 28:235–242
- Bindea G, Mlecnik B, Hackl H, et al (2009) ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25:1091–1093. doi: 10.1093/bioinformatics/btp101
- Bledsoe RK, Montana VG, Stanley TB, et al (2002) Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. *Cell* 110:93–105. doi: 10.1016/S0092-8674(02)00817-6
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120

- Bray F, Ferlay J, Soerjomataram I, et al (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 68:394–424. doi: 10.3322/caac.21492
- Carithers LJ, Ardlie K, Barcus M, et al (2015) A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreserv Biobank* 13:311–319. doi: 10.1089/bio.2015.0032
- Cavga AD, Karahan N, Keskin O, Gursoy A (2015) Taming Oncogenic Signaling at Protein Interfaces: Challenges and Opportunities. *Curr Top Med Chem*
- Cavga AD, Tardu M, Korkmaz T, et al (2019) Cryptochrome deletion in p53 mutant mice enhances apoptotic and anti-tumorigenic responses to UV damage at the transcriptome level. *Funct Integr Genomics*. doi: 10.1007/s10142-019-00680-5
- Chan S, Debono M (2010) Review: Replication of cortisol circadian rhythm: New advances in hydrocortisone replacement therapy. *Ther. Adv. Endocrinol. Metab.* 1:129–138
- Chandrasekar T, Yang J, Gao A, Evans CP (2015) Mechanisms of resistance in castration-resistant prostate cancer (CRPC). *Transl Androl Urol* 4:365–80. doi: 10.3978/j.issn.2223-4683.2015.05.02
- Chang JC, Temple GF, Trecartin RF, Kan YW (1979) Suppression of the nonsense mutation in homozygous β 0 thalassaemia [31]. *Nature* 281:602–603
- Chen CD, Sawyers CL (2002) NF- B Activates Prostate-Specific Antigen Expression and Is Upregulated in Androgen-Independent Prostate Cancer. *Mol Cell Biol* 22:2862–2870. doi: 10.1128/mcb.22.8.2862-2870.2002
- Chu C, Zhang QC, Da Rocha ST, et al (2015) Systematic discovery of Xist RNA binding proteins. *Cell* 161:404–416. doi: 10.1016/j.cell.2015.03.025
- Claessens F, Joniau S, Helsen C (2017) Comparing the rules of engagement of androgen and glucocorticoid receptors. *Cell. Mol. Life Sci.* 74:2217–2228
- Clark EL, Coulson A, Dalgliesh C, et al (2008) The RNA helicase p68 is a novel androgen receptor coactivator involved in splicing and is overexpressed in prostate cancer. *Cancer Res* 68:7938–7946. doi: 10.1158/0008-5472.CAN-08-0932
- Clark EL, Hadjimichael C, Temperley R, et al (2013) p68/Ddx5 Supports β -Catenin & RNAP II during Androgen Receptor Mediated Transcription in Prostate Cancer. *PLoS One* 8.: doi: 10.1371/journal.pone.0054150
- Cleutjens KBJM, Van Eekelen CCEM, Van Der Korput HAGM, et al (1996) Two androgen response regions cooperate in steroid hormone regulated activity of the prostate-specific antigen promoter. *J Biol Chem* 271:6379–6388. doi: 10.1074/jbc.271.11.6379
- Coornaert B, Carpentier I, Beyaert R (2009) A20: Central gatekeeper in inflammation and immunity. *J. Biol. Chem.* 284:8217–8221
- Cukuroglu E, Gursoy A, Keskin O (2011) HotRegion: a database of predicted hot spot clusters. *Nucleic Acids Res* 40:D829–D833. doi: 10.1093/nar/gkr929
- Dahlman-Wright K, Wright A, Gustafsson JÅ, Carlstedt-Duke J (1991) Interaction of the glucocorticoid receptor DNA-binding domain with DNA as a dimer is mediated by a short segment of five amino acids. *J Biol Chem* 266:3107–3112
- Dalal K, Ban F, Li H, et al (2018) Selectively targeting the dimerization interface of human androgen receptor with small-molecules to treat castration-resistant prostate cancer. *Cancer Lett* 437:35–43. doi: 10.1016/j.canlet.2018.08.016
- Dalal K, Roshan-Moniri M, Sharma A, et al (2014) Selectively targeting the DNA-binding domain of the androgen receptor as a prospective therapy for prostate cancer. *J Biol Chem* 289:26417–26429. doi: 10.1074/jbc.M114.553818
- daSilva LF, Beckedorff FC, Ayupe AC, et al (2018) Chromatin Landscape Distinguishes the Genomic Loci of Hundreds of Androgen-Receptor-Associated LincRNAs From the Loci of Non-associated LincRNAs. *Front Genet* 9:132. doi: 10.3389/fgene.2018.00132
- Davey RA, Grossmann M (2016) Androgen Receptor Structure, Function and Biology: From Bench to Bedside. *Clin Biochem Rev* 37:3–15

- Davies TH, Ning YM, Sánchez ER (2002) A new first step in activation of steroid receptors. Hormone-induced switching of FKBP51 and FKBP52 immunophilins. *J Biol Chem*. doi: 10.1074/jbc.C100531200
- Deep G, Panigrahi GK (2016) Hypoxia-Induced Signaling Promotes Prostate Cancer Progression: Exosomes Role as Messenger of Hypoxic Response in Tumor Microenvironment. *Crit Rev Oncog* 20:419–434. doi: 10.1615/critrevoncog.v20.i5-6.130
- Dehm SM, Schmidt LJ, Heemers H V., et al (2008) Splicing of a novel androgen receptor exon generates a constitutively active androgen receptor that mediates prostate cancer therapy resistance. *Cancer Res* 68:5469–5477. doi: 10.1158/0008-5472.CAN-08-0594
- Derfoul A, Robertson NM, Lingrel JB, et al (1998) Regulation of the human Na/K-ATPase β 1 gene promoter by mineralocorticoid and glucocorticoid receptors. *J Biol Chem* 273:20702–20711. doi: 10.1074/jbc.273.33.20702
- Djulbegovic M, Beyth RJ, Neuberger MM, et al (2010) Screening for prostate cancer: systematic review and meta-analysis of randomised controlled trials. *BMJ* 341:
- Dobin A, Davis CA, Schlesinger F, et al (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21
- Donehower LA, Harvey M, Slagle BL, et al (1992) Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* 356:215–221. doi: 10.1038/356215a0
- Dong J, Wu Z, Wang D, et al (2019) Hsp70 Binds to the Androgen Receptor N-terminal Domain and Modulates the Receptor Function in Prostate Cancer Cells. *Mol Cancer Ther* 18:39–50. doi: 10.1158/1535-7163.mct-18-0432
- Dorselaer A Van, Clerc P, Avner P, et al (2010) 2-D Structure of the A Region of Xist RNA and Its Implication for PRC2 Association. 8:1–16. doi: 10.1371/journal.pbio.1000276
- Duhovny D, Nussinov R, Wolfson HJ (2002) Efficient Unbound Docking of Rigid Molecules. pp 185–200
- Engin HB, Keskin O, Nussinov R, Gursoy A (2012) A strategy based on protein-protein interface motifs may help in identifying drug off-targets. *J Chem Inf Model* 52:2273–2286. doi: 10.1021/ci300072q
- Evans RM (1988) The steroid and thyroid hormone receptor superfamily. *Science* (80-) 240:889–895. doi: 10.1126/science.3283939
- Farman N, Oblin ME, Lombes M, et al (2017) Immunolocalization of gluco- and mineralocorticoid receptors in rabbit kidney. *Am J Physiol Physiol* 260:C226–C233. doi: 10.1152/ajpcell.1991.260.2.c226
- Fawell SE, Lees JA, White R, Parker MG (1990) Characterization and colocalization of steroid binding and dimerization activities in the mouse estrogen receptor. *Cell* 60:953–962. doi: 10.1016/0092-8674(90)90343-D
- Filipski E, King VM, Li X, et al (2002) Host circadian clock as a control point in tumor progression. *J Natl Cancer Inst* 94:690–697. doi: 10.1093/jnci/94.9.690
- Fontemaggi G, Kela I, Amariglio N, et al (2002) Identification of direct p73 target genes combining DNA microarray and chromatin immunoprecipitation analyses. *J Biol Chem* 277:43359–43368. doi: 10.1074/jbc.M205573200
- Frankish A, Bignell A, Berry A, et al (2018) GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* 47:D766–D773. doi: 10.1093/nar/gky955
- Freeman LC (1977) A Set of Measures of Centrality Based on Betweenness. *Sociometry* 40:35–41
- Froesch BA, Takayama S, Reed JC (1998) BAG-1L protein enhances androgen receptor function. *J Biol Chem* 273:11660–11666. doi: 10.1074/jbc.273.19.11660
- Fu L, Pelicano H, Liu J, et al (2002) The circadian gene Period2 plays an important role in tumor suppression and DNA damage response in vivo. *Cell* 111:41–50. doi: 10.1016/S0092-8674(02)00961-3
- Fuller PJ, Lim-Tio SS, Brennan FE (2000) Specificity in mineralocorticoid versus

- glucocorticoid action. In: *Kidney International*. pp 1256–1264
- Gabay M, Li Y, Felsher DW (2014) MYC activation is a hallmark of cancer initiation and maintenance. *Cold Spring Harb Perspect Med* 4.: doi: 10.1101/cshperspect.a014241
- Gao W, Bohl CE, Dalton JT, et al (2013) Targeting the Binding Function 3 (BF3) Site of the Androgen Receptor Through Virtual Screening. 2. Development of 2-((2-phenoxyethyl)thio)-1H-benzimidazole derivatives. *J Med Chem* 55:1136–48. doi: 10.1021/jm3015712
- Gauger MA, Sancar A (2005) Cryptochrome, circadian cycle, cell cycle checkpoints, and cancer. *Cancer Res* 65:6828–6834. doi: 10.1158/0008-5472.CAN-05-1119
- Gekakis N, Staknis D, Nguyen HB, et al (1998) Role of the CLOCK Protein in the Mammalian Circadian Mechanism. *Science* (80-) 280:1564–1569 %U <http://www.sciencemag.org/content/280>. doi: 10.1126/science.280.5369.1564
- Girling JS, Whitaker HC, Mills IG, Neal DE (2007) Pathogenesis of prostate cancer and hormone refractory prostate cancer. *Indian J Urol* 23:35–42. doi: 10.4103/0970-1591.30265
- Gomez-Sanchez E, Gomez-Sanchez CE (2014) The multifaceted mineralocorticoid receptor. *Compr Physiol* 4:965–994. doi: 10.1002/cphy.c130044
- Grunewald TGP, Bach H, Cossarizza A, Matsumoto I (2012) The STEAP protein family: Versatile oxidoreductases and targets for cancer immunotherapy with overlapping and distinct cellular functions. *Biol Cell* 104:641–657. doi: 10.1111/boc.201200027
- Guo Z, Yang X, Sun F, et al (2009) A novel androgen receptor splice variant is up-regulated during prostate cancer progression and promotes androgen depletion-resistant growth. *Cancer Res* 69:2305–2313. doi: 10.1158/0008-5472.CAN-08-3795
- Guryanova OA, Lieu YK, Garrett-Bakelman FE, et al (2016) Dnmt3a regulates myeloproliferation and liver-specific expansion of hematopoietic stem and progenitor cells. *Leukemia* 30:1133–1142. doi: 10.1038/leu.2015.358
- Guttman M, Donaghey J, Carey BW, et al (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477:295–300. doi: 10.1038/nature10398
- Han F, Ozawa H, Matsuda KI, et al (2005) Colocalization of mineralocorticoid receptor and glucocorticoid receptor in the hippocampus and hypothalamus. *Neurosci Res* 51:371–381. doi: 10.1016/j.neures.2004.12.013
- Harms K, Nozell S, Chen X (2004) The common and distinct target genes of the p53 family transcription factors. *Cell Mol Life Sci* 61:822–842. doi: 10.1007/s00018-003-3304-4
- Hasegawa Y, Brockdorff N, Kawano S, et al (2010) The matrix protein hnRNP U is required for chromosomal localization of xist RNA. *Dev Cell* 19:469–476. doi: 10.1016/j.devcel.2010.08.006
- Hashiramoto A, Yamane T, Tsumiyama K, et al (2010) Mammalian Clock Gene Cryptochrome Regulates Arthritis via Proinflammatory Cytokine TNF- . *J Immunol* 184:1560–1565. doi: 10.4049/jimmunol.0903284
- Heberle H, Meirelles VG, da Silva FR, et al (2015) InteractiVenn: A web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* 16.: doi: 10.1186/s12859-015-0611-3
- Herriges MJ, Swarr DT, Morley MP, et al (2014) Long noncoding RNAs are spatially correlated with transcription factors and regulate lung development. *Genes Dev* 28:1363–1379. doi: 10.1101/gad.238782.114
- Hirst MA, Hinck L, Danielsen M, Ringold GM (2006) Discrimination of DNA response elements for thyroid hormone and estrogen is dependent on dimerization of receptor DNA binding domains. *Proc Natl Acad Sci* 89:5527–5531. doi: 10.1073/pnas.89.12.5527
- Hofmann JW, Zhao X, De Cecco M, et al (2015) Reduced Expression of MYC Increases Longevity and Enhances Healthspan. *Cell* 160:477–488. doi: <http://dx.doi.org/10.1016/j.cell.2014.12.016>
- Hsieh C-L, Fei T, Chen Y, et al (2014) Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proc Natl Acad Sci U S A* 111:7319–24.

- doi: 10.1073/pnas.1324151111
- Huang KH, Chow KC, Chang HW, et al (2011) ATPase family AAA domain containing 3A is an anti-apoptotic factor and a secretion regulator of PSA in prostate cancer. *Int J Mol Med* 28:9–15. doi: 10.3892/ijmm.2011.670
- Hudson WH, Youn C, Ortlund EA (2014) Crystal structure of the mineralocorticoid receptor DNA binding domain in complex with DNA. *PLoS One* 9.: doi: 10.1371/journal.pone.0107000
- Ilic D, Djulbegovic M, Jung JH, et al (2018) Prostate cancer screening with prostate-specific antigen (PSA) test: A systematic review and meta-analysis. *BMJ* 362.: doi: 10.1136/bmj.k3519
- Jensen LJ, Kuhn M, Stark M, et al (2009) STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* 37 :D412–D416. doi: 10.1093/nar/gkn760
- Jensen LJ, Kuhn M, Stark M, et al (2018) 487 items (mouse) - STRING interaction network. <https://version-10-5.string-db.org/cgi/network.pl?networkId=CeX9QRITi9g5>. Accessed 19 Oct 2018
- Jenster G, van der Korput HAGM, van Vroonhoven C, et al (1991) Domains of the Human Androgen Receptor Involved in Steroid Binding, Transcriptional Activation, and Subcellular Localization. *Mol Endocrinol* 5:1396–1404. doi: 10.1210/mend-5-10-1396
- Jin MD, Zhi YW, Dao CS, et al (2007) SIRT1 interacts with p73 and suppresses p73-dependent transcriptional activity. *J Cell Physiol* 210:161–166. doi: 10.1002/jcp.20831
- Jochum T, Ritz ME, Schuster C, et al (2012) Toxic and non-toxic aggregates from the SBMA and normal forms of androgen receptor have distinct oligomeric structures. *Biochim Biophys Acta - Mol Basis Dis* 1822:1070–1078. doi: 10.1016/j.bbadis.2012.02.006
- Johnson P (1989) Eukaryotic Transcriptional Regulatory Proteins. *Annu Rev Biochem* 58:799–839. doi: 10.1146/annurev.biochem.58.1.799
- Kadmiel M, Cidlowski JA (2013) Glucocorticoid receptor signaling in health and disease. *Trends Pharmacol. Sci.* 34:518–530
- Karpievitch Y V., Dabney AR, Smith RD (2012) Normalization and missing value imputation for label-free LC-MS analysis. *BMC Bioinformatics* 13 Suppl 1.: doi: 10.1186/1471-2105-13-S16-S5
- Kauppi B, Jakob C, Färnegårdh M, et al (2003) The Three-dimensional Structures of Antagonistic and Agonistic Forms of the Glucocorticoid Receptor Ligand-binding Domain. *J Biol Chem* 278:22748–22754. doi: 10.1074/jbc.m212711200
- Kavakli IH, Baris I, Tardu M, et al (2017) The Photolyase/Cryptochrome Family of Proteins as DNA Repair Enzymes and Transcriptional Repressors. *Photochem. Photobiol.* 93:93–103
- Keeton EK, Wang Q, Li W, et al (2007) A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. *Mol Cell* 27:380–392. doi: 10.1016/j.molcel.2007.05.041
- Kim H-J, Park YI, Dong M-S (2006) Comparison of prostate cancer cell lines for androgen receptor-mediated reporter gene assays. *Toxicol Vitro* 20:1159–1167. doi: <https://doi.org/10.1016/j.tiv.2006.03.003>
- Kita K, Shiota M, Tanaka M, et al (2017) Heat shock protein 70 inhibitors suppress androgen receptor expression in LNCaP95 prostate cancer cells. *Cancer Sci* 108:1820–1827. doi: 10.1111/cas.13318
- Korkmaz, Aygenli F, Emisoglu H, et al (2018) Opposite Carcinogenic Effects of Circadian Clock Gene BMAL1. *Sci Rep*. doi: 10.1038/s41598-018-34433-410.1038/s41598-018-34433-4
- Korkmaz CG, Korkmaz KS, Kurys P, et al (2005) Molecular cloning and characterization of STAMP2, an androgen-regulated six transmembrane protein that is overexpressed in prostate cancer. *Oncogene* 24:4934–4945. doi: 10.1038/sj.onc.1208677
- Kozakov D, Hall DR, Xia B, et al (2017) The ClusPro web server for protein-protein docking.

- Nat Protoc 12:255–278. doi: 10.1038/nprot.2016.169
- Krieger E, Joo K, Lee J, et al (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins Struct. Funct. Bioinforma.* 77:114–122
- Kumar A, Mikolajczyk SD, Goel AS, et al (1997) Expression of pro form of prostate-specific antigen by mammalian cells and its conversion to mature, active form by human kallikrein 2. *Cancer Res* 57:3111–3114
- Kume K, Zylka MJ, Sriram S, et al (1999) mCRY1 and mCRY2 are essential components of the negative limb of the circadian clock feedback loop. *Cell* 98:193–205. doi: 10.1016/S0092-8674(00)81014-4
- Lack NA, Axerio-Cilies P, Tavassoli P, et al (2011) Targeting the binding function 3 (BF3) site of the human androgen receptor through virtual screening. *J Med Chem* 54:8563–8573. doi: 10.1021/jm201098n
- Lambert SA, Jolma A, Campitelli LF, et al (2018) The Human Transcription Factors. *Cell* 172:650–665
- Lang F, Shumilina E (2013) Regulation of ion channels by the serum- and glucocorticoid-inducible kinase SGK1. *FASEB J.* 27:3–12
- Laskowski RA, Hutchinson EG, Michie AD, et al (1997) PDBsum: A Web-based database of summaries and analyses of all PDB structures. *Trends Biochem. Sci.* 22:488–490
- Latchman DS (1997) Transcription factors: An overview. *Int J Biochem Cell Biol* 29:1305–1312. doi: 10.1016/S1357-2725(97)00085-X
- Lee EG, Boone DL, Chai S, et al (2000) Failure to Regulate TNF-Induced NF- κ B and Cell Death Responses in A20-Deficient Mice. *Science* (80-) 289:2350 LP – 2354
- Lee JH, Gaddameedhi S, Ozturk N, et al (2013) DNA damage-specific control of cell death by cryptochrome in p53-mutant ras-transformed cells. *Cancer Res* 73:785–791. doi: 10.1158/0008-5472.CAN-12-1994
- Lee JH, Sancar A (2011a) Circadian clock disruption improves the efficacy of chemotherapy through p73-mediated apoptosis. *Proc Natl Acad Sci* 108:10668–10672. doi: 10.1073/pnas.1106284108
- Lee JH, Sancar A (2011b) Regulation of apoptosis by the circadian clock through NF- κ B signaling. *Proc Natl Acad Sci* 108:12036–12041. doi: 10.1073/pnas.1108125108/-/DCSupplemental. www.pnas.org/cgi/doi/10.1073/pnas.1108125108
- Lee TI, Young RA (2013) Transcriptional regulation and its misregulation in disease. *Cell* 152:1237–1251
- Lensink MF, Velankar S, Wodak SJ (2017) Modeling protein–protein and protein–peptide complexes: CAPRI 6th edition. *Proteins Struct Funct Bioinforma* 85:359–377. doi: 10.1002/prot.25215
- Li H, Ban F, Dalal K, et al (2014a) Discovery of small-molecule inhibitors selectively targeting the DNA-binding domain of the human androgen receptor. *J Med Chem* 57:6458–6467. doi: 10.1021/jm500802j
- Li X, Brock GN, Rouchka EC, et al (2017) A comparison of per sample global scaling and per gene normalization methods for differential expression analysis of RNA-seq data. *PLoS One* 12:e0176185. doi: 10.1371/journal.pone.0176185
- Li YF, Wang QZ, Zhang TT, et al (2014b) Low dose of interferon- α improves the clinical outcomes of docetaxel in patients with castration-resistant prostate cancer: A pilot study. *Oncol Lett* 7:125–130. doi: 10.3892/ol.2013.1653
- Lilja H (2003) Biology of prostate-specific antigen. *Urology* 62:27–33. doi: 10.1016/S0090-4295(03)00775-1
- Lin YL, Sengupta S, Gurdziel K, et al (2009) p63 and p73 transcriptionally regulate genes involved in DNA repair. *PLoS Genet* 5:1–13. doi: 10.1371/journal.pgen.1000680
- Liu C, Wu HT, Zhu N, et al (2016) Steroid receptor RNA activator: Biologic function and role in disease. *Clin. Chim. Acta* 459:137–146

- Liu W, Wang J, Sauter NK, Pearce D (1995) Steroid receptor heterodimerization demonstrated in vitro and in vivo. *Proc Natl Acad Sci* 92:12480–12484. doi: 10.1073/pnas.92.26.12480
- Liu W, Wang J, Yu G, Pearce D (1996) Steroid receptor transcriptional synergy is potentiated by disruption of the {DNA}-binding domain dimer interface. *Mol Endocrinol* 10:1399–1406
- Livak KJ, Schmittgen TD (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2- $\Delta\Delta$ CT Method. *Methods* 25:402–408. doi: <http://dx.doi.org/10.1006/meth.2001.1262>
- Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550. doi: 10.1186/s13059-014-0550-8
- Lowrey PL, Takahashi JS (2004) MAMMALIAN CIRCADIAN BIOLOGY: Elucidating Genome-Wide Levels of Temporal Organization. *Annu Rev Genomics Hum Genet* 5:407–441. doi: 10.1146/annurev.genom.5.061903.175925
- Lucafo M, Iudicibus S, Silvestre A, et al (2015) Long Noncoding RNA GAS5: A Novel Marker Involved in Glucocorticoid Response. *Curr Mol Med* 15:94–99. doi: 10.2174/1566524015666150114122354
- Luo J, Nikolaev AY, Imai S ichiro, et al (2001) Negative control of p53 by Sir2 α promotes cell survival under stress. *Cell* 107:137–148. doi: 10.1016/S0092-8674(01)00524-4
- Luy B, Rottenberg JT, Al-Lazikani B, et al (2017) Development of Bag-1L as a therapeutic target in androgen receptor-dependent prostate cancer. *Elife* 6.: doi: 10.7554/elife.27159
- Ma A, Malynn B a (2012) A20: linking a complex regulator of ubiquitylation to immunity and human disease. *Nat Rev Immunol* 12:774–85. doi: 10.1038/nri3313
- Magee JA, Chang LW, Stormo GD, Milbrandt J (2006) Direct, androgen receptor-mediated regulation of the FKBP5 gene via a distal enhancer element. *Endocrinology*. doi: 10.1210/en.2005-1001
- Marignol L, Coffey M, Lawler M, Hollywood D (2008) Hypoxia in prostate cancer: A powerful shield against tumour destruction? *Cancer Treat. Rev.* 34:313–327
- Marzolla V, Armani A, Zennaro MC, et al (2012) The role of the mineralocorticoid receptor in adipocyte biology and fat metabolism. *Mol. Cell. Endocrinol.* 350:281–288
- Mazaira GI, Zgajnar NR, Lotufo CM, et al (2018) The Nuclear Receptor Field: A Historical Overview and Future Challenges. *Nucl Recept Res* 5.: doi: 10.11131/2018/101320
- McHugh CA, Chen C-K, Chow A, et al (2015) The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* 521:232–236. doi: 10.1038/nature14443
- Mifsud KR, Reul JM (2016) Acute stress enhances heterodimerization and binding of corticosteroid receptors at glucocorticoid target genes in the hippocampus. *Proc Natl Acad Sci* 113:11336–11341. doi: 10.1073/pnas.1605246113
- Mohammed H, Russell IA, Stark R, et al (2015) Progesterone receptor modulates ER α action in breast cancer. *Nature* 523:313–317. doi: 10.1038/nature14583
- Mohammed H, Taylor C, Brown GD, et al (2016) Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. *Nat Protoc* 11:316
- Moldes M, Lasnier F, Gauthereau X, et al (2001) Tumor Necrosis Factor- α -induced Adipose-related Protein (TIARP), a Cell-surface Protein That Is Highly Induced by Tumor Necrosis Factor- α and Adipose Conversion. *J Biol Chem* 276:33938–33946. doi: 10.1074/jbc.M105726200
- Monks DA, Rao P, Mo K, et al (2008) Androgen receptor and Kennedy disease/spinal bulbar muscular atrophy. *Horm Behav* 53:729–740. doi: 10.1016/j.yhbeh.2007.12.009
- Mootha VK, Lindgren CM, Eriksson K-F, et al (2003) PGC-1[α]-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 34:267–273
- Moses MA, Kim YS, Rivera-Marquez GM, et al (2018) Targeting the Hsp40/Hsp70 chaperone axis as a novel strategy to treat castration-resistant prostate cancer. *Cancer Res* 78:4022–

4035. doi: 10.1158/0008-5472.CAN-17-3728
- Narayanan S, Srinivas S, Feldman D (2016) Androgen-glucocorticoid interactions in the era of novel prostate cancer therapy. *Nat. Rev. Urol.* 13:47–60
- Nazareth L V., Weigel NL (1996) Activation of the human androgen receptor through a protein kinase A signaling pathway. *J Biol Chem* 271:19900–19907. doi: 10.1074/jbc.271.33.19900
- Nelson PS, Clegg N, Arnold H, et al (2002) The program of androgen-responsive genes in neoplastic prostate epithelium. *Proc Natl Acad Sci.* doi: 10.1073/pnas.182376299
- Network TCGAR, Chang K, Creighton CJ, et al (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 45:1113
- Ni L, Llewellyn R, Kesler CT, et al (2013) Androgen Induces a Switch from Cytoplasmic Retention to Nuclear Import of the Androgen Receptor. *Mol Cell Biol* 33:4766–4778. doi: 10.1128/MCB.00647-13
- Nishi M (2004) Visualization of Glucocorticoid Receptor and Mineralocorticoid Receptor Interactions in Living Cells with GFP-Based Fluorescence Resonance Energy Transfer. *J Neurosci* 24:4918–4927. doi: 10.1523/jneurosci.5495-03.2004
- Oishi K, Fukui H, Ishida N (2000) Rhythmic Expression of BMAL1 mRNA Is Altered in Clock Mutant Mice: Differential Regulation in the Suprachiasmatic Nucleus and Peripheral Tissues. *Biochem Biophys Res Commun* 268:164–171. doi: 10.1006/bbrc.1999.2054
- Olefsky JM (2001) Nuclear Receptor Minireview Series. *J. Biol. Chem.* 276:36863–36864
- Ou XM, Storrington JM, Kushwaha N, Albert PR (2001) Heterodimerization of Mineralocorticoid and Glucocorticoid Receptors at a Novel Negative Response Element of the 5-HT1A Receptor Gene. *J Biol Chem* 276:14299–14307. doi: 10.1074/jbc.M005363200
- Ozturk N, Lee JH, Gaddameedhi S, Sancar A (2009) Loss of cryptochrome reduces cancer risk in p53 mutant mice. *Proc Natl Acad Sci U S A* 106:2841–2846. doi: 10.1073/pnas.0813028106
- Palusa S, Reddy A (2013) Analysis of RNA-protein Interactions Using Electrophoretic Mobility Shift Assay (Gel Shift Assay). *BIO-PROTOCOL* 3.: doi: 10.21769/bioprotoc.967
- Panda AC, Martindale JL, Gorospe M (2016) Affinity Pulldown of Biotinylated RNA for Detection of Protein-RNA Complexes. *Bio-protocol* 6:e2062. doi: 10.21769/BioProtoc.2062
- Panda S, Hogenesch JB, Kay SA (2002) Circadian rhythms from flies to human. *Nature* 417:329–335
- Pandiyani P, Younes SA, Ribeiro SP, et al (2016) Mucosal regulatory T cells and T helper 17 cells in HIV-associated immune activation. *Front Immunol* 7.: doi: 10.3389/fimmu.2016.00228
- Pearce D, Yamamoto KR (1993) Mineralocorticoid and glucocorticoid receptor activities distinguished by nonreceptor factors at a composite response element. *Science* (80-) 259:1161–1165. doi: 10.1126/science.8382376
- Penny GD, Kay GF, Sheardown SA, et al (1996) Requirement for Xist in X chromosome inactivation. *Nature* 379:131–137. doi: 10.1038/379131a0
- Penson DF, Armstrong AJ, Concepcion R, et al (2016) Enzalutamide versus bicalutamide in castration-resistant prostate cancer: The STRIVE trial. *J Clin Oncol* 34:2098–2106. doi: 10.1200/JCO.2015.64.9285
- Pintacuda G, Wei G, Roustan C, et al (2017) hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. *Mol Cell* 68:955–969.e10. doi: 10.1016/j.molcel.2017.11.013
- Pinter SF, Sadreyev RI, Yildirim E, et al (2012) Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations. 1864–1876. doi: 10.1101/gr.133751.111
- Planey SL, Derfoul A, Steplewski A, et al (2002) Inhibition of glucocorticoid-induced apoptosis in 697 pre-B lymphocytes by the mineralocorticoid receptor N-terminal domain. *J Biol*

- Chem 277:42188–42196. doi: 10.1074/jbc.M205085200
- Popenda M, Szachniuk M, Antczak M, et al (2012) Automated 3D structure composition for large RNAs. *Nucleic Acids Res* 40.: doi: 10.1093/nar/gks339
- Presman DM, Ganguly S, Schiltz RL, et al (2016) DNA binding triggers tetramerization of the glucocorticoid receptor in live cells. *Proc Natl Acad Sci* 113:8236–8241. doi: 10.1073/pnas.1606774113
- Pu H, Horbinski C, Hensley PJ, et al (2014) PARP-1 regulates epithelial-mesenchymal transition (EMT) in prostate tumorigenesis. *Carcinogenesis* 35:2592–2601. doi: 10.1093/carcin/bgu183
- Rapaport F, Khanin R, Liang Y, et al (2013) Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol* 14.: doi: 10.1186/gb-2013-14-9-r95
- Ravagnan L, Gurbuxani S, Susin SA, et al (2001) Heat-shock protein 70 antagonizes apoptosis-inducing factor. *Nat Cell Biol* 3:839–843. doi: 10.1038/ncb0901-839
- Reményi A, Schöler HR, Wilmanns M (2004) Combinatorial control of gene expression. *Nat. Struct. Mol. Biol.* 11:812–815
- Reul JMHM, De Kloet ER (1985) Two receptor systems for corticosterone in rat brain: Microdistribution and differential occupation. *Endocrinology* 117:2505–2511. doi: 10.1210/endo-117-6-2505
- Rinn JL, Kertesz M, Wang JK, et al (2007) Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell* 129:1311–1323. doi: 10.1016/j.cell.2007.05.022
- Rossow KL, Janknecht R (2003) Synergism between p68 RNA helicase and the transcriptional coactivators CBP and p300. *Oncogene* 22:151–156. doi: 10.1038/sj.onc.1206067
- Roth A, Diederichs S (2015) Rap and chirp about X inactivation. *Nature* 521:170–171. doi: 10.1038/521170a
- Rother M, Rother K, Puton T, Bujnicki JM (2011) ModeRNA: A tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res* 39:4007–4022. doi: 10.1093/nar/gkq1320
- Rowlatt C, Chesterman FC, Sheriff MU (1976) Lifespan, age changes and tumour incidence in an ageing C57BL mouse colony. *Lab Anim* 10:419–42. doi: 10.1258/002367776780956917
- Salehi S, Taheri MN, Azarpira N, et al (2017) State of the art technologies to explore long non-coding RNAs in cancer. *J Cell Mol Med* 21:3120–3140. doi: 10.1111/jcmm.13238
- Samaan S, Tranchevent LC, Dardenne E, et al (2014) The Ddx5 and Ddx17 RNA helicases are cornerstones in the complex regulatory array of steroid hormone-signaling pathways. *Nucleic Acids Res* 42:2197–2207. doi: 10.1093/nar/gkt1216
- Sarayloo E, Tardu M, Unlu YS, et al (2017) Understanding lipid metabolism in high-lipid-producing *Chlorella vulgaris* mutants at the genome-wide level. *Algal Res* 28:244–252. doi: 10.1016/j.algal.2017.11.009
- Savory JGA, Prefontaine GG, Lamprecht C, et al (2001) Glucocorticoid Receptor Homodimers and Glucocorticoid-Mineralocorticoid Receptor Heterodimers Form in the Cytoplasm through Alternative Dimerization Interfaces. *Mol Cell Biol* 21:781–793. doi: 10.1128/mcb.21.3.781-793.2001
- Scardoni G, Laudanna C (2012) Centralities Based Analysis of Complex Networks. *New Front graph thory* 323–348. doi: 10.5772/35846
- Scher HHHHHH, Beer TM, Higano CSCS, et al (2010) Antitumour activity of MDV3100 in castration-resistant prostate cancer: a phase 1-2 study. *Lancet* 375:1437–1446. doi: 10.1016/S0140-6736(10)60172-9. Antitumour
- Schiewer MJ, Goodwin JF, Han S, et al (2012) Dual roles of PARP-1 promote cancer growth and progression. *Cancer Discov* 2:1134–1149. doi: 10.1158/2159-8290.CD-12-0120
- Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2005) PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Res* 33.: doi: 10.1093/nar/gki481

- Schoenborn JR, Wilson CB (2007) Regulation of Interferon- γ During Innate and Adaptive Immune Responses. In: Immunology BT-A in (ed). Academic Press, pp 41–101
- Shanmugam I, Cheng G, Terranova PF, et al (2007) Serum/glucocorticoid-induced protein kinase-1 facilitates androgen receptor-dependent cell survival. *Cell Death Differ* 14:2085–2094. doi: 10.1038/sj.cdd.4402227
- Shannon P, Markiel A, Ozier O, et al (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* 13:2498–2504. doi: 10.1101/gr.1239303
- Shearman LP (2000) Interacting Molecular Loops in the Mammalian Circadian Clock. *Science* (80-) 288:1013–1019. doi: 10.1126/science.288.5468.1013
- Singh H, Khan AA, Dinner AR (2014) Gene regulatory networks in the immune system. *Trends Immunol.* 35:211–218
- Smedley D, Haider S, Durinck S, et al (2015) The BioMart community portal: An innovative alternative to large, centralized data repositories. *Nucleic Acids Res* 43:W589–W598. doi: 10.1093/nar/gkv350
- Smith CL, Blake JA, Kadin JA, et al (2018) Mouse Genome Database (MGD)-2018: Knowledgebase for the laboratory mouse. *Nucleic Acids Res* 46:D836–D842. doi: 10.1093/nar/gkx1006
- Smyth GK, Shi W, Liao Y (2013) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30:923–930. doi: 10.1093/bioinformatics/btt656
- Sollberger G, Strittmatter GE, Garstkiewicz M, et al (2014) Caspase-1: The inflammasome and beyond. *Innate Immun* 20:115–125. doi: 10.1177/1753425913484374
- Srinivasan S, Kumar R, Koduru S, et al (2010) Inhibiting TNF-mediated signaling: A novel therapeutic paradigm for androgen independent prostate cancer. *Apoptosis* 15:153–161. doi: 10.1007/s10495-009-0416-9
- Stechschulte LA, Sanchez ER (2011) FKBP51 - A selective modulator of glucocorticoid and androgen sensitivity. *Curr. Opin. Pharmacol.* 11:332–337
- Stelloo S, Nevedomskaya E, Kim Y, et al (2018) Endogenous androgen receptor proteomic profiling reveals genomic subcomplex involved in prostate tumorigenesis. *Oncogene* 37:313–322. doi: 10.1038/onc.2017.330
- Subramanian A, Tamayo P, Mootha VK, et al (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* 102:15545–15550. doi: 10.1073/pnas.0506580102
- Sun S, Sprenger CCT, Vessella RL, et al (2010) Castration resistance in human prostate cancer is conferred by a frequently occurring androgen receptor splice variant. *J Clin Invest* 120:2715–2730. doi: 10.1172/JCI41824
- Takahashi JS, Hong H-K, Ko CH, McDearmon EL (2008) The genetics of mammalian circadian order and disorder: implications for physiology and disease. *Nat Rev Genet* 9:764–775. doi: 10.1038/nrg2430
- Takayama TK, Fujikawa K, Davie EW (1997) Characterization of the precursor of prostate-specific antigen. Activation by trypsin and by human glandular kallikrein. *J Biol Chem* 272:21582–21588. doi: 10.1074/jbc.272.34.21582
- Tan MHE, Li J, Xu HE, et al (2014) Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacol Sin* 36:3
- Taplin M-E, Bubley GJ, Shuster TD, et al (2002) Mutation of the Androgen-Receptor Gene in Metastatic Androgen-Independent Prostate Cancer. *N Engl J Med* 332:1393–1398. doi: 10.1056/nejm199505253322101
- Tardu M, Bulut S, Kavakli IH (2017) MerR and ChrR mediate blue light induced photo-oxidative stress response at the transcriptional level in *Vibrio cholerae*. *Sci Rep* 7:40817. doi: 10.1038/srep40817
- Tardu M, Dikbas UM, Baris I, Kavakli IH (2016) RNA-seq analysis of the transcriptional

- response to blue and red light in the extremophilic red alga, *Cyanidioschyzon merolae*. *Funct Integr Genomics* 16:657–669. doi: 10.1007/s10142-016-0521-0
- Trachtenberg J, Blackledge G (2002) Looking to the future: Advances in the management of hormone-refractory prostate cancer. *Eur Urol Suppl* 1:44–53. doi: 10.1016/S1569-9056(02)00087-8
- Tran C, Ouk S, Clegg NJ, et al (2009) Development of a Second-Generation Antiandrogen for Treatment of Advanced Prostate Cancer. *Science* (80-) 324:787 LP – 790. doi: 10.1126/science.1168175
- Trapnell C, Roberts A, Goff L, et al (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7:562–78. doi: 10.1038/nprot.2012.016
- Trapp T, Rupprecht R, Castren M, et al (1994) Heterodimerization between mineralocorticoid and glucocorticoid receptor: A new principle of glucocorticoid action in the CNS. *Neuron* 13:1457–1462. doi: 10.1016/0896-6273(94)90431-6
- Tuncbag N, Gursoy A, Nussinov R, Keskin O (2011) Predicting protein-protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM. *Nat Protoc* 6:1341–1354
- Tuncbag N, Kar G, Gursoy A, et al (2009) Towards inferring time dimensionality in protein-protein interaction networks by integrating structures: the p53 example. *Mol Biosyst* 5:1770–8. doi: 10.1039/B905661K
- Tyanova S, Cox J (2018) Perseus: A bioinformatics platform for integrative analysis of proteomics data in cancer research. *Methods Mol Biol* 1711:133–148. doi: 10.1007/978-1-4939-7493-1_7
- Tyanova S, Temu T, Sinitcyn P, et al (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* 13:731–740
- Uhlén M, Fagerberg L, Hallström BM, et al (2015) Tissue-based map of the human proteome. *Science* (80-) 347.: doi: 10.1126/science.1260419
- Uhlen M, Oksvold P, Fagerberg L, et al (2010) Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* 28:1248–1250
- van Kempen LCL, Coussens LM (2016) MMP9 potentiates pulmonary metastasis formation. *Cancer Cell* 2:251–252. doi: 10.1016/S1535-6108(02)00157-5
- Velonas VM, Woo HH, dos Remedios CG, Assinder SJ (2013) Current status of biomarkers for prostate cancer. *Int. J. Mol. Sci.* 14:11034–11060
- Vitaterna MH, Selby CP, Todo T, et al (1999) Differential regulation of mammalian Period genes and circadian rhythmicity by cryptochromes 1 and 2. *Proc Natl Acad Sci* 96:12114–12119. doi: 10.1073/pnas.96.21.12114
- Wade Harper J, Adami GR, Wei N, et al (1993) The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell* 75:805–816. doi: [http://dx.doi.org/10.1016/0092-8674\(93\)90499-G](http://dx.doi.org/10.1016/0092-8674(93)90499-G)
- Wagner M, Rid R, Maier CJ, et al (2012) DDX5 is a multifunctional co-activator of steroid hormone receptors. *Mol Cell Endocrinol* 361:80–91. doi: 10.1016/j.mce.2012.03.014
- Wahli W, Martinez E (2018) Superfamily of steroid nuclear receptors: positive and negative regulators of gene expression. *FASEB J* 5:2243–2249. doi: 10.1096/fasebj.5.9.1860615
- Wajant H, Pfizenmaier K, Scheurich P Tumor necrosis factor signaling. *Cell Death Differ* 10:45–65
- Wang J, Mager J, Chen Y, et al (2001) Imprinted X inactivation maintained by a mouse Polycomb group gene. *Nat Genet* 28:371–375. doi: 10.1038/ng574
- Wang JC (2005) Finding primary targets of transcriptional regulators. *Cell Cycle* 4:356–358
- Weinberg SE, Singer BD, Steinert EM, et al (2019) Mitochondrial complex III is essential for suppressive function of regulatory T cells. *Nature* 565:495–499
- Welsbie DS, Xu J, Chen Y, et al (2009) Histone deacetylases are required for androgen receptor function in hormone-sensitive and castrate-resistant prostate cancer. *Cancer Res* 69:958–

966. doi: 10.1158/0008-5472.CAN-08-2216
- Wilson BJ, Bates GJ, Nicol SM, et al (2004) The p68 and p72 DEAD box RNA helicases interact with HDAC1 and repress transcription in a promoter-specific manner. *BMC Mol Biol* 5.: doi: 10.1186/1471-2199-5-11
- Wisdom R, Johnson RS, Moore C (1999) c-Jun regulates cell cycle progression and apoptosis by distinct mechanisms. *EMBO J* 18:188–97. doi: 10.1093/emboj/18.1.188
- Wortham NC, Ahamed E, Nicol SM, et al (2009) The DEAD-box protein p72 regulates ER α -oestrogen-dependent transcription and cell growth, and is associated with improved survival in ER α -positive breast cancer. *Oncogene* 28:4053–4064. doi: 10.1038/onc.2009.261
- Xu D, Zhan Y, Qi Y, et al (2015) Androgen receptor splice variants dimerize to transactivate target genes. *Cancer Res* 75:3663–3671. doi: 10.1158/0008-5472.CAN-15-0381
- Yang L, Lin C, Jin C, et al (2013) lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. *Nature* 500:598–602. doi: 10.1038/nature12451
- Yu H, Kim PM, Sprecher E, et al (2007) The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Comput Biol* 3:e59. doi: 10.1371/journal.pcbi.0030059
- Yu J, Yu J, Mani RS, et al (2010) An Integrated Network of Androgen Receptor, Polycomb, and TMPRSS2-ERG Gene Fusions in Prostate Cancer Progression. *Cancer Cell* 17:443–454. doi: 10.1016/j.ccr.2010.03.018
- Yu X, Gupta A, Wang Y, et al (2005) Foxa1 and Foxa2 interact with the androgen receptor to regulate prostate and epididymal genes differentially. In: *Annals of the New York Academy of Sciences*. pp 77–93
- Zhang P, Leu JIJ, Murphy ME, et al (2014) Crystal structure of the stress-inducible human heat shock protein 70 substrate-binding domain in complex with peptide substrate. *PLoS One* 9.: doi: 10.1371/journal.pone.0103518
- Zhang S, Murtha PE, Young CYF (1997) Defining a functional androgen responsive element in the 5' far upstream flanking region of the prostate-specific antigen gene. *Biochem Biophys Res Commun* 231:784–788. doi: 10.1006/bbrc.1997.6197

Resume & Publications

I was born in Istanbul and went to high school in Robert College. I completed my undergraduate and master's studies on Molecular Biology and Genetics at Bilkent University, working on liver cancer in the laboratory of Dr. Mehmet Öztürk. I did internships at the ULB in Brussels in the laboratory of Dr. Decio Eizirik on diabetes, at the BCCA in Vancouver with Dr. Yuzhuo Wang on prostate cancer, and at the NCI in Maryland with Dr. Ruth Nussinov on B Cell Receptor signaling. I completed the PhD work described here in the Computational Systems Biology Laboratory of Koç University, under the supervision of Dr. Özlem Keskin and Dr. Attila Gürsoy, in collaboration with Dr. Nathan Lack on the prostate cancer work. Below are the articles we were able to publish from my PhD work;

Cavga A.D., Tardu M., Korkmaz, T., Keskin, O., Ozturk, N., Gursoy, A., Kavakli, I.H.. **Cryptochrome deletion in p53 mutant mice enhances apoptotic and anti-tumorigenic responses to UV damage at the transcriptome level.** Functional & Integrative Genomics, 2019. doi:10.1007/s10142-019-00680-5.

Cavga A.D., Karahan N., Keskin O., Gursoy A.. **Taming Oncogenic Signaling at Protein Interfaces: Challenges and Opportunities.** Current Topics in Medicinal Chemistry.2015;15(20):2005-2018.
doi:http://dx.doi.org/10.2174/1568026615666150519101956.

DeepCOP – Deep Learning–Based Approach to Predict Gene Regulating Effects of Small Molecules

Godwin Woo, Michael Fernandez, Michael Hsing, Nathan A. Lack, **Ayse Derya Cavga**, Artem Cherkasov
accepted, Bioinformatics

Beyond the Heterodimer Model of Interaction for Mineralocorticoid and Glucocorticoid Receptors in the Nucleus and at DNA

John R. Pooley, Caroline A. Rivers, Michael T. Kilcooley, Susana N. Paul, **Ayse Derya Cavga**, Luke M. Davis, Yvonne M. Kershaw, Serena Muratcioglu, R. Louis Schiltz, Attila Gursoy, Ozlem Keskin, Stafford L. Lightman
under revision, PLOS ONE

Appendix A

Protein Interactors of lncRNA KLK3e

KLK3e sequence from publication (Hsieh et al. 2014)

TGGGACAACTTGCAAACCTGCTCAGCCTTTGTCTCTGATGAAGATATTATCTTCATG
 ATCTTGGATTGAAAACAGACCTACTCTGGAGGAACATATTGTATCGATTGTCCTTGA
 CAGTAAACAAATCTGTTGTAAGAGACATTATCTTTATTATCTAGGACAGTAAGCAA
 GCCTGGATCTGAGAGAGATATCATCTTGCAAGGATGCCTGCTTTACAAACATCCTT
 GAAACAACAATCCAGAAAAAAAAGGTGTTGCTGTCTTTGCTCAGAAGACACACA
 GATACGTGACAGAACCATGGAGAATTGCCTCCCAACACTGTTTCAGCCAGAGCCTTC
 CACCCTTGTCTGCAGGACAGTCTCAACGTTCCACCATTAAATACTTCTTCTGTCA
 TCCTGCTTATTTATGCCTAACCAAGGTTCTAGGTCCCGATCGACTGTGTCTGGCAGC
 ACTCCACTGCCAAACCCAGAATAAGGCAGCGCTCAGGATCCCGAAGGGGCATGGC
 TGGGGATCAGAACTTCTGGGTTTGAGTGAGGAGTGGGTCCACCCTCTTGAATTTCA
 AAGGAGGAAGAGGCTGGATGTGAAGGAAGTGGGGGAGGGAAAGTGTCAAGTTCCGA
 ACTCTTAGGTCAATGAGGGAGGAGACTGGTAAGGTCCCAGTCCCAGAGTACTGAT
 GTGGGAATGGCCTAAGAATCTCATATCCTCAGGAAGAAGGTGCTGGAATCCTGAGG
 GGTAGAGTTCTGGGTATATTTGTGGCTTAAGGCTCTTTGGCCCCTGAAGGGCAGAG
 GCTGGAACCATTAGGTCCAGGGTTTGGGGTGATAGTAATGGGATCTCTTGATTCCTC
 AAGAGTCTGAGGATCGAGGGTTGCCATTCTTCCATCTTGCCACCTAATCCTTACTC
 CACTTGAGGGTATCACCAGCCCTTCTAGTCCATGAAGGTGCCCTTGGGCAAGCAC
 AATCTGAGCATGAAAGATGCCCCAGAGGCCTTGGGTGTCATCCACTCATCATCCAG
 CATCCACACTCTGAGGGTGTGGCCAGCACCATGACGTCATGTTGCTGTGACTATCCC
 TGCAGCGTGCCTCTCCAGCCACCTGCCAACCGTAGAGCTGCCGACATCCTCCTCTGG
 TGGGAGTGGCCTGCATGGTGCCAGGCTGAGGCCTAGTGTGACAGAGGGAGCCTGG
 AATCATAGGGATCCAGGACTCAAAAGTGCTAGAGAATGGCCATATGTCACCATCCA
 TGAAATCTCAAGGGCTTCTGGGTGGAGGGCACAGGGACCTGAACCTTATGGGTTTTTC
 CCCAAGTCTATTGCTCTCCCAAGTGAGTCTCCAGATACGAGGCACTGTGCCAGCA
 TCAGCCTTATCTCCACCACATCTTGTAAGGGACTACCCAGGGCCCTGATGAACA
 CCATGGTGTGTACAGGAGTAGGGGGTGGAGGCACGGAATCCTGTGAGGTCACAGC
 CAAGGGAGCATCATCATGGGTGGGGAGGAGGCAATGGACAGGCTTGAGAACGGGG
 ATGTGGTTGTATTTGGTTTTCTTTGGTTAGATAAAGTGCTGGGTATAGGATTGAGAG
 TGGAGTATGAAGACCAGTTAGGATGGAGGATCAGATTGGAGTTGGGTAGAGATG
 GGGTAAAATTGTGCTTCGGATGAGTTTGGGATTGACACTGTGGAGGTGGTTTGGGA
 TGGCATGGCTTTGGGATGGAAATAGATTTGTTTTGATGTTGGCTCAGACATCCTTGG
 GGATTGAACTGGGGATGAAGCTGGGTTTGATTTTGGAGGTAGAAGACGTGGAAGTA
 GCTGTGAGATTTGACAGTGGCCATGAGTTTTGTTTGTATGGGGAATCAAACAATGGG
 GGAAGACATAAGGGTTGGCTTGTAGGTTAAGTTGCGTTGGGTGATGGGGTCGGG
 GCTGTGTATAATGCAGTTGGATTGGTTTGTATTAAATTGGGTGGGTGAGGTTTGG
 TTGAGGATGAGTTGAGGATATGCTTGGGGACACCGGATCCATGAGGTTCTCACTGG
 AGTGGAGACAACTTCCTTCCAGGATGAATCCGGGGAAGCCTTAATTCACGTGTA
 GGGGAGGTCAGGCCACTGGCTAAGTATATCCTTCCACTCCAGCTCTAAGATGGTCT
 TAAATTGTGATTATCTATATCCACCTCTGTCTCCCTCA

PCR amplified KLK3e transcript sequencing result

>181224-066_M24_KLK_2K_T7_T7promoter.ab1 1514

CAGGCTTCATGTCTCTGATGATAATTATCTTGATGATCTTGGATTGAGAC
AGACCTACTCTGGATGAACTTATTGTATCGATTGTCGTTGACAGTAAACA
AATCTGTTGTAAGAGACATTATCTTTATTATCTAGGACAGTAAGCAAGCC
TGGATCTGAGAGAGATATCATCTTGCAAGGATGCCTGCTTTACAAACATC
CTTGAAACAACAATCCAGAAAAAAGGTGTTGCTGTCTTTGCTCAGAA
GACACACAGATACGTGACAGAACCATGGAGAATTGCCTCCCAACACTGTT
CAGCCAGAGCCTTCCACCCTTGTCTGCAGGACAGTCTCAACGTTCCACCA
TTAAATACTTCTTCTGTACATCCTGCTTATTTATGCCTAACCAAGGTTT
TAGGTCCCGATCGACTGTGTCTGGCAGCACTCCACTGCCAAACCCAGAAT
AAGGCAGCGCTCAGGATCCCGAAGGGGCATGGCTGGGGATCAGAACTTCT
GGGTTTGAGTGAGGAGTGGGTCCACCCTCTTGAATTTCAAAGGAGGAATA
GGCTGGATGTGAAGGTACTGGGGGAGGGAAAGTGTCAGTTCCGAACCTCT
AGGTCAATGAGGGAGGAGACTGGTAAGGTCCCAGCTCCCGAGGTACTGAT
GTGGGAATGGCCTAAGAATCTCATATCCTCAGGAAGAAGGTGCTGGAATC
CTGAGGGGTAGAGTTCTGGGTATATTTGTGGCTTAAGGCTCTTTGGCCCC
TGAAGGGCAGAGGCTGGAACCATTAGGTCCAGTGTTTGGGGTGATAGTAA
TGGGATCTCTTGATTCTCAAGAGTCTGAGGATCGAGGGTTGCCATTCT
TCCATCTTGCCACCTAATCCTTACTCCACTTGAGGGTATTACCAGCCCTT
CTAGCTCCATGAAGGTGCCCCTGGGCAAGCACAATCTGAGCATGAAAGAT
GCCCCAGAGGCCTTGGGTGTCATCCACTCATCATCCAGCATCCACACTCT
GAAGGGTGTGGCCAGCACCATGACGTCATGTTGCTGTGACTATCCCTGCA
ACGTGGCTTCTCCAGCCACCTGCCAACCGTAAAGCTGCCGACATCCTCCT
CCTGGTGGGAATGGCCTGCATGGTGCCAGGCTGAAGCCTATTGTCAAACA
GGGAACCTGGAATCCTAGGGATCCAGGACTCAAATTTGCTAAGAAATGGC
CCAAGGTCACCATCCCTGGAATCTCAAGGGCTTTCTGGGTTGAAAGGCCA
AGGAAATTGAACCTTATTGGTTTTTCCCAAGTCTATTGGCTTCCCAAT
GGAATCTCCCAAATAAAAGGGACGTTGGGCCAAATAAGCCTTTATTCCCC
CCAATCTTTGTAAAAAGAAAATAACCCCGGCCCCCTATAAAACCAACTG
GTTGTTTCAAAGAAAAAAGGGGGGTGAAAGACAGCACACTCCTTTTGGGT
GTTACCCCCCCCACGGAAAATTCCTCTCTCTCGTTGGGGGGGAAAAAAA
GAGGAAAACCCCAA

PCR amplified KLK3e transcript sequencing result: BLAST alignment with KLK3e sequence from publication (Hsieh et al. 2014)

KLK3e_sense_S1_S7

Sequence ID: Query_138825 Length: 2170 Number of Matches: 1

Range 1: 30 to 1378

Score	Expect	Identities	Gaps	Strand
2093 bits(2320)	0.0	1290/1360(95%)	15/1360(1%)	Plus/Plus
Query 10	TGTCTCTGATGATAAT--TATCTTGATGATCTTGGATTGAGA-CAGACCTACTCTGGATG	66		
Sbjct 30	TGTCTCTGATGAAGATATTATCTTCATGATCTTGGATTGAAAACAGACCTACTCTGGAGG	89		
Query 67	AACCTATTGTATCGATTGTGCGTTGACAGTAAACAAATCTGTTGTAAGAGACATTATCTTT	126		
Sbjct 90	AACATATTGTATCGATTGTCTTGACAGTAAACAAATCTGTTGTAAGAGACATTATCTTT	149		
Query 127	ATTATCTAGGACAGTAAGCAAGCCTGGATCTGAGAGAGATATCATCTTGCAAGGATGCCT	186		
Sbjct 150	ATTATCTAGGACAGTAAGCAAGCCTGGATCTGAGAGAGATATCATCTTGCAAGGATGCCT	209		
Query 187	GCTTTACAAACATCCTTGAAACAACAATCCAGAAAAAAGGTGTTGCTGTCTTTGCTC	246		
Sbjct 210	GCTTTACAAACATCCTTGAAACAACAATCCAGAAAAAAGGTGTTGCTGTCTTTGCTC	269		
Query 247	AGAAGACACACAGATACGTGACAGAACCATGGAGAATTGCCTCCCAACACTGTTACGCCA	306		
Sbjct 270	AGAAGACACACAGATACGTGACAGAACCATGGAGAATTGCCTCCCAACACTGTTACGCCA	329		
Query 307	GAGCCTTCCACCCCTTGCTGTCAGGACAGTCTCAACGTTCCACCATTAAATACTTCTCTG	366		
Sbjct 330	GAGCCTTCCACCCCTTGCTGTCAGGACAGTCTCAACGTTCCACCATTAAATACTTCTCTG	389		
Query 367	TCACATCCTGCTTATTTATGCCTAACCAAGGTTCTAGGTCCCGATCGACTGTGTCTGGCA	426		
Sbjct 390	TCACATCCTGCTTATTTATGCCTAACCAAGGTTCTAGGTCCCGATCGACTGTGTCTGGCA	449		
Query 427	GCACTCCACTGCCAAACCCAGAATAAGGCAGCGCTCAGGATCCCGAAGGGGCATGGCTGG	486		
Sbjct 450	GCACTCCACTGCCAAACCCAGAATAAGGCAGCGCTCAGGATCCCGAAGGGGCATGGCTGG	509		
Query 487	GGATCAGAACTTCTGGGTTTGAGTGAGGAGTGGGTCCACCTCTTGAATTTCAAAGGAGG	546		
Sbjct 510	GGATCAGAACTTCTGGGTTTGAGTGAGGAGTGGGTCCACCTCTTGAATTTCAAAGGAGG	569		
Query 547	AATAGGCTGGATGTGAAGGTACTGGGGGAGGGAAAGTGTGAGTTCCGAACCTTTAGGTCA	606		
Sbjct 570	AAGAGGCTGGATGTGAAGGAAGTGGGGGAGGGAAAGTGTGAGTTCCGAACCTTTAGGTCA	629		
Query 607	ATGAGGGAGGAGACTGGTAAGGTCCCAGCTCCCAGGTAAGTGTGGGAATGGCCTAAG	666		
Sbjct 630	ATGAGGGAGGAGACTGGTAAGGTCCCAGCTCCCAGGTAAGTGTGGGAATGGCCTAAG	689		
Query 667	AATCTCATATCCTCAGGAAGAAGGTGCTGGAATCCTGAGGGGTAGAGTTCTGGGTATATT	726		
Sbjct 690	AATCTCATATCCTCAGGAAGAAGGTGCTGGAATCCTGAGGGGTAGAGTTCTGGGTATATT	749		
Query 727	TGTGGCTTAAGGCTCTTTGGCCCCGTAAGGGCAGAGGCTGGAACCATTAGGTCCAGTGTT	786		
Sbjct 750	TGTGGCTTAAGGCTCTTTGGCCCCGTAAGGGCAGAGGCTGGAACCATTAGGTCCAGGGTT	809		
Query 787	TGGGGTGATAGTAATGGGATCTCTTGATTCTCAAGAGTCTGAGGATCGAGGGTTGCCCA	846		
Sbjct 810	TGGGGTGATAGTAATGGGATCTCTTGATTCTCAAGAGTCTGAGGATCGAGGGTTGCCCA	869		
Query 847	TTCTTCCATCTTGCCACCTAATCCTTACTCCACTTGAGGGTATTACCAGCCCTTCTAGCT	906		
Sbjct 870	TTCTTCCATCTTGCCACCTAATCCTTACTCCACTTGAGGGTATCACCAGCCCTTCTAGCT	929		
Query 907	CCATGAAGGTGCCCTGGGCAAGCACAATCTGAGCATGAAAGATGCCCCAGAGGCTTGG	966		
Sbjct 930	CCATGAAGGTGCCCTGGGCAAGCACAATCTGAGCATGAAAGATGCCCCAGAGGCTTGG	989		
Query 967	GTGTCATCCACTCATCATCCAGCATCCACACTCTGAAGGGTGTGGCCAGCACCATGACGT	1026		
Sbjct 990	GTGTCATCCACTCATCATCCAGCATCCACACTCTG-AGGGTGTGGCCAGCACCATGACGT	1048		
Query 1027	CATGTTGCTGTGACTATCCCTGCAACGTGGCTTCTCCAGCCACCTGCCAACCCTAAAGCT	1086		
Sbjct 1049	CATGTTGCTGTGACTATCCCTGCAGCGT-GCCTCTCCAGCCACCTGCCAACCCTAGAGCT	1107		
Query 1087	GCCGACATCCTCCTCTGGTGGGAATGGCCTGCATGGTGCCAGGCTGAAGCCTATTGTCA	1146		
Sbjct 1108	GCCGACATCCTCCT-CTGGTGGGAGTGGCCTGCATGGTGCCAGGCTGAGGCCTAGTGTCA	1166		
Query 1147	AACAGGGAACCTGGAATCCTAGGGATCCAGGACTCAAAATTGCTAAGAAATGGCCCAAGG	1206		


```

Sbjct  1870  TTGCTGTGACTATCCCTGCAGCGTGCC-TCTCCAGCCACCTGCCAACCGTAGAGCTGCCG 1928
Query  1091  ACATCCTCCTCCTGGTGGGAATGGCCTGCATGGTGCCAGGCTGAAGCCTATTGTCAAACA 1150
      |||||
Sbjct  1929  ACATCCTCCT-CTGGTGGGAGTGGCCTGCATGGTGCCAGGCTGAGGCCTAGTGTGAGACA 1987
Query  1151  GGGAACCTGGAATCCTAGGGATCCAGGACTCAAAATTGCTA-AGAAATGGCCCAAGGTCA 1209
      |||||
Sbjct  1988  GGGAGCCTGGAATCATAGGGATCCAGGACTCAAAAGTGCTAGAGAA-TGGCCATATGTCA 2046
Query  1210  CCATCCCTGGAATCTCAAGGGCTTCTGGGTGAAAGGCCAAGGAAATTGAACCTTATTG 1269
      |||||
Sbjct  2047  CCATCCATGAAATCTCAAGGGCTT-CTGGGTGGAG-GGCACAGGGACCTGAAC-TTATGG 2103
Query  1270  GTTTTCCCAAGTCTATTGGCTTCCCCAATGGAATCTCCCAAATAAAAGGGACGTTGGG 1329
      |||||
Sbjct  2104  G-TTTTCCCAAGTCTATTG-CTCTCCCAAGTGAGTCTCCAGATACGAGGCAC-T-GTG 2159
Query  1330  CCAA-ATAAGCCTTTAT-TCCCCCAATCTTTGTAAAA 1365
      |||
Sbjct  2160  CCAGCATCAGCCTT-ATCTCCACCACATCTT-GTAAAA 2195

```

Homo sapiens uncharacterized LOC105372441 (LOC105372441), transcript variant 1, long non-coding RNA

Sequence ID: [NR_131205.1](#) Length: 463 Number of Matches: 1

Score	Expect	Identities	Gaps	Strand
433 bits(234)	4e-118	234/234(100%)	0/234(0%)	Plus/Plus
Query 135	GGACAGTAAGCAAGCCTGGATCTGAGAGAGATATCATCTTGCAAGGATGCCTGCTTTACA	194		
Sbjct 230	GGACAGTAAGCAAGCCTGGATCTGAGAGAGATATCATCTTGCAAGGATGCCTGCTTTACA	289		
Query 195	AACATCCTTGAAACAACAATCCAGAAAAAAGGTGTTGCTGTCTTTGCTCAGAAGACA	254		
Sbjct 290	AACATCCTTGAAACAACAATCCAGAAAAAAGGTGTTGCTGTCTTTGCTCAGAAGACA	349		
Query 255	CACAGATACGTGACAGAACCATGGAGAATTGCCTCCCAACACTGTTTCAGCCAGAGCCTTC	314		
Sbjct 350	CACAGATACGTGACAGAACCATGGAGAATTGCCTCCCAACACTGTTTCAGCCAGAGCCTTC	409		
Query 315	CACCCCTGTCTGCAGGACAGTCTCAACGTTCCACCATTAAATACTTCTTCTGTC	368		
Sbjct 410	CACCCCTGTCTGCAGGACAGTCTCAACGTTCCACCATTAAATACTTCTTCTGTC	463		

Homo sapiens KLK3 upstream enhancer/promoter region (LOC106128905) on chromosome 19

Sequence ID: [NG_042790.1](#) Length: 5990 Number of Matches: 1

Range 1: 1720 to 3068

Score	Expect	Identities	Gaps	Strand
2126 bits(1151)	0.0	1295/1362(95%)	19/1362(1%)	Plus/Plus
Query 10	TGTCTCTGATG-ATA-ATTATCTTGATGATCTTGGATTG-AGACAGACCTACTCTGGATG	66		
Sbjct 1720	TGTCTCTGATGAAGATATTATCTTCATGATCTTGGATTGAAAACAGACCTACTCTGGAGG	1779		
Query 67	AACTTATTGTATCGATTGTGCGTTGACAGTAAACAATCTGTTGTAAGAGACATTATCTTT	126		
Sbjct 1780	AACATATTGTATCGATTGTCCTTGACAGTAAACAATCTGTTGTAAGAGACATTATCTTT	1839		
Query 127	ATTATCTAGGACAGTAAGCAAGCCTGGATCTGAGAGAGATATCATCTTGCAAGGATGCCT	186		
Sbjct 1840	ATTATCTAGGACAGTAAGCAAGCCTGGATCTGAGAGAGATATCATCTTGCAAGGATGCCT	1899		
Query 187	GCTTTACAAACATCCTTGAAACAACAATCCAGAAAAAAGGTGTTGCTGTCTTTGCTC	246		
Sbjct 1900	GCTTTACAAACATCCTTGAAACAACAATCCAGAAAAAAGGTGTTGCTGTCTTTGCTC	1959		
Query 247	AGAAGACACACAGATACGTGACAGAACCATGGAGAATTGCCTCCCAACACTGTTTCAGCCA	306		
Sbjct 1960	AGAAGACACACAGATACGTGACAGAACCATGGAGAATTGCCTCCCAACACTGTTTCAGCCA	2019		
Query 307	GAGCCTTCCACCCTTGTCTGCAGGACAGTCTCAACGTTCCACCATTAAATACTTCTTCTG	366		
Sbjct 2020	GAGCCTTCCACCCTTGTCTGCAGGACAGTCTCAACGTTCCACCATTAAATACTTCTTCTG	2079		
Query 367	TCACATCCTGCTTATTTATGCCTAACCAAGGTTCTAGGTCCCGATCGACTGTGTCTGGCA	426		

Sbjct	2080		TCACATCCTGCTTATTTATGCCTAACCAAGGTTCTAGGTCCC	GATCGACTGTGTCTGGCA	2139
Query	427		GCACTCCACTGCCAAACCCAGAATAAGGCAGCGCTCAGGATCCC	GAAGGGGCATGGCTGG	486
Sbjct	2140		GCACTCCACTGCCAAACCCAGAATAAGGCAGCGCTCAGGATCCC	GAAGGGGCATGGCTGG	2199
Query	487		GGATCAGAACTTCTGGGTTTGAGTGAGGAGTGGGTCCACCCTCT	TGAATTTCAAAGGAGG	546
Sbjct	2200		GGATCAGAACTTCTGGGTTTGAGTGAGGAGTGGGTCCACCCTCT	TGAATTTCAAAGGAGG	2259
Query	547		AATAGGCTGGATGTGAAGGTACTGGGGGAGGGAAAGTGT	CAGTTCCGAACCTCTTAGGTCA	606
Sbjct	2260		AAGAGGCTGGATGTGAAGGAACTGGGGGAGGGAAAGTGT	CAGTTCCGAACCTCTTAGGTCA	2319
Query	607		ATGAGGGAGGAGACTGGTAAGGTCCCAGCTCCCAGGTA	CTGATGTGGGAATGGCCTAAG	666
Sbjct	2320		ATGAGGGAGGAGACTGGTAAGGTCCCAGCTCCCAGGTA	CTGATGTGGGAATGGCCTAAG	2379
Query	667		AATCTCATATCCTCAGGAAGAAGGTGCTGGAATCCTGAGGGG	TAGAGTTCTGGGTATATT	726
Sbjct	2380		AATCTCATATCCTCAGGAAGAAGGTGCTGGAATCCTGAGGGG	TAGAGTTCTGGGTATATT	2439
Query	727		TGTGGCTTAAGGCTCTTTGGCCCCGAAGGGCAGAGGCTGGA	ACCATTAGGTCCAGTGTT	786
Sbjct	2440		TGTGGCTTAAGGCTCTTTGGCCCCGAAGGGCAGAGGCTGGA	ACCATTAGGTCCAGGGTT	2499
Query	787		TGGGGTGATAGTAATGGGATCTCTTGATTCTCAAGAGTCTG	AGGATCGAGGGTTGCCCA	846
Sbjct	2500		TGGGGTGATAGTAATGGGATCTCTTGATTCTCAAGAGTCTG	AGGATCGAGGGTTGCCCA	2559
Query	847		TTCTTCCATCTTGCCACCTAATCCTTACTCCACTTGAGGGT	TATTACCAGCCCTTCTAGCT	906
Sbjct	2560		TTCTTCCATCTTGCCACCTAATCCTTACTCCACTTGAGGGT	TATCACCAGCCCTTCTAGCT	2619
Query	907		CCATGAAGGTGCCCTGGGCAAGCACAATCTGAGCATGAAAG	ATGCCCCAGAGGCTTGG	966
Sbjct	2620		CCATGAAGGTGCCCTGGGCAAGCACAATCTGAGCATGAAAG	ATGCCCCAGAGGCTTGG	2679
Query	967		GTGTCATCCACTCATCATCCAGCATCCACACTCTGAAGGGT	GTGGCCAGCACCATGACGT	1026
Sbjct	2680		GTGTCATCCACTCATCATCCAGCATCCACACTCTG-AGGGT	GTGGCCAGCACCATGACGT	2738
Query	1027		CATGTTGCTGTGACTATCCCTGCAACGTGGCTTCTCCAGCC	ACCTGCCAACCGTAAAGCT	1086
Sbjct	2739		CATGTTGCTGTGACTATCCCTGCAGCGTGCC-TCTCCAGCC	ACCTGCCAACCGTAGAGCT	2797
Query	1087		GCCGACATCCTCCTCCTGGTGGGAATGGCTGCATGGTGCC	AGGCTGAAGCCTATTGTCA	1146
Sbjct	2798		GCCGACATCCTCCT-CTGGTGGGAGTGGCTGCATGGTGCC	AGGCTGAGGCCTAGTGTCA	2856
Query	1147		AACAGGGAACCTGGAATCCTAGGGATCCAGGACTCAAAAT	TGCTA-AGAAATGGCCCAAG	1205
Sbjct	2857		GACAGGGAGCCTGGAATCATAGGGATCCAGGACTCAAAAGT	GCTAGAGAA-TGGCCATAT	2915
Query	1206		GTCACCATCCCTGGAATCTCAAGGGCTTTCTGGGTTGAAAG	GCCAAGGAAATTGAACCTT	1265
Sbjct	2916		GTCACCATCCATGAAATCTCAAGGGCTT-CTGGGTGGAG-G	GCACAGGGACCTGAAC-TT	2972
Query	1266		ATTGGTTTTTCCCCAAGTCTATTGGCTTCCCCAATGGAAT	CTCCAAATAAAAGGGACGT	1325
Sbjct	2973		ATGGG-TTTTCCCCAAGTCTATTG-CTCTCCCAAGTGAGT	CTCCAGATACGAGGCAC-T	3029
Query	1326		TGGGCCAA-ATAAGCCTTTAT-TCCCCCAATCTTTGTAAAA	1365	
Sbjct	3030		-GTGCCAGCATCAGCCTT-ATCTCCACCACATCTT-GTAAAA	3068	

PCR amplified Xist transcript sequencing result

>190312-075_C21_Xist_PCRcln_T7promoter.ab 1710

CAACATGTATGGCGTATTTCTTTAAAAATCACCTAAATTCCATAAAAATATTT
TTTTAAATTCTATACTTTCTCCTAGTGTCTTCTTGACACGTCCTCCATATTTTT
TTAAAGAAAGTATTTGGAATATTTTGAGGCAATTTTAAATATTTAAGGAATT
TTTCTTTGGAATCATTTTTTGGTTGACATCTCTGTTTTTTGTGGATCAGTTTTTT
ACTCTTCCACTCTCTTTTCTATATTTTGCCCATCGGGGCTGCGGATACCTGGT
TTTATTATTTTTTTCTTTGCCCAACGGGGCCGTGGATACCTGCCTTTTAATTCT
TTTTTATTTCGCCCATCGGGGCCGCGGATACCTGCTTTTTTATTTTTTTTTTCCTTA
GCCCATCGGGGTATCGGATACCTGCTGATTCCCTTCCCCTCTGAACCCCCAA
CACTCTGGCCCATCGGGGTGACGGATATCTGCTTTTTTAAAAATTTTCTTTTTT
TGGCCCATCGGGGCTTCGGAACCTGCTTTTTTTTTTTTTTATTTTTTCCTTGCCC
CCCGGGGCCCCGGAACCGGGTTTAATTTTTGTTTTTTGGGCCCCCGGGGC
CCCGGAAACCCGCTTTGATTTTTTTTTTTCCCCCCCCCGGGGGGTTTTTTGG
GAGGAAAAAAGGTGGTTTTTGGGGGTGGTGGCCCCCCCCCGGAAATTAACC
ATTTTTTTTGTGGGGGAAAATTGGGGGGGGGGGGGGGGGGGCCCTCGCTTTT
GGCAAACAAAAAATTCCGCGTTAAACCTAGGGGGGGGTTATAATTGTGGGA
GAAGAAGGGGCCGCCTTTTTATAATTGCGCCTAACAAGGCCTAAGGGATAT
CTTTTTGTGTGAGTTTAACATGCGGGGGGGAGGGGGTGAATTCCTGTCCGGC
AGAGAAAATCATGGGGTTGCCGGGGACAAACCCTCCCGAAAGAAATGATTT
GGGGCTGGTAGGACCTTCGCCGATGGGGTGGCGGGGGGCGGGCAAAAAAT
CCCTTTTTTCCTTCCCAGGGGAAAAGTTAAAAATGGGGGTGGGAAGAAAAA
GGGCCGCGGCCTAAAATGGGGGGATTTTATATTTTGGGGAATTTTTCCCCCA
AGGGGGGGGGGGGGGCTTTTTTTAGTTTTTTTTATTGGTTGAAACCACCTCTTG
CGCCCGCGGGTGGCGAGAGGGGGGAAAAAAGGACCCCCCGCGCGGGGCG
GCGGCCTTTTTTTTAGTTATTAATAATGGGGGTGGTGGCGGGGTCCCCCCCC
AAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAATAAACACACCGCAGCAAAC
AGAGAAACCCCGATATTATTTTTTTTATTTTTTTTTTTTTTTGTTTTGTTTGTCC
TGGCGGGAGGGAAGGAGAGGGGGGGGGGAAGAATGAAAAAAAAATATTTTT
CTTTCCCCTCACACAGGAGAGTGGAATATATTAATAAGGAGGGGTGTGTG
CCCCTGTATGCCGACAACAAAAAGAGAGAGGGGGGGGCGGGGGGCGGGGT
GGTGGCTACCCCCCTCTCACTCCGCGGGGGGGGGGGGAGAAAATAAGAAA
GAATAAAAAATAAAAAATAAACTCCATAATTTTATTATATATATAATTATA
ATAAATAAAAAATTAGGGAGTTAAGAGAAGAGAATGTGTGGTTGAGAGGA
AGAAATGGAGTAGAAGAGAGACGAGAGAGAAGAGAGCGACACACGCCCGC
CGCC

Appendix Figure A.1: Xist sequencing quality low towards end.

PCR amplified Xist transcript sequencing result: BLAST alignments of matches from NCBI database Homo Sapiens Nucleotide Collection

Homo sapiens X inactive specific transcript (non-protein coding) (XIST), long non-coding RNA

Sequence ID: [NR_001564.2](#) Length: 19296 Number of Matches: 2

Range 1: 137 to 1198

Score	Expect	Identities	Gaps	Strand
1095 bits(1214)	0.0	898/1069(84%)	21/1069(1%)	Plus/Plus
Query 8		TATGGCGTATTTCTTTAAAAA--TCACCTAAATTCATAAAATA		65
Sbjct 137		TATGGCGTATTTCTTTAAAAAATCACCTAAATTCATAAAATATTTT		196
Query 66		ACTTTCTCCTAGTGTCTTCTTGACACGTCCTCCAtat		125
Sbjct 197		ACTTTCTCCTAGTGTCTTCTTGACACGTCCTCCATATTTT		256
Query 126		tat		185
Sbjct 257		TATTTTGAGGCAATTTTAAATATTTAAGGAATTTTCTTTGGAATCATT		316
Query 186		tctctgtttttgtggatcagtttttACTCTTCCACTCTCTTTTCTATATTTGCCCAT		245
Sbjct 317		TCTCTGTTTTTTGTGGATCAGTTTTTACTCTTCCACTCTCTTTTCTATATTTGCCCAT		376
Query 246		CGGGGCTGCGGATACCTGGTTTTATTATTTTTCTTTGCCAACGGGGCGTGGATACCT		305
Sbjct 377		CGGGGCTGCGGATACCTGGTTTTATTATTTTTCTTTGCCAACGGGGCGTGGATACCT		436
Query 306		GCCTTTTAATTCTTTTTATTTCGCCCATCGGGGCGCGGATACCTGC		365
Sbjct 437		GCCTTTTAATTCTTTTTATTTCGCCCATCGGGGCGCGGATACCTGC		496
Query 366		ttCCTTAGCCCATCGGGGTATCGGATACCTGCTGATTCCCTTCCCTCTGAACCCCAAC		425
Sbjct 497		TTCTTAGCCCATCGGGGTATCGGATACCTGCTGATTCCCTTCCCTCTGAACCCCAAC		556
Query 426		ACTCTGGCCCATCGGGGTGACGGATATCTGCTtttttaaaaat		485
Sbjct 557		ACTCTGGCCCATCGGGGTGACGGATATCTGCTTTTAAAAATTTCTTTT		616
Query 486		CGGGGCTTCGGAACCTGCT-tttttttttttat		544
Sbjct 617		CGGGGCTTCGGAACCTGCTTTTTTTTTTTTATTTTCTTGC		676
Query 545		aaccgggtttaattttgttttttgggcccccgggccccggAAACCCGCTTTGA		604
Sbjct 677		TACCTGCTTTAATTTTGTCTTTTCTGGCCATCGGGGCGCGGATACCTGCTTTGATTTT		736
Query 605		ttttttcccccccccggggggttttttgggagaaaaaaggtggttttttgggggtggt		664
Sbjct 737		TTTTTTTCATCGCCCATCGGTGCTTTTATGGATGAAAAATGTTGGTTTGTGGGTTGT		796
Query 665		ggccccccccGAAAAATTAACCA		724
Sbjct 797		TGCACTCTCTGGAATATCTACACTTTTTTTTGCTGCTGATCATTGGTGGTGTGTGAGTG		856
Query 725		gCCCCCTCGCTTTTGGCAACAAAAAATCCGCGTTAAACCTAggggggggTTATAATTGT		784
Sbjct 857		TACCTACCGCTTTGGCAGA-GAATGACTCTGCAGTTAAGCTAAGGGCGTGTTCAGATTGT		915
Query 785		GGGAGAAGAAGGGCCGCCTTTTATAATTG-CGCCTAACAAAGGCTAAGGGATATCTTT		843
Sbjct 916		-GGAGGAAAAGTGGCCGCCATTTTAGACTTGCCGCATAACTCGG-CTTAGGGCTAGTCGT		973
Query 844		TTGTG-TGAGTTTAACATGCgggggggag-ggggTGAATTCCTGTCCGGCAGAGAAAATC		901
Sbjct 974		TTGTGCTAAGTTAAAC-TAGGGAGGCAAGATGGATGATAGCAGGTCAGGCAGAGGAAGTC		1032
Query 902		ATG-GGGTGGCCGGGACAAACCTCCCGAAAGAAATGATTGGGGC-TGGTAGGACCTT		959
Sbjct 1033		ATGTGCATTGCATGAGCTAAACCTATCTGAATGAATTGATTGGGGCTTGTAGGAGCTT		1092
Query 960		CGC-CGA-TGGGGTGGCGGGGGCGGGCAAAAAATC-CCTTTTTTCCTTCCCAggggaaa		1016
Sbjct 1093		TGCGTGATTGTTGTATCGGGAGGCAG--TAAGAATCATCTTTATCAGT-ACAAGGGACT		1149
Query 1017		agttaaaaaatgg-gggg--gggaagaaaaagggcgcgccctaaaaatgg		1062
Sbjct 1150		AGTTAAAAATGGAAGGTAGGAAAGACTAAGGTGCAGGGCTTAAATGG		1198

KLK3e Interacting Proteins Mass Spectrometry Results

Appendix Table A.1: Mass spectrometry coverage values of KLK3e interacting proteins detected in two biological replicates.

Gene	Cut Band Mass Spec		Whole Lane Mass Spec			SCORE	Molecular Weight (kDa)
	KLK3e	Xist	KLK3e	Xist	Cher	ΣCoverage (KLK3e-Xist-Cher)	
RPS4Y1	99.24	0	26.62	0	0	125.86	29
RPL7A	82.7	0	30.45	0	15.78	97.37	30
RPL19	85.21	0	43.37	14.28	19.38	94.92	23
H1FX	77	0	31.46	5.63	20.66	82.17	22
PRPF6	88.63	13.29	2.56	0	1.28	76.62	107
SART1	72	4.51	12	0	8.5	70.99	90
HIST1H4A	46.6	0	23.3	0	0	69.90	11
HSPA8	57.12	0	6.81	0	2.63	61.30	71
DDX5	66.45	21	47.88	20.19	20.35	52.79	69
CHTOP	43.95	0	6.45	0	0	50.40	26
ATAD3A	58.36	14.2	13.56	5.21	5.36	47.15	71
SYNCRIP	68.7	25.53	26.16	9.95	13.32	46.06	70
RPL31	29.6	0	14.4	0	0	44.00	14
ATAD3B	43.83	6.79	9.41	0	4.78	41.67	73
DDX17	39.5	0	15.64	3.02	10.56	41.56	80
RPL29	28.94	0	9.43	0	0	38.37	18
G3BP2	24.9	0	4.56	0	0	29.46	54
HNRNPF	8.92	10.6	60.24	0	29.4	29.16	46
EIF2AK2	24.86	0	14.16	0	11.25	27.77	62
LIG3	25.97	3.76	2.58	0	0	24.79	113
DDX54	20.89	0	2.61	0	0	23.50	99
HNRNPA0	5.24	0	37.38	20.98	0	21.64	31
UTP23	8.04	0	10.44	0	0	18.48	28
DDX3X	8.76	0	9.66	0	0	18.42	73
HNRNPA1	10.76	0	81.72	51.08	23.65	17.75	39
PSPC1	7.46	0	8.8	0	0	16.26	59
DDX21	50.45	43.17	14.68	0	7.41	14.55	87
KIF23	7.5	0	5.42	0	1.67	11.25	110
RORA	10.89	3.63	3.63	0	0	10.89	59
DHX36	11.21	6.05	4.86	0	0	10.02	115
HNRNPUL2-BSCL2	12.34	4.56	11.26	3.48	5.9	9.66	85
RPL15	3.43	0	11.76	0	5.88	9.31	24
PARP1	4.94	0	3.36	0	0	8.30	113
FAM111A	6.22	0	2.94	1.47	0	7.69	70
RPS25	5.6	0	40	20.8	17.6	7.20	14
DGCR8	2.46	0	3.36	0	0	5.82	86
PRPF39	8.52	5.68	2.84	0	0	5.68	78
CPSF1	1.52	0	3.81	1.11	0	4.22	161
ZFR	1.3	0	4.94	1.4	2.8	2.04	117
ATM	0.52	0	0.26	0	0	0.78	350
DYNC1H1	0.32	0	0.34	0	0	0.66	532

Xist Interacting Proteins Mass Spectrometry Results**Appendix Table A.2: Mass spectrometry coverage values of known Xist interactor proteins detected in Xist specific bands cut from gel.**

UniprotID	GeneName	Description	Σ Coverage Xist total	Σ Coverage KLK total	MW [kDa]
P61978	HNRNPK	Heterogeneous nuclear ribonucleoprotein K	51.18	6.26	50.9
P43243	MATR3	Matrin-3	129.16	36.48	94.6
Q00839	HNRNPU	Heterogeneous nuclear ribonucleoprotein U	73.82	44.01	90.5



Appendix Table A.3: GO and Pathway enrichment results of 41 KLK3e interacting proteins. Obtained from STRING database. (Jensen et al. 2009)

Biological Process (GO)			
GO-term	description	count in gene set	false discovery rate
GO:0016071	mRNA metabolic process	20 of 667	3.16e-16
GO:0090304	nucleic acid metabolic process	32 of 3941	1.44e-13
GO:0006396	RNA processing	19 of 825	1.44e-13
GO:0016070	RNA metabolic process	30 of 3430	3.26e-13
GO:0010467	gene expression	29 of 3733	2.22e-11

Molecular Function (GO)			
GO-term	description	count in gene set	false discovery rate
GO:1901363	heterocyclic compound binding	31 of 5305	3.21e-09
GO:0097159	organic cyclic compound binding	31 of 5382	3.21e-09
GO:0003723	RNA binding	15 of 850	3.21e-09
GO:0003676	nucleic acid binding	24 of 3332	3.42e-08
GO:0004004	ATP-dependent RNA helicase activity	6 of 66	2.47e-07

Cellular Component (GO)			
GO-term	description	count in gene set	false discovery rate
GO:1990904	ribonucleoprotein complex	17 of 770	7.00e-12
GO:0005730	nucleolus	15 of 926	1.29e-08
GO:0044428	nuclear part	27 of 4359	3.83e-08
GO:0043232	intracellular non-membrane-bounded organelle	26 of 4005	3.83e-08
GO:0031981	nuclear lumen	26 of 4030	3.83e-08

KEGG Pathways			
pathway	description	count in gene set	false discovery rate
hsa03010	Ribosome	6 of 130	1.13e-05
hsa03040	Spliceosome	5 of 130	0.00014
hsa03410	Base excision repair	2 of 33	0.0308

Reactome Pathways			
pathway	description	count in gene set	false discovery rate
HSA-8953854	Metabolism of RNA	16 of 652	1.11e-11
HSA-72203	Processing of Capped Intron-Containing Pre-mRNA	9 of 234	1.00e-07
HSA-72163	mRNA Splicing - Major Pathway	8 of 178	1.98e-07
HSA-168254	Influenza Infection	7 of 148	9.22e-07
HSA-975956	Nonsense Mediated Decay (NMD) independent of the Exon ...	6 of 92	9.56e-07

UniProt Keywords			
keyword	description	count in gene set	false discovery rate
KW-1017	Isopeptide bond	23 of 1713	3.10e-13
KW-0694	RNA-binding	17 of 661	3.10e-13
KW-0832	Ubl conjugation	22 of 2380	1.74e-09
KW-0687	Ribonucleoprotein	10 of 278	4.55e-09
KW-0597	Phosphoprotein	35 of 8066	7.80e-09

PFAM Protein Domains			
domain	description	count in gene set	false discovery rate
PF00270	DEAD/DEAH box helicase	6 of 75	7.11e-07
PF00271	Helicase conserved C-terminal domain	6 of 106	2.51e-06
PF00076	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	6 of 200	6.01e-05
PF00645	Poly(ADP-ribose) polymerase and DNA-Ligase Zn-finger reg...	2 of 2	0.00034
PF12037	Domain of unknown function (DUF3523)	2 of 3	0.00045

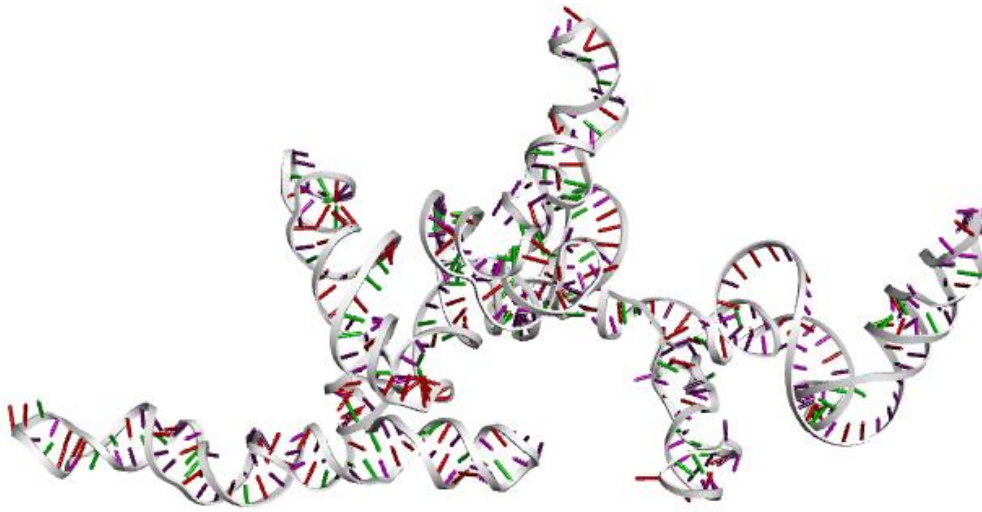
Appendix Table A.4: Performance of docking servers in CAPRI. (Lensink et al. 2017)

<u>SERVER</u>	RANK			
	<u>2017(a)</u>	<u>2017(b)</u>	<u>2016(c)</u>	<u>2013(d)</u>
CLUSPRO	2	1/1	2	1
DOCK/PIERR	NA	6	6	4#
GALAXYPPDOCK	NA	5	NA	NA
GRAMM-X	9	6	4	9
HADDOCK	7	3/2	1	2
HEXSEVER	NA	NA	NA	5
LZERD	3	2/3	5	7
MEGADOCK	NA	NA/3	NA	NA
PATCHDOCK/FIBERDOCK	NA	6	NA	9/6
PYDOCKWEB		4/1	NA	NA
SURFIT	NA	NA	NA	8
SWARMDOCK	6	3/3	3	3
TACOS	NA	NA	NA	9

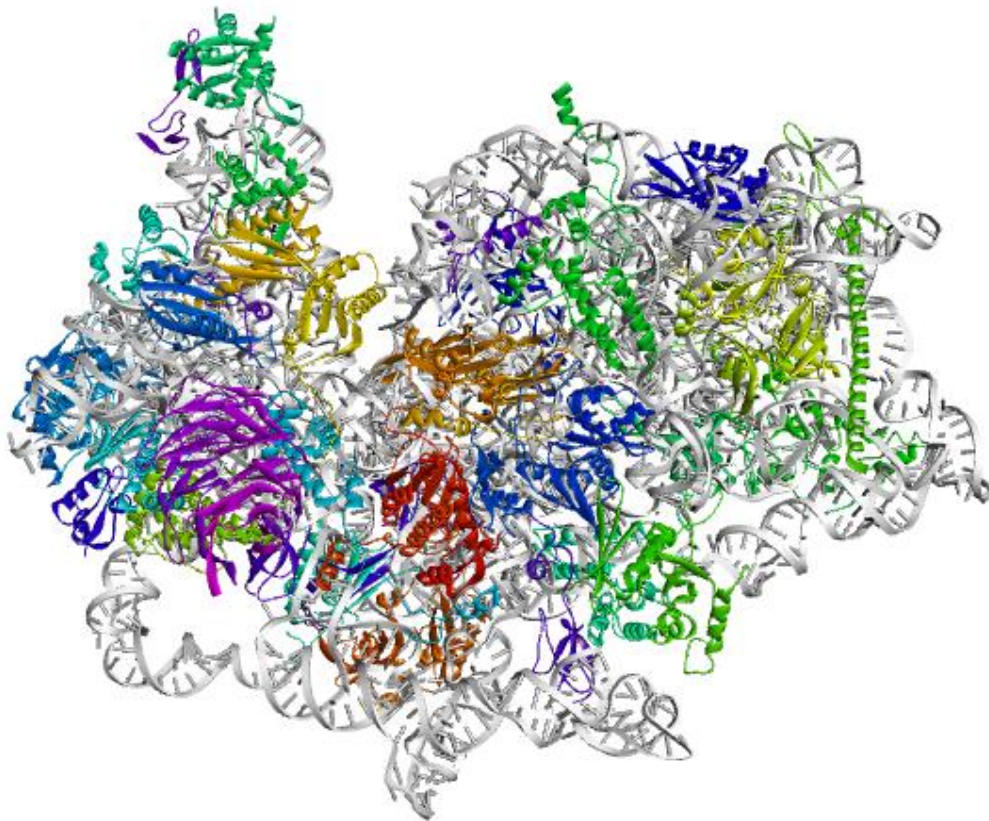
Table adapted from;

<http://www.capri-docking.org/resources/#performance-of-docking-servers-in-capri>

Appendix Figure A.2: First 500 bases of KLK3e modeled with RNA composer (Popenda et al. 2012) lacks the structure, motifs and cavities of the full size model.



Appendix Figure A.3: A sample electron microscopy structure of an RNA-protein complex of an RNA of similar size with KLK3e. PDB ID 5FLX shows the 1869 nucleotides long hepatitis C viral RNA bound by human ribosomal proteins, the RNA acting like sponge and allowing docking of many proteins.



Appendix B

Effects of Small Molecules Targeting the Androgen Receptor

Treatment of LNCaP cells with small molecules, RIME and RNA-seq preparation

Cell treatments were performed at the Vancouver Prostate Centre by Flora Huang. LNCaP cells (6×10^6) were seeded in RPMI 1640 media containing 10% FBS and 1% penicillin/streptomycin. Following overnight incubation, the media was changed to RPMI 1640 containing 5% CSS and the cells were then grown for 48hrs at 37°C with 5% CO₂. On the day of treatment, CSS media was refreshed and the cells were treated with either ethanol and one of the below, or DHT (10nM) and one of the below;

- 1) DMSO (Control)
- 2) 1.5 μ M Enzalutamide (MDV3100 (Tran et al. 2009; Scher et al. 2010; Beer et al. 2014)
- 3) 3 μ M molecule 13789 (Lack et al. 2011; Gao et al. 2013)
- 4) 6 μ M molecule 17005 (Dalal et al. 2018)
- 5) 6 μ M molecule 14449 (Dalal et al. 2014; Li et al. 2014a)
- 6) 6 μ M molecule 220010 (Antonarakis et al. 2016)

Concentration of the compound was selected based on the published IC₉₀ and RT-qPCR analysis of AR target genes provided in Appendix Figure B.7. Cells were treated for 6 hours.

For RIME analysis, Protein interactions in the cells were fixated using low dose formaldehyde, and cells were lysed. RIME protocol (Mohammed et al. 2016) was applied using antibodies targeting androgen receptor (AR).

For RNA sequencing, total RNA was extracted using Trizol and the Qiagen RNeasy RNA Extraction Kit. Three biological replicas were collected for each condition. The concentrations and the qualities of the RNA samples were checked by Nanodrop and tape station. RNA was then subjected to Poly-A selection prior to sequencing at the BGI in China.

**Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME)
analysis of AR interactors that are disrupted by small molecules**

Appendix Table B.1: LFQ intensity readings for proteins that are not present in IgG negative control (AR8), not present in EtOH (AR1), present in DHT (AR2) and not present with one or more of the compounds. (AR3-7).

Gene names	LFQ intensity AR-1 EtOH control	LFQ intensity AR2 DHT treated	LFQ intensity AR-3 Enza. LBD	LFQ intensity AR-4 13789 LBD BF3	LFQ intensity AR-5 17005 DBD	LFQ intensity AR-6 14449 DBD	LFQ intensity AR-7 220010 NTD	LFQ intensity AR-8 IgG
DBT	0	1415200	0	0	1101500	0	1275900	0
AES	0	1898300	5899800	6803500	5127300	1019700	0	0
SMARCC2	0	4827500	1734200	0	2153500	1246700	2950200	0
CSNK2A1;CSNK2A3	0	5958600	0	0	1843100	2791600	1243000	0
FOXA1	0	2937500	0	0	1403500	9725600	5042400	0
CSTF1	0	2440800	1131500	0	0	5409300	4841400	0
TAF15	0	5685100	0	1150700	1489900	1794000	5247600	0
VDAC1	0	2797300	0	0	1482200	5379300	2235400	0
ETFB	0	2389400	0	0	1014700	3563400	4984200	0
SMARCA4	0	1067500	0	1891200	3862500	1213700	1127400	0
FKBP5	0	2606600	0	0	9904400	3862300	2444200	0
DLD	0	2972700	0	0	6694400	0	1124300	0
MRPL24	0	522470	0	0	580380	0	2448200	0
ECH1	0	2797200	0	0	8261700	0	7694300	0
TRIM24	0	1661000	0	0	4517400	3344400	1600900	0
MRPL23	0	1505200	0	0	4633500	1980100	1647000	0
CS	0	538340	0	0	0	0	2940700	0
MAK16	0	2961700	0	0	0	1571500	0	0
RNF20	0	2510100	0	0	5095900	3545200	4613400	0
WDR55	0	1701600	0	0	4098200	4823300	0	0
TBL1XR1	0	1230200	0	0	0	4294100	1744200	0
ACO2	0	720440	0	0	1262500	0	2226600	0
PDXDC1	0	2107400	0	0	0	0	1010500	0
SREK1	0	1281400	0	0	1646400	3048300	1271800	0
PYCRL	0	4440000	5085900	0	4812500	6300100	4312100	0
CAT	0	7471900	0	0	1842300	0	1539200	0
TBL3	0	3051900	0	0	0	1126600	2445600	0
FIP1L1	0	2411500	0	0	2742700	5186900	2887400	0
ESRP1	0	935750	0	0	1456500	1595300	1091300	0
NKX3-1	0	2097600	0	0	3816500	4360100	938820	0
YLP1M1	0	3704400	3464400	0	4445400	3452000	3440700	0
EIF2S1	0	3703700	0	0	7152500	3585900	3933300	0
SUPT16H	0	6125200	0	0	7506800	1030300	6235200	0
NDUFA9	0	758030	0	0	1295900	0	1640800	0
PDCD6IP	0	4153800	0	0	6804200	3961800	4968900	0
SRRM1	0	8385000	8347900	0	8196900	6726300	8266800	0
RBM25	0	5106700	0	0	8469600	6809800	3897700	0
DAZAP1	0	8973400	0	0	1312300	9317500	1105700	0
SNRPD2	0	4673400	0	0	1205200	0	5118500	0
SNW1	0	4196800	0	0	9876300	0	5200400	0
OLA1	0	2089500	0	0	2686700	2339400	2478600	0
DDX18	0	1141000	0	0	0	2782400	1258700	0

MTA1	0	1922700	0	0	2594100	1833600	2375200	0
RPA3	0	8676700	0	0	2099500	0	9631200	0
CHCHD3	0	1030000	0	0	1664300	0	1953100	0
CUL4B	0	2470800	0	0	4748800	0	3689300	0
SMC3	0	4493300	0	0	5398300	3935700	5891900	0
PSMC2	0	3144900	0	0	3639500	2524200	4412700	0
PPIL4	0	4039200	0	0	5012000	6213300	2077600	0
MRPL13	0	2286000	0	0	2959300	0	4472700	0
CCT5	0	5316900	0	0	5643800	5059600	6533800	0
CPNE1	0	7757600	3334800	0	8235300	7213600	6196600	0
PSMC1	0	6678800	0	0	4952400	0	1640700	0
NHP2	0	3476700	0	0	0	5921800	5009200	0
HK1	0	3051300	0	0	2515200	0	6961500	0
ANXA7	0	2018300	0	0	0	3170200	3076900	0
RBBP4	0	1074200	8824000	0	7975200	6593100	9469900	0
BANF1	0	6145500	0	0	8415400	6989100	3379900	0
RPL14	0	1529900	0	0	1676200	1746900	1251400	0
CHD4	0	3737700	0	0	6659600	0	4677900	0
CCAR1	0	5012900	0	0	5354500	4927200	4879200	0
CIRBP	0	7310800	0	0	7567500	8099100	6232500	0
EPM2AIP1	0	4411100	0	0	0	5579700	7413400	0
KARS	0	3361000	2661400	0	3098300	0	4095000	0
DDX50	0	799030	0	0	2326900	0	0	0
MRPL14	0	1382600	0	0	2169300	1841400	0	0
WDR1	0	5789600	0	0	8764900	0	8009300	0
FLOT2	0	3115300	0	0	2453000	2975700	3561200	0
NARS	0	1201200	0	0	4412500	6718500	2348000	0
PRDX5	0	1053600	0	0	1006900	1485900	5328200	0
PSMA5	0	4472300	0	0	0	3341500	9426300	0
CCT4	0	7532300	0	0	7576500	0	1379700	0
ADAR	0	8504100	0	0	8494700	8869200	6469300	0
TCOF1	0	8506400	0	0	0	1505200	8655600	0
TUBB4A;TUBB6	0	983210	0	0	0	0	2733900	0
MORC2	0	610990	0	0	748930	941480	0	0
U2AF2	0	2510400	0	0	3453200	1652700	1800300	0
TOP2B	0	1319400	0	0	1327900	9980300	1294500	0
RANGAP1	0	3566700	0	0	4347400	0	5410500	0
LRRFIP1	0	1289000	0	0	0	0	3486800	0
PRPF4B	0	4769800	0	0	4307600	3965300	4566700	0
GART	0	1271400	0	0	0	0	3421600	0
RUVBL2	0	7174600	0	0	5225500	4410800	9600100	0
DBN1	0	2145300	0	0	0	0	5731200	0
LIMA1	0	973240	0	0	0	0	2589800	0
RTF1	0	1534000	0	0	0	0	4055200	0
SAFB	0	5462300	0	0	4284100	3554600	6546500	0
DDOST	0	1865400	0	0	2421600	0	2407900	0
HACD3	0	1321200	0	0	2032000	0	1330000	0
PSMD3	0	3365100	0	0	0	3241100	5294800	0
SAP18	0	1663300	0	0	1504900	1453600	1256300	0
FAF2	0	1153100	0	0	1267600	0	1651600	0
NACC1	0	1631700	0	0	2850600	0	1259800	0
GNAS;GNAO1;GNA	0	1681000	0	0	0	0	4222800	0
LUC7L3	0	3062300	0	0	7591600	0	0	0
SUGP2	0	7627000	0	0	5688400	7885800	5325100	0
GLO1	0	1433000	0	0	0	0	3525500	0
SARS	0	4473800	0	0	1905400	0	9055600	0
ABCD3	0	3186800	0	0	2748400	0	4943800	0
UBTF	0	6672100	0	0	3441300	5897200	6740600	0
PSMD12	0	1387600	0	0	0	0	3236200	0
PDHB	0	941700	0	0	0	0	2195000	0
CTNND1	0	9428000	0	0	5776000	0	1618800	0

SLC12A2	0	907310	0	0	0	0	2111900	0
SF3B4	0	3604100	0	0	4844600	0	3536500	0
U2SURP	0	3603000	0	0	4821500	0	3480800	0
PKP3	0	5181100	0	0	0	2593200	9300100	0
OGT	0	2334900	0	0	1581900	1405100	2346600	0
HNRNPH3	0	3211400	0	0	4796200	0	2530600	0
CANX	0	3030900	0	0	2999700	0	3892000	0
EIF2A	0	2499000	0	0	2478800	0	3160600	0
LIG3	0	3593200	0	0	0	4733900	3322800	0
CSE1L	0	2527300	0	0	1725700	1170500	2746300	0
NPLOC4	0	2509200	0	0	1722100	1376100	2500000	0
UTP15	0	3050700	0	0	0	3868600	2854000	0
CTTN	0	1527600	0	0	0	1436500	1907500	0
EMG1	0	4846900	0	0	0	5869800	4732400	0
PSMC5	0	5442600	0	0	0	0	1186000	0
LARP7	0	3820800	0	0	2399500	2867000	3039400	0
VPS35	0	8426600	0	0	2959700	0	1518000	0
P4HB	0	5754400	0	0	4339000	0	8023600	0
CRNKL1	0	1085200	0	0	1070700	0	1257500	0
WDR75	0	3979100	0	0	0	4002800	4510300	0
AP2B1	0	7121300	0	0	4469200	0	1073300	0
OSBPL8	0	3833900	0	0	3905600	0	4246700	0
FECH	0	844690	0	0	0	0	1795800	0
BUB3	0	6765500	0	5906600	0	3279600	5161300	0
HCFC1	0	6805600	0	0	4117000	2721700	7574400	0
11.Eyl	0	2123300	0	0	0	2104200	2354600	0
EIF1	0	5109000	0	0	0	0	1069100	0
ALDH18A1	0	1555800	0	0	0	0	3238800	0
SQSTM1	0	1361900	0	0	0	0	2828200	0
LASP1	0	1524900	0	0	0	0	3159200	0
ACIN1	0	8390800	0	0	4341300	5309700	7543800	0
SUB1	0	2845800	0	0	0	1818600	3958300	0
NPM3	0	4368700	0	0	0	0	8837700	0
LDHA	0	1062300	0	0	5855000	5437500	1011400	0
ANK3	0	1350200	2482200	0	1130900	1433100	2209800	0
SF3A1	0	1050100	0	0	8608800	0	1224800	0
COPB1	0	2364500	0	0	6485300	0	4018000	0
RPL31	0	2172100	0	0	2634800	0	1643600	0
TIAL1;TIA1	0	7832900	0	0	0	9130900	6295900	0
CEBPZ	0	4455400	0	0	0	5587100	3149400	0
FMR1	0	4186000	0	0	3624700	0	4552300	0
HSD17B10	0	1126300	0	0	0	0	2192100	0
PDIA6	0	502270	0	0	0	0	973020	0
MCM5	0	9188900	0	0	6026100	6843400	4836000	0
BCLAF1	0	8523100	0	0	7851000	0	8552400	0
GMPS	0	6737800	0	0	0	0	1289000	0
DCTN1	0	4137900	0	0	0	3794400	4086800	0
EIF4H	0	6110700	0	0	3560300	4117600	3953500	0
RACGAP1	0	1156700	0	0	0	0	2199400	0
RPRD1B	0	961880	0	0	0	0	1824100	0
COPZ1	0	1134500	0	0	0	0	2150300	0
HLTF	0	1065900	0	0	0	2019700	0	0
NIP7	0	2308100	0	0	0	3329100	1030400	0
TOMM70A	0	2509100	0	0	1828300	0	2898900	0
NOP9	0	1259400	0	0	0	1436600	932060	0
SURF6	0	2237800	0	0	0	0	4178900	0
KIF5B	0	9696200	0	0	4073200	0	1388300	0
MYL12A;MYL12B;M	0	5760400	0	0	0	0	1063900	0
PFDN2	0	2184400	0	0	0	0	4031000	0
GLRX3	0	3501300	0	0	2632400	0	3807200	0
SSB	0	2074700	0	0	6384700	1381500	1757300	0

POGZ	0	1690800	0	0	0	0	3059700	0
UBXN1	0	6406100	0	0	0	0	1157000	0
AP2A1	0	2857100	0	0	0	0	5151300	0
TFRC	0	1038000	0	0	0	0	1861700	0
WDR18	0	980230	0	0	0	0	1757300	0
PRCC	0	564130	0	0	0	0	1010900	0
MAD1L1	0	607690	0	0	0	0	1080000	0
EIF3F	0	5356600	0	0	6260400	0	3213900	0
ABCE1	0	2405400	0	0	1673500	0	2575300	0
POLR2B	0	1448700	0	0	0	0	2552400	0
EIF3D	0	5411000	0	0	0	3435600	6072400	0
DIDO1	0	1428400	0	0	0	0	2506800	0
SMARCA5	0	7295800	0	0	3442200	0	9359800	0
VTA1	0	8995400	0	0	4509200	4587700	6603000	0
HSPA4	0	2429700	0	0	0	0	4224100	0
SEC22B	0	5116400	0	0	2413000	0	6451800	0
LMAN1	0	3444300	0	0	0	0	5955000	0
CKMT1A;CKMT1B	0	7416600	0	0	6596100	0	6120500	0
RBM47	0	3965000	0	0	3825600	0	2954100	0
TNKS1BP1	0	8799400	0	0	5235900	0	9758400	0
EEA1	0	5753200	0	0	0	0	9742000	0
UTP18	0	2107700	0	0	0	3553000	0	0
RNH1	0	8911700	0	0	7531400	0	7472600	0
NUP155	0	9542000	0	0	4735200	0	1132700	0
DIAPH1	0	1622300	0	0	0	0	2710400	0
RCL1	0	1550200	0	0	0	0	2587900	0
PRKRA	0	1666800	0	0	985460	0	1793300	0
WIBG	0	355250	0	0	0	0	590670	0
EIF5	0	3556900	0	0	0	0	5835200	0
UBXN6	0	2100900	0	0	0	0	3434500	0
FLOT1	0	5345500	0	0	0	0	8737200	0
COPB2	0	6557900	0	0	4554900	0	6147500	0
ACSL3;ACSL4	0	2009800	0	0	1450400	0	1824100	0
SF3A3	0	1076700	0	0	4456800	5561900	7497300	0
STAT1	0	2830900	0	0	0	0	4580300	0
TPM1	0	712450	0	0	0	0	1148000	0
MTA2	0	604270	0	0	0	0	966880	0
MIA3	0	3219400	4157400	0	1114200	4856800	3122100	0
RAB7A	0	3445800	0	0	0	0	5438800	0
XPO5	0	2345200	0	0	1821900	0	1843200	0
STRAP	0	3609500	0	0	0	2657400	2963200	0
ESYT1	0	7401500	0	0	5868200	0	5644800	0
NME3	0	570880	0	0	0	0	885910	0
BCAS2	0	3248400	0	0	2422500	0	2613900	0
DDX42	0	8174400	0	0	6627300	0	6007300	0
NAMPT	0	9353200	0	0	5692100	0	8722300	0
KIAA1598	0	626050	0	0	0	0	962840	0
GALK2	0	1320300	0	0	0	0	2025200	0
RAD50	0	1062400	0	0	5704300	3846900	6702100	0
CSNK1A1;CSNK1A1	0	2253600	0	0	1525200	0	1916600	0
DNM1L	0	5764200	0	0	0	0	8798300	0
AP1M2	0	2098900	0	0	1437400	0	1753700	0
PRPF3	0	1685400	1249300	0	0	0	1305100	0
RAE1	0	1312500	0	0	0	0	1979400	0
RDX	0	1064800	0	0	4276000	0	1174700	0
UGDH	0	1220300	0	0	0	0	1815700	0
C7orf50	0	3126400	0	0	0	0	4638200	0
COG4	0	769340	0	0	0	0	1141200	0
EPS15L1	0	1269900	0	0	0	0	1883600	0
TXNL1	0	2019000	0	0	0	0	2992900	0
PSMD13	0	4364500	0	0	0	0	6433600	0

NKRF	0	1940300	0	0	0	2854600	0	0
THOC2	0	8168500	0	0	6000000	0	6004900	0
PLAA	0	2801000	0	0	0	0	4115600	0
PSMC4	0	5331300	0	0	0	0	7811300	0
SNRPA;SNRPB2	0	1005100	0	0	7247100	0	7427300	0
INTS3	0	1093400	0	0	0	0	1596300	0
HIST1H2AC;HIST3H	0	1475900	0	0	0	0	2140000	0
PSMB1	0	1414900	0	0	0	1084600	965470	0
SEC23IP	0	2103800	0	0	0	0	3035700	0
IQGAP2	0	1782800	0	0	0	0	2569600	0
CNOT1	0	3557300	0	0	2334900	0	2790900	0
TRIM25	0	1748700	0	0	9661600	0	1551000	0
RANBP2	0	1896200	0	0	1367200	0	1360700	0
PRPF31	0	5723100	0	0	0	4403500	3822200	0
COPG2	0	2402800	0	0	0	0	3451000	0
YARS	0	7020300	0	0	0	0	9977800	0
RRP1	0	2256100	0	0	0	0	3198700	0
POLDIP3;PDIP46	0	4231400	0	0	3782800	0	2204700	0
HUWE1	0	1640100	0	0	0	0	2320100	0
POLR1C	0	3326800	0	0	0	0	4691900	0
PDS5B	0	5133000	0	0	0	3462000	3728400	0
AARS	0	1731500	0	0	729960	3422400	2005900	0
ADD1	0	1064600	0	0	0	0	1488500	0
NRP1	0	589860	0	0	0	0	823200	0
DNAJC7	0	4400500	0	0	0	0	6115600	0
MFAP1	0	1725700	0	0	0	0	2396100	0
FXR2	0	8978900	4456600	0	0	0	7961600	0
RPS20	0	3370900	0	0	0	0	4632900	0
PRUNE2	0	5226500	0	0	0	0	7171300	0
ABCD1	0	599030	0	0	0	0	821700	0
TLN1	0	2115500	0	0	1813800	692420	2650300	0
NXF1	0	5436700	0	0	2625100	2181000	2625700	0
PRRC2A	0	2741000	0	0	1417400	0	2321800	0
PPP1R10	0	7697600	0	0	2932900	3255100	4292800	0
PFKL	0	1437700	0	0	0	0	1938900	0
USP10	0	6989200	0	0	0	2931300	6449500	0
GULP1	0	2661200	0	0	0	0	3560100	0
O2.Eyl	0	1286600	0	0	7334100	0	9860800	0
RAB11B;RAB11A	0	3225000	0	0	0	0	4304000	0
PRPF6	0	1015400	0	0	6476400	0	7063500	0
MCM2	0	4876900	0	0	0	0	6490700	0
ACP1	0	1747900	0	0	0	0	2315400	0
RRP9	0	1135200	0	0	0	7276500	7710700	0
FKBP3	0	2747800	0	0	0	0	3620100	0
AIP	0	1922700	0	0	0	0	2524000	0
STAU2	0	2170100	0	0	1392300	0	1428000	0
SEH1L	0	1132000	0	0	0	0	1471000	0
TCP1	0	5383100	0	0	0	0	6984000	0
CLIC1	0	2515000	0	0	0	0	3261300	0
SF3B6	0	1037800	0	0	0	0	1343800	0
SEC23A	0	2438300	0	0	0	0	3154900	0
USP5	0	1117700	0	0	6287400	0	8173100	0
SARNP	0	2062700	0	0	0	0	2668600	0
RPP30	0	731080	0	0	0	0	945660	0
API5	0	5488800	0	0	0	0	7089700	0
TOLLIP	0	5441500	0	0	0	0	6971000	0
CYFIP2;CYFIP1	0	961000	0	0	0	0	1219900	0
CORO1B	0	3753400	0	0	0	0	4719700	0
GARS	0	1952100	0	0	0	0	2435100	0
SEC16A	0	7795400	0	0	0	0	9721100	0
LBR	0	1040100	0	0	0	0	1291200	0

LRRC47	0	7841800	0	0	0	0	9720500	0
EIF3L	0	4830700	0	0	2435800	0	3531000	0
STEAP2	0	1758200	0	0	0	0	2164500	0
EIF4B	0	8876200	0	0	4088100	0	6823900	0
GTF2F1	0	1853600	0	0	0	0	2264000	0
CORO1C	0	884810	0	0	0	0	1076300	0
QARS	0	9281500	0	0	3388100	0	7876900	0
RRM1	0	1273200	0	0	0	0	1542100	0
NCBP1	0	3546500	0	0	0	0	4285700	0
PRKAR2A	0	1988300	0	0	0	0	2398800	0
SMU1	0	5895000	0	0	0	0	7096200	0
FAM98A	0	1732700	0	0	0	0	2079900	0
HNRNPUL1	0	4744200	0	0	3321900	0	2363000	0
THOC1	0	2880100	0	0	0	0	3450100	0
RPL23	0	2186400	0	0	0	0	2614800	0
CHERP	0	3177000	0	0	0	0	3797100	0
RPL21	0	2132100	0	0	0	0	2542300	0
CPSF3	0	1892300	0	0	0	0	2256200	0
CALD1	0	1050900	0	0	0	0	1251900	0
CMBL	0	4601000	0	0	0	0	5449300	0
SPTAN1	0	1026800	0	0	0	0	1213600	0
HNRNPUL2-	0	2884700	0	0	1068200	0	2317000	0
HDGF	0	2072700	0	0	0	9435400	1477800	0
NOMO1;NOMO3;N	0	622320	0	0	0	0	726630	0
NUP205	0	1082700	0	0	0	0	1260100	0
GBF1	0	1720000	0	0	0	0	1994200	0
KRR1	0	5507200	0	0	0	0	6364200	0
DYNC112	0	1305600	0	0	0	0	1497500	0
AIMP2	0	5535300	0	0	0	0	6331700	0
ATXN10	0	5620400	0	0	0	0	6417500	0
RSL24D1	0	2039400	0	0	0	0	2321000	0
RNF213	0	3983500	0	0	1804100	0	2691000	0
IPO4	0	1664300	0	0	0	0	1871800	0
DDB1	0	1218600	0	0	0	0	1356000	0
MTDH	0	1209100	0	0	3028300	3159500	7262500	0
IPO7	0	2903500	0	0	0	0	3224900	0
PSMC6	0	2118100	0	0	0	0	2352300	0
MAGEA10;MAGEA4	0	1674200	0	0	0	0	1855300	0
XPOT	0	1178900	0	0	3640600	3722800	5636100	0
MCM4	0	3504100	0	0	0	0	3852000	0
PSMD2	0	7605400	0	0	0	0	8313100	0
PAF1	0	2517300	0	0	0	0	2712600	0
EDF1	0	3318600	0	0	0	0	3573600	0
RBM12	0	4763300	0	0	0	0	5113400	0
CDC5L	0	9478000	0	0	0	0	1016700	0
CUL2	0	794260	0	0	0	0	846290	0
FOLH1;FOLH1B	0	1374800	0	0	0	0	1463800	0
APEX1	0	4284300	0	0	0	0	4540700	0
RECQL	0	6209800	0	0	0	0	6565700	0
TJP1	0	1611200	0	0	5029000	0	1199800	0
DCTN2	0	4095300	0	0	0	0	4295100	0
GLYR1	0	5664100	0	0	2255900	0	3672300	0
GPI	0	2399300	0	0	0	0	2506800	0
RPS21	0	3450600	0	0	0	0	3561800	0
DNAJC8	0	3576100	0	0	0	0	3689800	0
ADSL	0	2292400	0	0	0	0	2364800	0
EIF5B	0	5028300	0	0	0	0	5167800	0
RPL7L1	0	9094400	0	0	0	0	9341700	0
MYEF2	0	3111700	0	0	0	0	3194200	0
ATL3	0	1603500	0	0	0	0	1639000	0
UNC45A	0	8201900	0	0	0	0	8374700	0

MON2	0	1908500	0	0	0	0	1947700	0
H2AFY2	0	8761200	0	0	2044200	2056400	4833700	0
ZNF326	0	1466200	0	0	0	0	1492300	0
HNRNPLL;HNRPLL	0	1564000	0	0	0	0	1585400	0
FAM98B	0	5828900	0	0	0	0	5883000	0
ZC3HAV1	0	4534100	0	0	0	0	4558400	0
STAG2	0	2330100	0	0	0	0	2339600	0
SEC23B	0	1842800	0	0	0	0	1849800	0
NOC4L	0	2808100	0	0	0	0	2786900	0
COG3	0	2195700	0	0	0	0	2164900	0
CDC37	0	6619300	0	0	0	0	6514900	0
SUPT5H	0	2531800	0	0	0	0	2479000	0
EIF4G2	0	5929100	0	0	0	0	5804000	0
DHX37	0	1543300	0	0	0	0	1509600	0
MGEA5	0	1421700	0	0	0	0	1385100	0
PRKCD	0	2321600	0	0	0	0	2256400	0
PLCB4	0	1235900	0	0	0	0	1199400	0
RAD23B	0	2973900	0	0	0	0	2846200	0
DSG2	0	3027000	0	0	0	0	2884800	0
PML	0	1791300	0	0	0	0	1707000	0
MCM6	0	4384200	0	0	0	0	4173500	0
PRKAR1A	0	3344700	0	0	0	0	3170600	0
IPO5	0	2084800	0	0	0	0	1969100	0
PWP2	0	5826300	0	0	0	2336100	3149500	0
MYO1C	0	5922700	0	0	0	0	5572100	0
DDX41	0	1587700	0	0	0	0	1489700	0
TBCD	0	1039700	0	0	0	0	969500	0
ANP32E	0	9513700	0	0	0	0	8844300	0
CMAS	0	1824900	0	0	0	0	1691600	0
SCRIB	0	2171400	0	0	0	0	2000100	0
AIDA	0	2065500	0	0	0	0	1901400	0
PPP1CA;PPP1CC	0	3261600	0	0	0	0	3000300	0
CCDC124	0	4141200	0	0	0	0	3800800	0
CKAP5	0	6166800	0	0	0	0	5649800	0
DNM2;DNM1	0	5524500	0	0	2057200	0	2985600	0
DDX31	0	4116900	0	0	0	0	3756500	0
NUP62	0	3294700	0	0	0	0	3004300	0
PSMD1	0	2906000	0	0	0	0	2631500	0
PHF2;PHF8	0	1858300	0	0	0	0	1681000	0
SKIV2L	0	722660	0	0	0	0	651470	0
STAT3	0	1142000	0	0	0	0	1029100	0
USP14	0	2939000	0	0	0	0	2644500	0
IQGAP1	0	5433000	0	0	0	0	4888400	0
EIF6	0	1186600	0	0	0	0	1066800	0
NAP1L4	0	804000	0	0	0	0	721080	0
AHNAK	0	4549100	0	0	5231000	3650800	3187700	0
TRIP11	0	2156200	0	0	0	0	1913900	0
MYH14	0	2733000	0	0	0	0	2422400	0
SYNE2	0	1108800	0	0	0	0	980380	0
CDC73	0	3034800	0	0	0	0	2664200	0
EXOSC7	0	3470100	0	0	0	0	3026400	0
CHAMP1	0	4305600	0	0	0	0	3741600	0
UTRN	0	1441700	0	0	0	0	1247700	0
H1FX	0	5986700	0	0	0	0	5116300	0
SERBP1	0	1020000	0	0	0	0	8698100	0
VPS13C	0	436760	0	0	0	0	369290	0
MOV10	0	1811300	0	0	0	0	1523900	0
PRPF4	0	4838200	0	0	0	0	4038100	0
NUP88	0	4636600	0	0	0	0	3846900	0
NUP160	0	2447400	0	0	0	0	2018200	0
OSBP	0	1519200	0	0	0	0	1242500	0

EXOSC2	0	3308100	0	0	0	0	2697800	0
NSUN5	0	1324600	0	0	0	0	1077600	0
PPP1R12A	0	3797800	0	0	0	0	3082300	0
RBM15	0	1568700	0	0	0	0	1271200	0
C14orf166	0	1440600	0	0	0	3885900	7656500	0
ZC3H11A	0	1329000	0	0	0	0	1059400	0
TBL2	0	2406600	0	0	0	0	1917200	0
ZNF638	0	1641900	0	0	0	0	1306500	0
GOLGA3	0	3015000	0	0	0	0	2395800	0
LRBA	0	2651900	0	0	2837300	2276600	1595600	0
STT3A	0	3118700	0	0	0	0	2477200	0
FAM120C	0	2594900	0	0	0	0	2057200	0
EDC4	0	8592600	0	0	0	0	6808600	0
USP9X;USP9Y	0	2961800	0	0	0	0	2320100	0
TRMT112	0	1372800	0	0	0	0	1074400	0
DNAJB1	0	1954700	0	0	0	0	1513100	0
PFDN4	0	2479400	0	0	0	0	1916800	0
SNRNP40	0	3895300	0	0	0	0	3007100	0
GTF2E2	0	736550	0	0	0	0	556590	0
SART3	0	5176400	0	0	0	0	3901200	0
PSMB5	0	1216000	0	0	0	0	915770	0
AQR	0	2500200	0	0	0	0	1852900	0
EWSR1	0	7066600	0	0	0	0	5214700	0
GOLGB1	0	7577700	0	0	0	0	5573000	0
DNTTIP2	0	3561700	0	0	0	0	2606500	0
STUB1	0	7620300	0	0	0	0	5572400	0
PBRM1	0	9485100	0	0	0	0	6930500	0
RCC1	0	1268400	0	0	0	0	9174600	0
SPTBN2	0	2154900	0	0	0	0	1555400	0
RPS17	0	5848600	0	0	0	0	4135300	0
ARFGEF3	0	2394100	0	0	0	0	1687500	0
SEC24B	0	1495400	0	0	0	0	1046300	0
NUP133	0	1240900	0	0	0	0	8643400	0
RFC1	0	3043500	0	0	0	0	2093100	0
PNN	0	5475300	0	0	0	0	3744800	0
HDGFRP2;HDGFRP3	0	6744500	0	0	0	0	4571400	0
G3BP2	0	6721900	0	0	0	0	4548200	0
TSR1	0	1468200	0	0	0	0	992810	0
ABCF1	0	3919300	0	0	0	0	2642500	0
CUL5	0	1788300	0	0	0	0	1193500	0
YTHDC2	0	1365100	0	0	0	0	904590	0
GIGYF2	0	9028800	0	0	1958100	0	3932100	0
ECM29;KIAA0368	0	2752900	0	0	0	0	1777000	0
MACF1	0	2304300	0	0	0	0	1484800	0
MSI2;MSI1	0	1763600	0	0	0	0	1129300	0
ZNF598	0	1020100	0	0	0	0	652770	0
EIF2AK2	0	1487800	0	0	0	0	941600	0
BZW1	0	2460500	0	0	0	0	1549300	0
ETF1	0	5816400	0	0	0	0	3637200	0
CLUH	0	8230600	0	0	0	0	5069700	0
NUP214	0	7078100	0	0	0	0	4348700	0
FUBP3	0	6106300	0	0	0	0	3751500	0
DEK	0	5556300	0	0	0	0	3398100	0
PURA	0	3764000	0	0	0	0	2297900	0
CPSF7	0	1056100	0	0	0	0	6409500	0
EIF3E	0	4226000	0	0	0	0	2544700	0
SRP72	0	5612400	0	0	0	0	3271600	0
RNF40	0	1536600	0	0	0	0	889490	0
CAPRIN1	0	1671900	0	0	0	0	9571700	0
ABCF3	0	2129900	0	0	0	0	1213600	0
PELP1	0	4584500	0	0	0	0	2558300	0

NOC3L	0	3609100	0	0	0	0	1941400	0
LARP1	0	6140300	0	0	0	0	3289100	0
MRT04	0	4472300	0	0	846470	1529700	0	0
MCTS1	0	1768900	0	0	0	0	933120	0
TUBA4A	0	2479200	0	0	6865000	0	6072000	0
GSTM3	0	2397700	0	0	0	0	1249600	0
SRP54	0	4614200	0	0	0	0	2389800	0
USP39	0	2352300	0	0	0	0	1209000	0
DHX16	0	879410	0	0	0	0	450930	0
NELFB	0	1207500	0	0	0	0	616280	0
SNRPG;SNRPGP15	0	8298200	0	0	0	0	4232200	0
DRG1	0	3027000	0	0	0	0	1508300	0
LANCL1	0	2286900	0	0	0	0	1124500	0
BUD31	0	1952000	0	0	0	0	957650	0
GBE1	0	2187600	0	0	0	0	1052800	0
NSF	0	4601000	0	0	0	0	2096600	0
SCYL1	0	1535300	0	0	0	0	695360	0
PTPN11	0	1974800	0	0	0	0	884400	0
RTCA	0	727270	0	0	0	0	322610	0
POLR2A	0	4159400	0	0	0	0	1842500	0
TCERG1	0	5338500	0	0	0	0	2318600	0
RBM8A	0	3477200	0	0	0	0	1478200	0
AKAP1	0	8210600	0	0	0	0	3462700	0
ARVCF	0	1759200	0	0	0	0	731910	0
NUP98	0	3313200	0	0	0	0	1227600	0
AHCTF1	0	2714600	0	0	0	0	853980	0
AP2M1	0	3711100	0	0	0	0	1160500	0
RPS19	0	3735600	0	0	0	0	1113800	0
KRT79	0	3921100	0	0	0	0	0	0
PPAN-	0	4436400	0	0	0	0	0	0
GDI2	0	4317800	0	0	0	0	0	0
LAS1L	0	4109400	0	0	0	0	0	0
PDAP1	0	3771800	0	0	0	0	0	0
THOC6	0	3331200	0	0	0	0	0	0
VAPA	0	3317400	0	0	0	0	0	0
COG2	0	3094000	0	0	0	0	0	0
MDC1	0	2977300	0	0	0	0	0	0
POLR2C	0	2871100	0	0	0	0	0	0
CAD	0	2851800	0	0	0	0	0	0
SSR4	0	2705900	0	0	0	0	0	0
DDX39A	0	2676900	0	0	0	0	0	0
PAFAH1B2	0	2619500	0	0	0	0	0	0
CLASP1	0	2513200	0	0	0	0	0	0
SDSL	0	2467100	0	0	0	0	0	0
PPM1G	0	2453700	0	0	0	0	0	0
TCF25	0	2423800	0	0	0	0	0	0
SMARCC1	0	2379800	0	0	0	0	0	0
PHC2;PHC3	0	2222200	0	0	0	0	0	0
SNX9	0	2200800	0	0	0	0	0	0
FYTTD1	0	2197200	0	0	0	0	0	0
DUSP3	0	2128300	0	0	0	0	0	0
ANKFY1	0	2100000	0	0	0	0	0	0
ILVBL	0	2089900	0	0	0	0	0	0
ATRX	0	2032900	0	0	0	0	0	0
NVL	0	2012700	0	0	0	0	0	0
MAP1S	0	1987400	0	0	0	0	0	0
CDC42BPB	0	1977000	0	0	0	0	0	0
MBNL1	0	1920200	0	0	0	0	0	0
MPHOSPH6	0	1848100	0	0	0	0	0	0
BZW2	0	1767000	0	0	0	0	0	0
CCDC47	0	1721600	0	0	0	0	0	0

TNPO1	0	1703700	0	0	0	0	0	0
UFM1	0	1692500	0	0	0	0	0	0
SBDS	0	1673800	0	0	0	0	0	0
RBM17	0	1668500	0	0	0	0	0	0
NLE1	0	1668400	0	0	0	0	0	0
ARFGAP3	0	1644600	0	0	0	0	0	0
EML4	0	1636200	0	0	0	0	0	0
VCL	0	1618500	0	0	0	0	0	0
BTF3	0	1616000	0	0	0	0	0	0
TMEM33	0	1532800	0	0	0	0	0	0
FN1	0	1510000	0	0	0	0	0	0
SYMPK	0	1506200	0	0	0	0	0	0
THUMPD1	0	1349300	0	0	0	0	0	0
SMAD3	0	1310000	0	0	0	0	0	0
BOLA2	0	1277400	0	0	0	0	0	0
AATF	0	1274700	0	0	0	0	0	0
SEN3;SEN3-	0	1271400	0	0	0	0	0	0
CPSF1	0	1269500	0	0	0	0	0	0
DDX47	0	1258600	0	0	0	0	0	0
GTF3C5	0	1220200	0	0	0	0	0	0
RAB2A;RAB4B;RAB	0	1218000	0	0	0	0	0	0
RCOR1;RCOR3	0	1209300	0	0	0	0	0	0
NRBP1	0	1197300	0	0	0	0	0	0
PEBP1	0	1183500	0	0	0	0	0	0
BAIAP2	0	1171300	0	0	0	0	0	0
CUL1	0	1162700	0	0	0	0	0	0
HSPA14	0	1156900	0	0	0	0	0	0
SNX1	0	1128700	0	0	0	0	0	0
CC2D1A	0	1121600	0	0	0	0	0	0
GTF3C2	0	1113300	0	0	0	0	0	0
ARPC4;ARPC4-	0	1112400	0	0	0	0	0	0
GSE1	0	1066900	0	0	0	0	0	0
UBE2D1;UBE2D4	0	1050600	0	0	0	0	0	0
DMAP1	0	1049700	0	0	0	0	0	0
SUPT6H	0	984630	0	0	0	0	0	0
LSM1	0	983200	0	0	0	0	0	0
SLAIN2	0	930760	0	0	0	0	0	0
WDR5	0	916790	0	0	0	0	0	0
BMS1	0	904080	0	0	0	0	0	0
VWA9	0	897520	0	0	0	0	0	0
TPT1	0	891970	0	0	0	0	0	0
DNAJA2	0	869740	0	0	0	0	0	0
COIL	0	836170	0	0	0	0	0	0
GEMIN5	0	822540	0	0	0	0	0	0
NCAPD2	0	802240	0	0	0	0	0	0
CLASP2	0	801070	0	0	0	0	0	0
GOLGA1	0	784590	0	0	0	0	0	0
DPCD	0	783860	0	0	0	0	0	0
KRT73;KRT4	0	775430	0	0	0	0	0	0
MECP2	0	757550	0	0	0	0	0	0
LPP	0	753510	0	0	0	0	0	0
SETD3	0	732970	0	0	0	0	0	0
FLII	0	725990	0	0	0	0	0	0
NUP188	0	719960	0	0	0	0	0	0
USP15	0	685400	0	0	0	0	0	0
ANXA5	0	660430	0	0	0	0	0	0
PHAX	0	650440	0	0	0	0	0	0
HECTD4	0	590840	0	0	0	0	0	0
CBLL1	0	562600	0	0	0	0	0	0
EIF4G3	0	525340	0	0	0	0	0	0
ACTR1B	0	515300	0	0	0	0	0	0

PTPRF	0	509040	0	0	0	0	0	0
DHFR	0	507720	0	0	0	0	0	0
CD3EAP	0	495880	0	0	0	0	0	0
TM7SF2	0	485790	0	0	0	0	0	0
DOCK7	0	482660	0	0	0	0	0	0
PQBP1	0	418070	0	0	0	0	0	0
WDHD1	0	412450	0	0	0	0	0	0
MTAP	0	411470	0	0	0	0	0	0
EMC2	0	401390	0	0	0	0	0	0
RRP7A	0	399820	0	0	0	0	0	0
DDX19A;DDX19B	0	387190	0	0	0	0	0	0
RAB14	0	352570	0	0	0	0	0	0
ERBB2IP	0	350820	0	0	0	0	0	0
EIF2B5	0	350760	0	0	0	0	0	0
INTS1	0	348930	0	0	0	0	0	0
XIAP	0	338680	0	0	0	0	0	0
PPP2R5C;PPP2R5D	0	323620	0	0	0	0	0	0
ARAF	0	299820	0	0	0	0	0	0
SF3A2	0	293850	0	0	0	0	0	0
TOP2A	0	279460	0	0	0	0	0	0
PIK3CD;DKFZp779K	0	279450	0	0	0	0	0	0
TTC38	0	217100	0	0	0	0	0	0
ZNF519;ZNF721;ZN	0	214480	0	0	0	0	0	0
HYOU1	0	209310	0	0	0	0	0	0
ANKZF1	0	166490	0	0	0	0	0	0
EIF4E2	0	153720	0	0	0	0	0	0
KIAA1468	0	102830	0	0	0	0	0	0

Appendix Table B.2: List of proteins that appear to be interrupted by each small molecule, as represented in the Venn Diagram in Figure 5.1

Names	total
13789_LBD_BF3 14449_DBD 17005_DBD 220010_NTD Enzalutamide_LBD	121
13789_LBD_BF3 14449_DBD 17005_DBD Enzalutamide_LBD	292
13789_LBD_BF3 17005_DBD 220010_NTD Enzalutamide_LBD	4
13789_LBD_BF3 14449_DBD 220010_NTD Enzalutamide_LBD	2
13789_LBD_BF3 17005_DBD Enzalutamide_LBD	32
13789_LBD_BF3 14449_DBD Enzalutamide_LBD	85
13789_LBD_BF3 220010_NTD Enzalutamide_LBD	4
13789_LBD_BF3 14449_DBD 17005_DBD	2
13789_LBD_BF3 Enzalutamide_LBD	62
17005_DBD Enzalutamide_LBD	1
13789_LBD_BF3 17005_DBD	1
13789_LBD_BF3 14449_DBD	1
Enzalutamide_LBD	2
13789_LBD_BF3	8
14449_DBD	1
220010_NTD	1

Proteins from published Endogenous AR RIME analysis (Stelloo et al. 2018) that are disrupted by each drug in our preliminary RIME data analysis;

[AR_RIME] and [Enza_LBD]: SMARCA4

[AR_RIME] and [13789_LBD]: SMARCC2

[AR_RIME] and [Enza_LBD] and [13789_LBD]: CSE1L, EIF4H, HCFC1

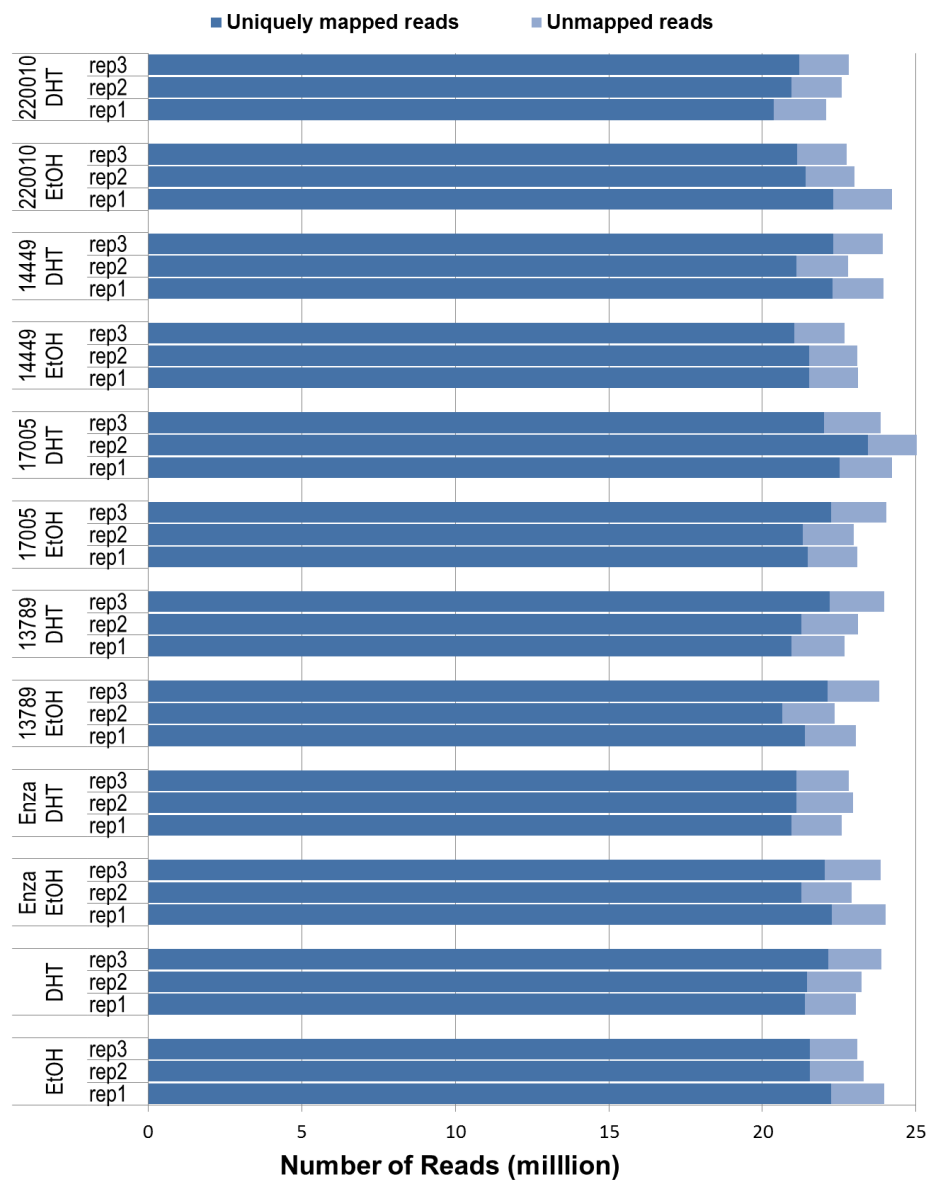
[AR_RIME] and [13789_LBD] and [17005_DBD]: CSTF1

[AR_RIME] and [Enza_LBD] and [13789_LBD] and [14449_DBD]: BCLAF1, CHD4, PRPF6

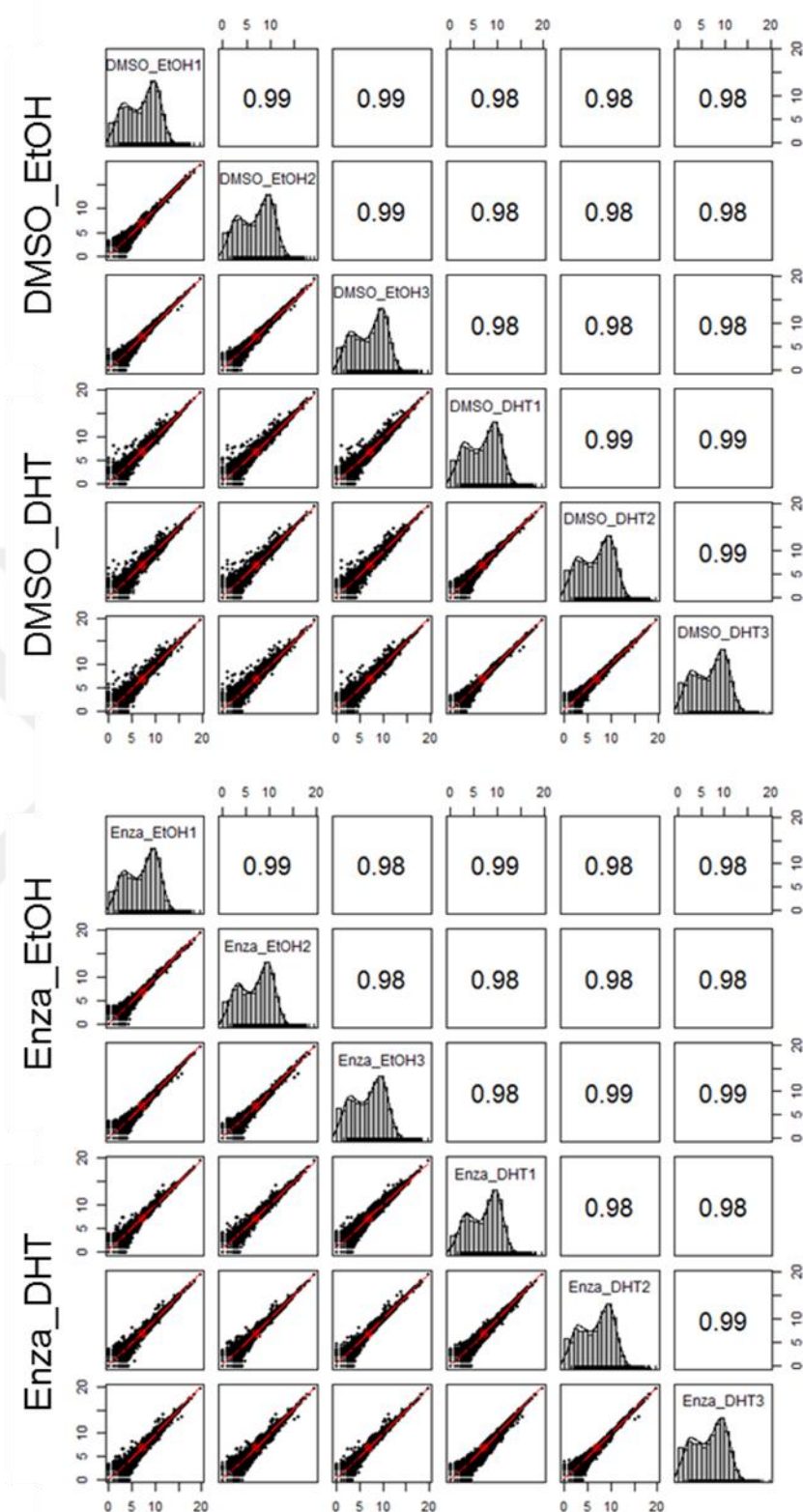
[AR_RIME] and [Enza_LBD] and [13789_LBD] and [17005_DBD] and [14449_DBD]: APEX1, EWSR1, RBM15, RCC1, RNF40, RPS20

[AR_RIME] and [Enza_LBD] and [13789_LBD] and [17005_DBD] and [14449_DBD] and [220010_NTD]: SMARCC1, WDHD1

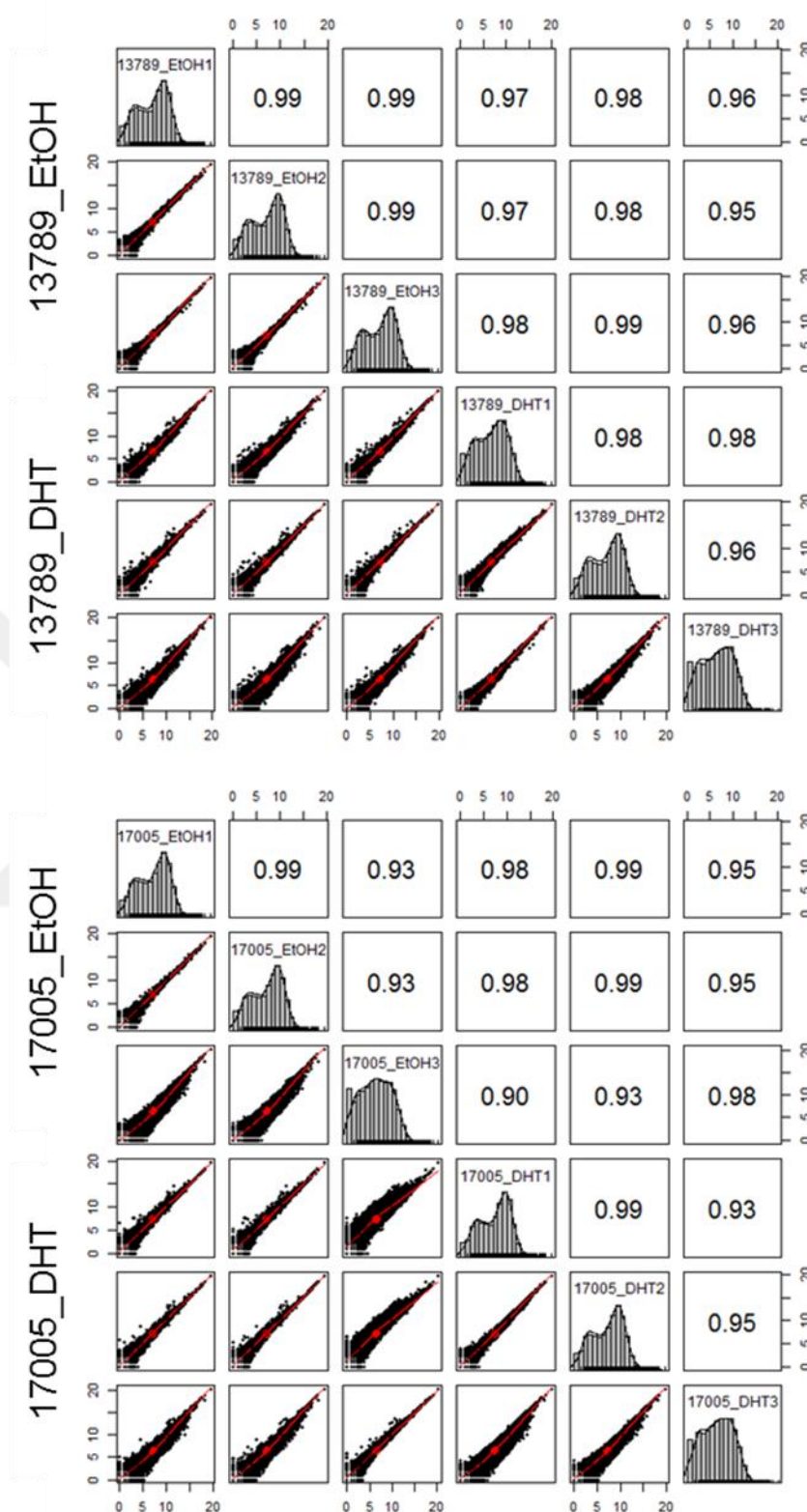
Analysis performed using InteractiVenn (Heberle et al. 2015)

RNA sequencing Analysis

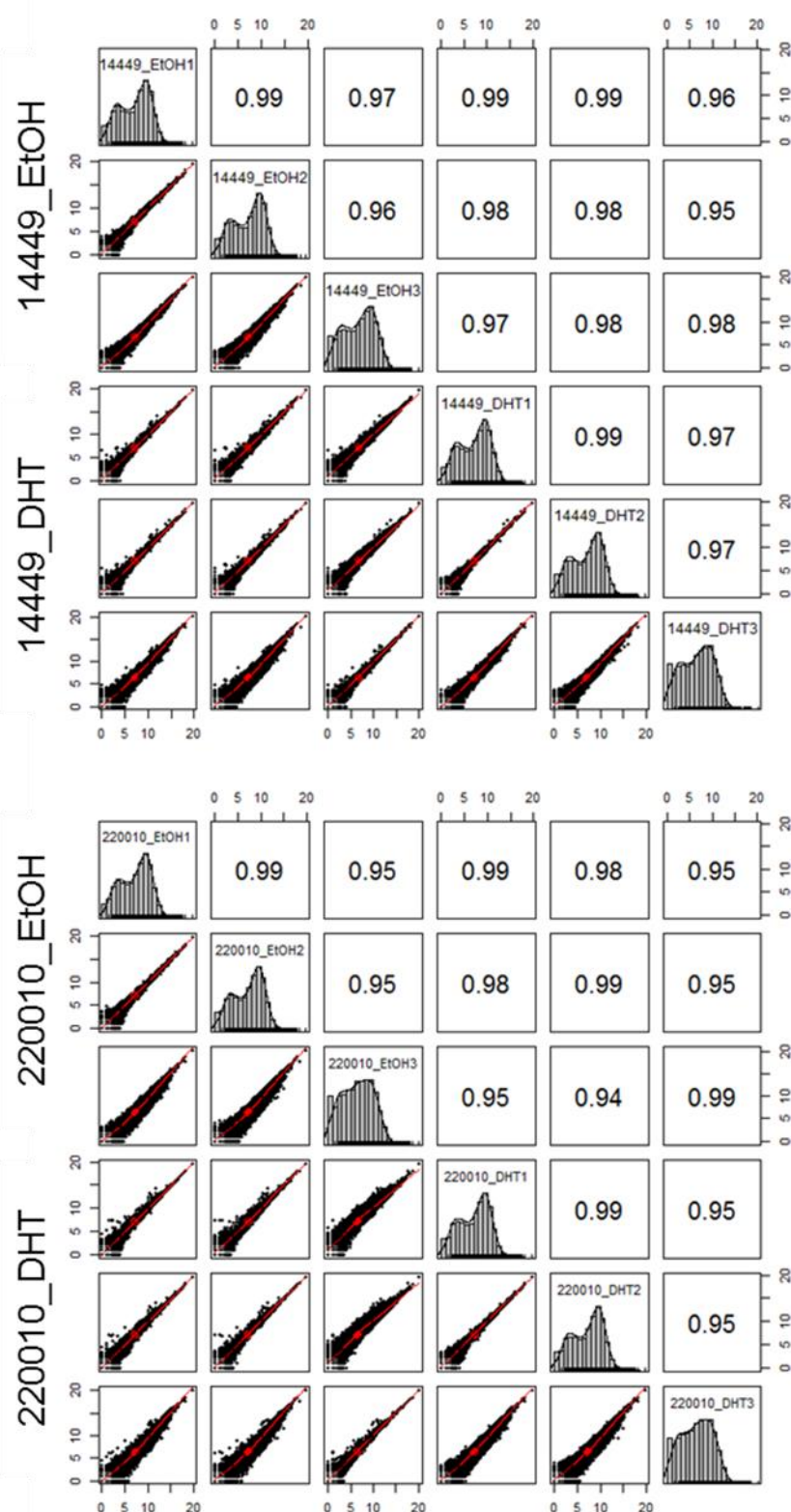
Appendix Figure B.1: The sequence reads had high ratios of uniquely mapped reads with an average of about 93%.



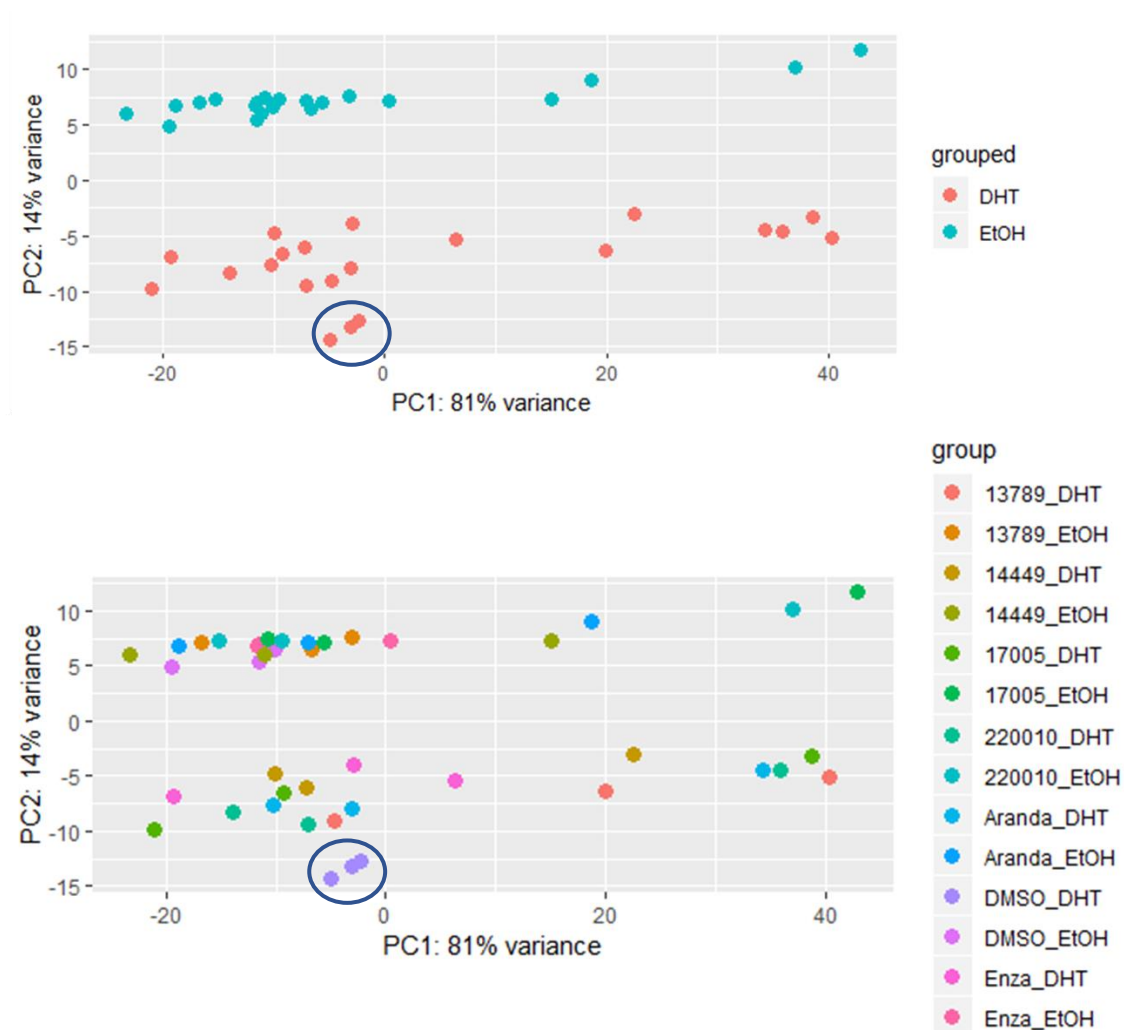
Appendix Figure B.2: Biological triplicates of DMSO controls and Enzalutamide treated samples had high Pearson correlation values.



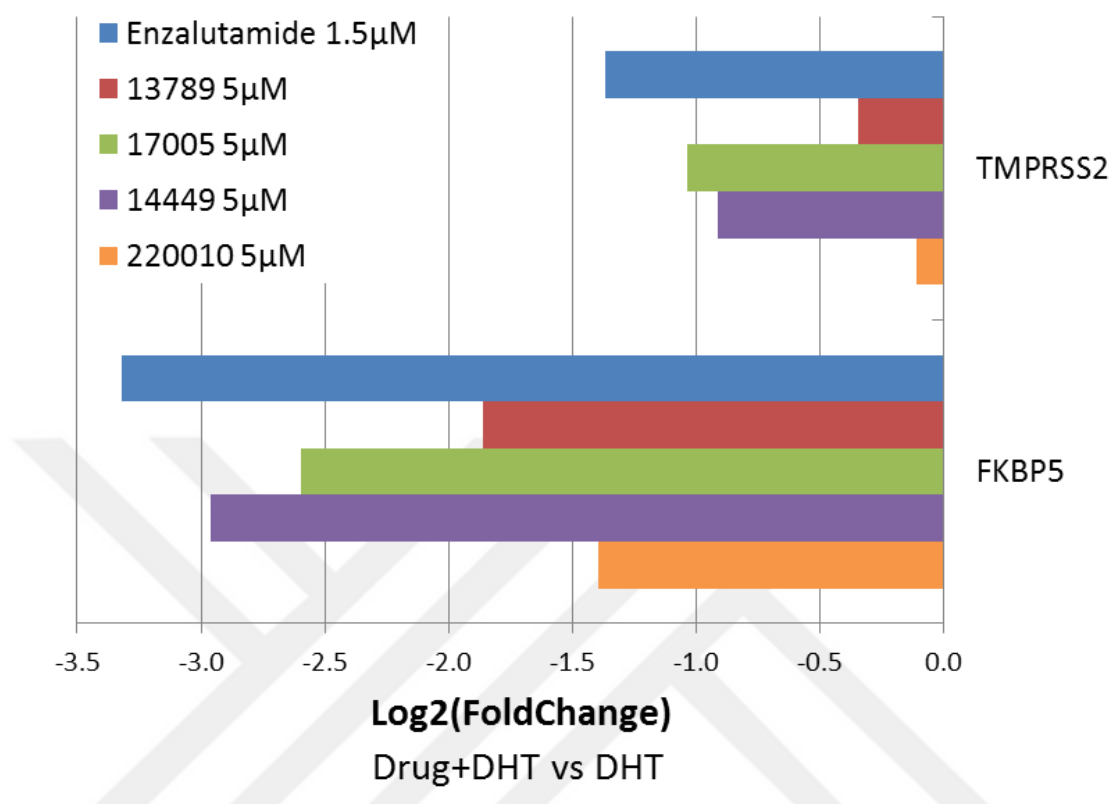
Appendix Figure B.3: Biological triplicates of 13789 and 17005 treated samples had high Pearson correlation values.



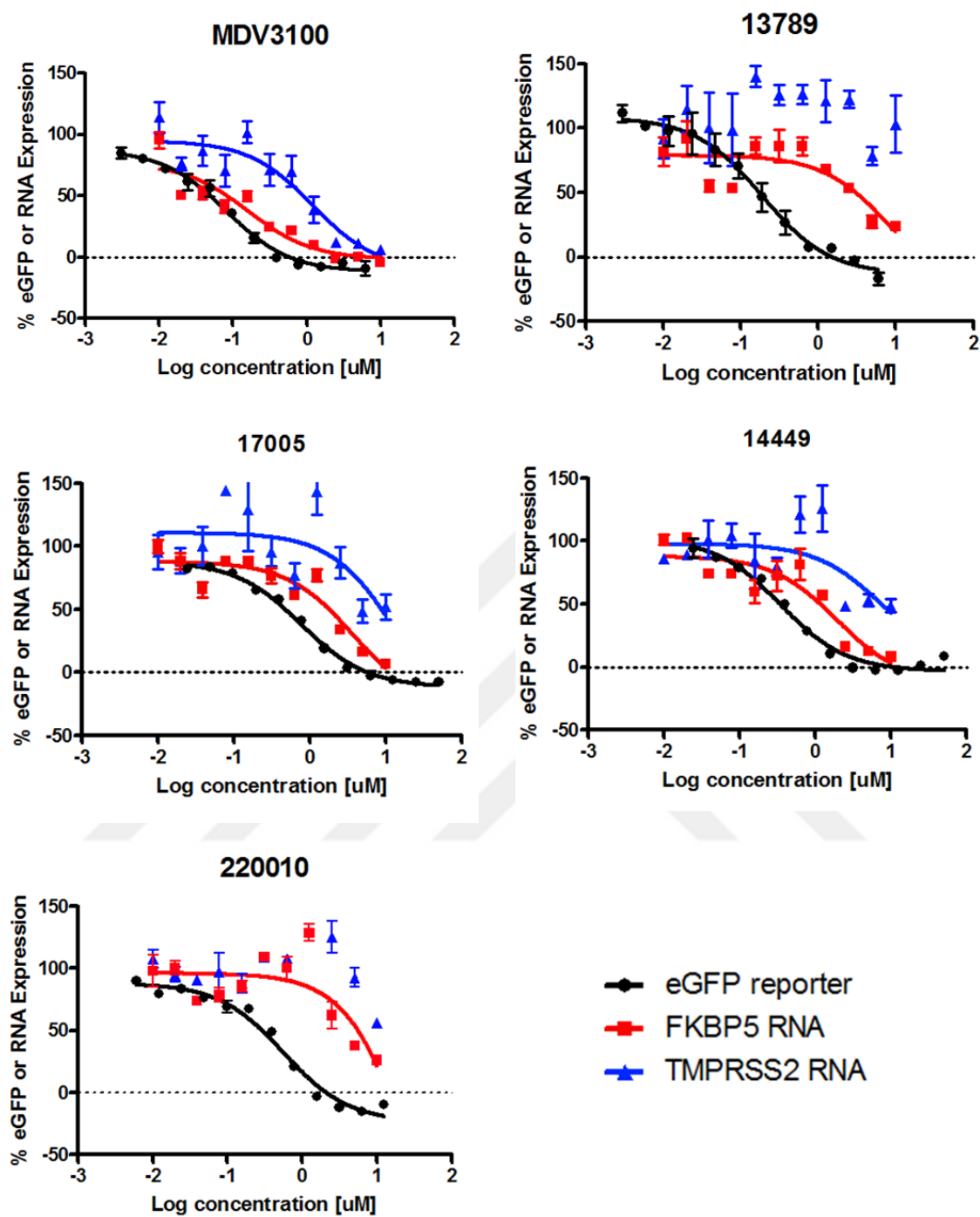
Appendix Figure B.4: Biological triplicates of 14449 and 220010 treated samples had high Pearson correlation values.



Appendix Figure B.5: Principle Component Analysis (PCA) showed that the DHT treated controls (DMSO_DHT) clustered together, the DHT + small molecule treated samples clustered separately, and the EtOH + small molecule treated samples clustered as a separate group. Replicas of each small molecule treatment did not cluster together, however, all drugs caused a trend towards the EtOH controls compared to the DMSO_DHT controls.



Appendix Figure B.6: RT-qPCR analyses of key AR-regulated genes at compound doses closest to selected from the dose response curves below in Appendix Figure B.7. Gene responses observed in the qPCR analyses correlate with RNA sequencing results described in the results section.

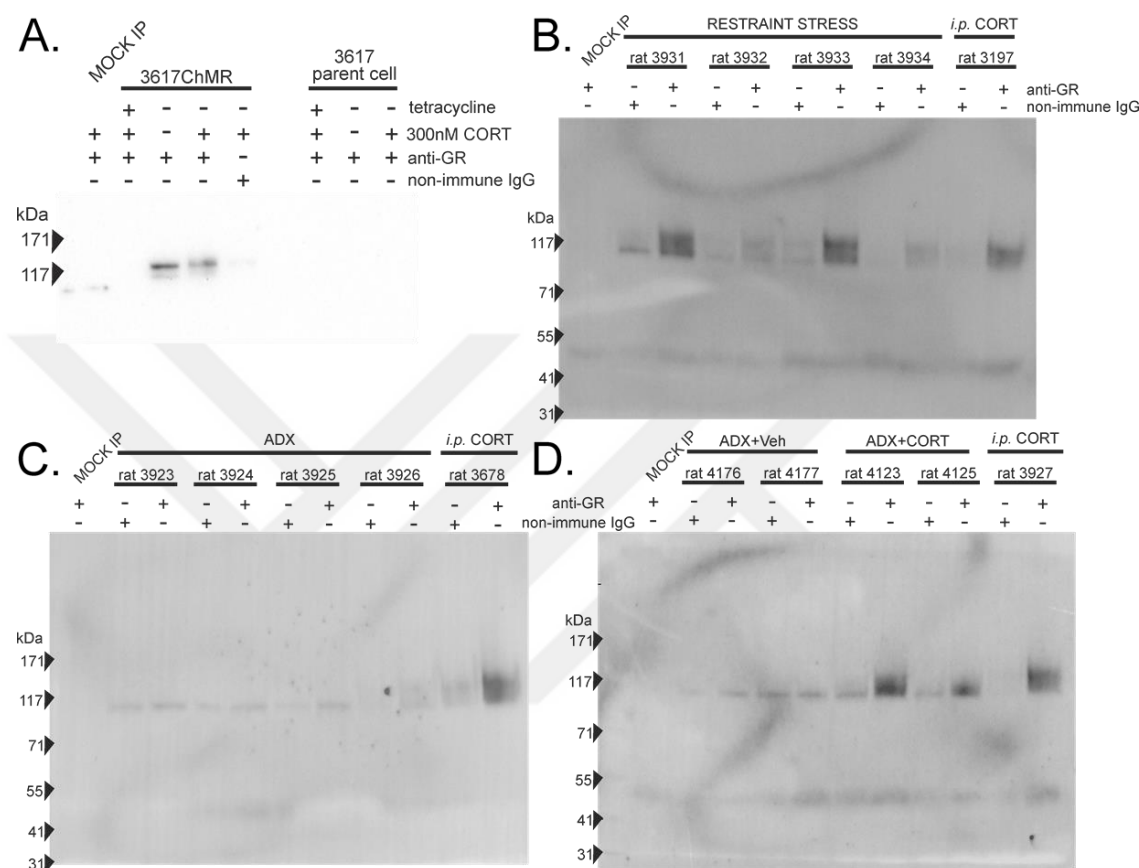


Appendix Figure B.7: RT-qPCR analyses of dose response of key AR-regulated genes to treatment with different compounds at different doses.

Appendix C

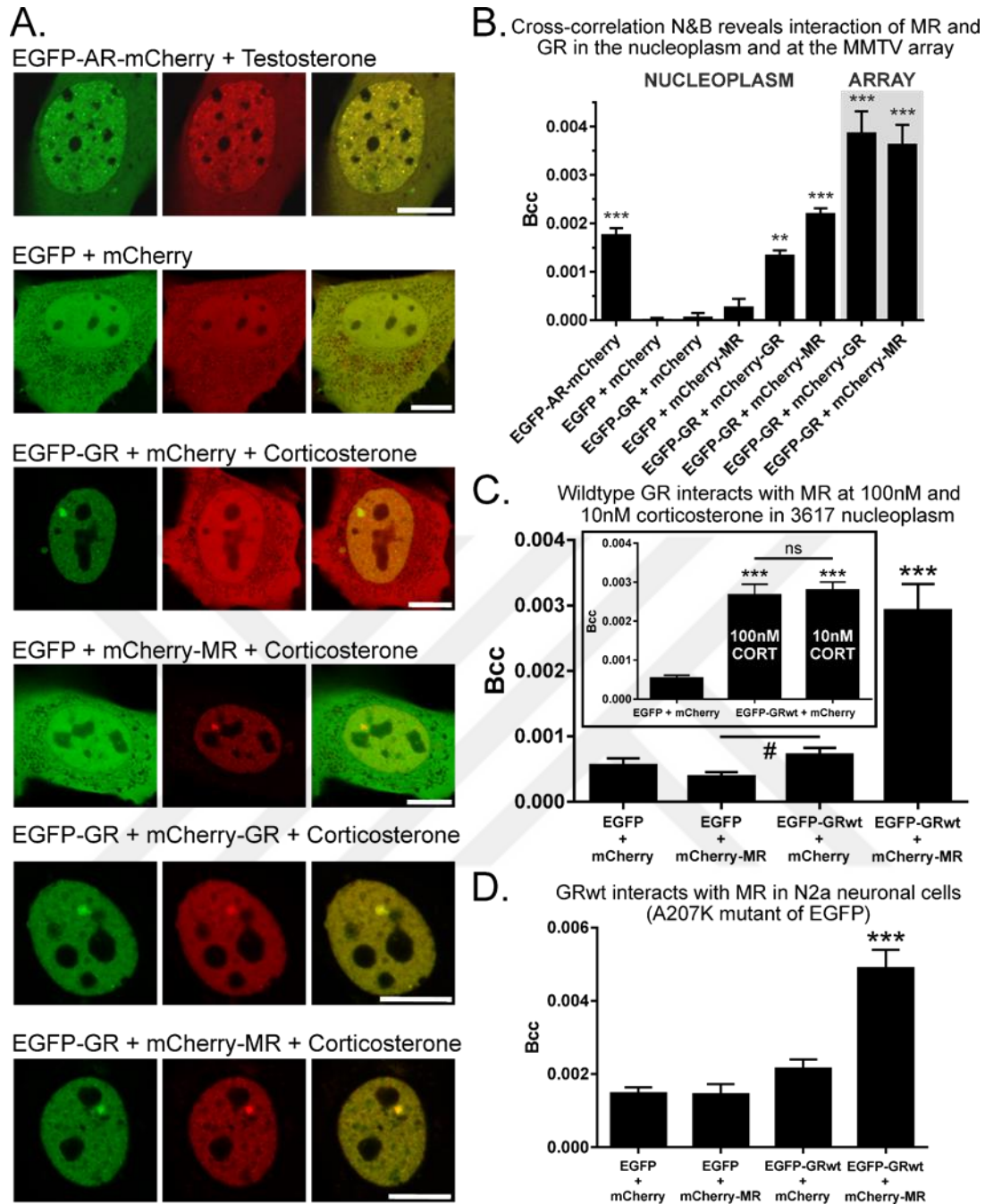
Oligomerization Potentials of Steroid Receptors

Glucocorticoid and mineralocorticoid receptors interact in hippocampus and cells



experiments performed by Dr. John Pooley from the University of Bristol

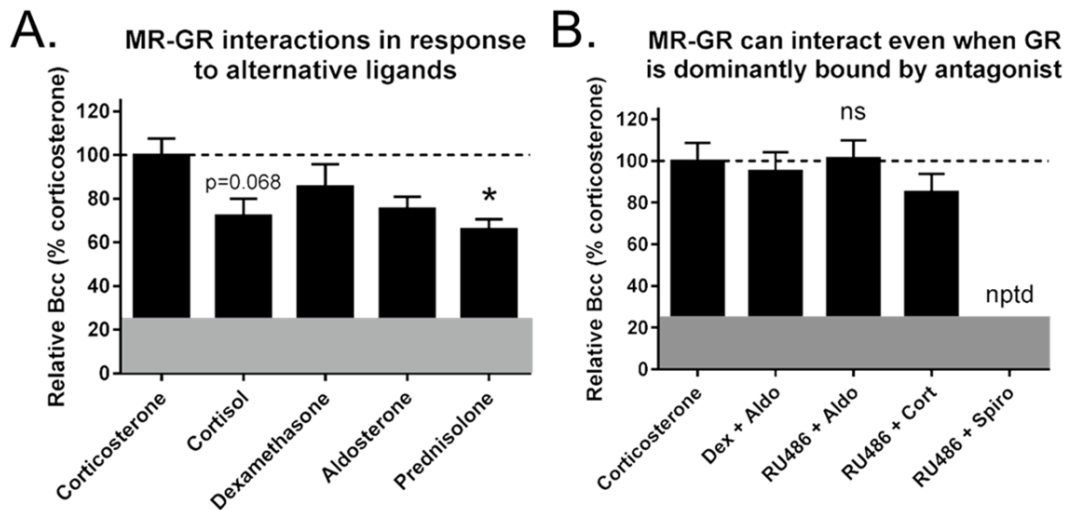
Appendix Figure C.1. Co-immunoprecipitation supports GR-MR interaction in a cell line and whole rat hippocampus. (A) GR immunoprecipitation pulls down mCherry-MR when cells are treated with vehicle or corticosterone. Negative controls are IgG antibody for IP and a MOCK IP with no lysate. (B) Immunoprecipitation of rat GR from whole hippocampus captures MR supporting interaction following stress. (C) MR does not co-immunoprecipitate with GR when animals are bilaterally adrenalectomized (ADX). Compare signal to positive control (i.p. CORT) animal. (D) Intraperitoneal corticosterone injection (i.p., 3 mg/kg) of adrenalectomized (ADX) rats restores MR co-immunoprecipitation indicating corticosterone dependence.



experiments performed by Dr. John Pooley from the University of Bristol

Appendix Figure C.2. Cross-correlation number and brightness assay (ccN&B) reveals GR-MR interactions in 3617 cell nucleus and at chromatin. 3617 cells +tetracycline transfected with constructs expressing the proteins indicated 24 hrs before imaging. Use of GR_{C656G} mirrored co-immunoprecipitation experiments. Nuclear localisation was driven by at least 30 min treatment with 1 μ M testosterone (EGFP-AR-

mCherry) or 300 nM corticosterone (all other conditions). **(A)** Representative cells (scale bars 10 μ m). **(B)** An androgen receptor (AR) tagged at one end with EGFP and at the other with mCherry provided a positive interaction control. AR does not bind the MMTV array but nucleoplasmic Bcc values were significantly higher than negative controls EGFP + mCherry, EGFP-GR_{C656G} + mCherry, and EGFP + mCherry-MR (N = 12-35 cells per control group). Bcc values were additionally obtained from GR+GR and MR+GR pairings in nucleoplasm (N = 47-65 cells per group) and at the MMTV array (N = 40-63 cells). Mean $\pm$ SEM. **(C)** Interaction of mCherry-MR and EGFP-GR_{wt} at 100 nM corticosterone in 3617 nucleoplasm. Welch's F test showed Bcc values between groups were significantly different ($F(3,48.17) = 14.52$, $p < 0.001$), with Games-Howell post hoc testing. Mean $\pm$ SEM, N = 24-25 cells per group. Inset compares Bcc values obtained using 100 nM or 10 nM corticosterone. Welch's F test revealed group means were significantly different ($F(2,35.6) = 77.46$, $p < 0.001$). There was no significant difference (ns) between groups with different corticosterone doses. Mean $\pm$ SEM, N = 24-27 cells per group. **(D)** Neuronal cell line N2a shows significant MR-GR interaction in the nucleoplasm relative to negative controls. Welch's F test revealed group means were significantly different ($F(3,56.44) = 14.95$, $p < 0.001$), with Games-Howell post hoc test. Mean $\pm$ SEM, N = 26-29 cells per group. EGFP_{A207K} prevents EGFP dimerization at higher concentrations. Post hoc tests indicate means significantly different from all negative control group(s) where ** $p < 0.01$, *** $p < 0.001$. # Significantly different from each other, $p = 0.023$.



Experiments performed by Dr. John Pooley from the University of Bristol

Appendix Figure C.3. GR and MR can interact independent of bound ligand.

(A) Experimental comparison of Bcc values determined by ccN&B for MR-GR where different endogenous or synthetic ligands have been applied to activate both receptors (100 nM). The Bcc value for corticosterone (known interaction) was set to 100%. The greyed-out region represents the maximum mean Bcc value obtained for non-interacting controls in previous experiments. **(B)** Experimentally determined Bcc values for combined treatments expressed relative to 100 nM corticosterone (known interaction). Doses were varied to attain dominant occupation of each receptor by the appropriate ligands. Dexamethasone 10 nM + aldosterone 10 nM, spironolactone + RU486 (1 μ M each), aldosterone + RU486 and corticosterone + RU486 (10 nM MR-targeted agonist, 1 μ M GR-targeted antagonist). Mean $\pm$ SEM, N = 20-23 cells per group.

*Significantly different, $p < 0.05$; ns, not significant; nptd, not possible to determine.

Appendix D

RNA-Seq Analysis of Cytochrome Deletion in p53 Mutants***Supplementary Methods***

The cell culturing, UV treatment, RNA extraction, RNA-seq library preparation, and RNA-seq quality check, and qPCR validation procedures described below were performed by Dr. Mehmet Tardu from the laboratory of Prof. Dr. Halil İ. Kavaklı in Koç University.

Cell culturing and UV-treatment

Mouse skin fibroblast (MSF) cells from $p53^{-/-}$ knockout and $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ mice in a C57BL/6J background were kindly provided from the laboratory of Aziz Sancar (University of North Carolina at Chapel Hill). Knockout MSF cells were maintained and grown in Dulbecco's Modified Eagle Medium (DMEM) supplemented with penicillin at 50 U/ml, streptomycin at 50 µg/ml, and 10% fetal bovine serum. The cells were maintained in an incubator at 37 °C under 5% CO₂.

For UV-treatment assays, cells were seeded in six-well plates at 4×10^5 cells per well and incubated in growth medium for 16-20 h. Then, the cells were washed with 1x PBS (Phosphate Buffered Saline) and exposed to UV light (General electric germicidal lamp GT5, emitting mainly at 254 nm) for 40 seconds, at a fluence of 40 J.m⁻² with a fluence rate of 1 J.m⁻².s⁻¹. UV light intensity was measured using a UVX Digital Radiometer (UVP, USA) at 264 nm. After UV-treatment, the cells were washed with 1xPBS, growth medium was added, and cells were incubated for 16 hours. After 16 hours incubation, cells were harvested for RNA extraction followed by RNA-seq library preparation. Non UV-treated cells were used as control samples. RNA-seq was performed using two biological replicates for each of the conditions shown in **Figure 3.4.A**.

RNA extraction and RNA-seq library preparation

Total RNA was extracted from 8×10^5 UV-treated and UV non-treated cells using RNeasy mini kit (Qiagen, USA) according to the manufacturer's instructions, and subjected to poly-A selection. The quality and quantity of RNA samples were assessed using Nanodrop 2000 spectrophotometer (Thermo Scientific, USA) and 2100 BioAnalyzer (Agilent, USA). Only the RNA samples with OD_{260/280} between 1.8 and 2.2, 28S:18S ≥ 1 and the RNA integrity number (RIN) ≥ 8 were used for library construction. After quality and quantity measurements, to remove residual genomic DNA, samples were treated with RNase-free DNase I (Thermo Scientific, USA) at a concentration of 1 U/ μ g. Then, RNA-seq library preparation was performed using TruSeq mRNA Sample Preparation Kit (Illumina, USA) as described in previous work (Tardu et al. 2016; Sarayloo et al. 2017). Briefly, polyadenylated RNAs were purified from the 1 μ g of total RNA using oligo (dT) magnetic beads and fragmented using fragmentation buffer. After that, the cleaved short RNA fragments were used as a template for first-strand cDNA synthesis using first strand synthesis mix, and the second strand was synthesized using second strand marking master mix. AMPure XP beads (Beckman Coulter, USA) were used to clean up the double stranded cDNAs. Then, 3' end adenine nucleotides were added and sequencing adaptors were ligated to the cDNA fragments followed by enrichment of libraries by PCR amplification. Enriched cDNA libraries were used for cluster generation and sequencing. 100X2 paired-end sequencing of four cDNA libraries with two biological replicates were performed using the Illumina HiSeq 2500 sequencing platform (Illumina, USA).

RNA sequence quality check

All sequence data were 2 $\times$ 100 bp in length. The high quality reads were saved in fastq files and deposited in the Gene Expression Omnibus (GEO) database at NCBI with the accession number GSE115102. The fluorescent images were processed into sequences, and base calling and quality value calculations were performed using the Illumina data processing pipeline (v1.5). The following steps were used for sequencing quality controls: (1) removing reads with adaptor contamination; and (2) removing low-quality bases (quality score < Q30) using Trimmomatic (v0.35) (Bolger et al. 2014). All downstream analyses were based on clean, high quality data.

Validation of RNA-seq results using qRT-PCR

We randomly selected 11 representative down- and up-regulated differentially expressed genes for validation of RNA-seq using qRT-PCR under identical conditions where we performed RNA-seq experiments. Total RNA was extracted from 8×10^5 UV non-treated and UV-treated cells using RNeasy mini kit (Qiagen, USA) according to the manufacturer's instructions. After DNase I treatment, 1 μ g of total RNA was used for first-strand cDNA synthesis using random hexamer oligos. qRT-PCR was performed with Luminaris HiGreen qRT-PCR Master Mix (Thermo Scientific, USA) on the CFX Connect Real-Time PCR Detection System (Biorad, USA). qRT-PCR was performed with gene-specific primers (Appendix Table D.1). *Gapdh* housekeeping gene whose expression level didn't change across UV(-) and UV(+) cells was used as an internal reference gene. The amplification programs were performed according to the following protocol: 95 °C for 10 min; 40 cycles of 95 °C for 2s, 56 °C for 10s and 72 °C for 10s, and followed by a thermal denaturing step to generate the melting curves. All qRT-PCR reactions were performed in biological triplicates (each triplicate with two technical replicates), and the results were expressed relative to the transcription level of *Gapdh* in each sample by using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001). The mRNA expression data analyzed by using IBM SPSS statistical 20.0 (SPSS Inc., USA). All data were presented as the means $\pm$ S.D. (n = 6) of the relative mRNA expression.

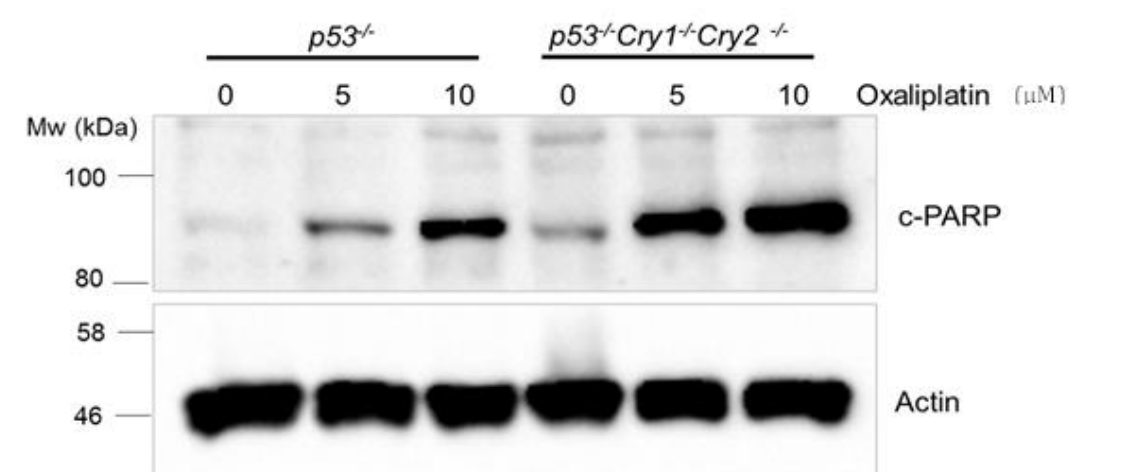
The qPCR validations were carried out by Dr. Mehmet Tardu from the laboratory of Prof. Dr. Halil İ. Kavaklı in Koç University, and by Tuba Korkmaz and Assoc. Prof. Nuri Öztürk in Gebze Technical University;

Table D.1: Primers used for qRT-PCR Analyses.

Bdkrb1-F	CCATCAGTCAGGACCGCTAC
Bdkrb1-R	CAGGGACGACTTTGACGGAA
Cst6-F	ACAGAGTGCCGAAAGACCAG
Cst6-R	GAGCTCAAAGTTGCAACGCA
Egr3-F	CAACGACATGGGCTCCATTC
Egr3-R	GGCCTTGATGGTCTCCAGTG
Gapdh-F	TGGCAAAGTGGAGATTGTTGCC
Gapdh-R	AAGATGGTGATGGGCTTCCCCG
H3F3a-F	GTGTGTAGGGGAAAGCCTCG
H3F3a-R	ACGAGCCATGGTAAGGACAC
Hoxa1-F	GATCCTGGAGTGATGTGGTCC
Hoxa1-R	AACCACTGTAGTCCAGCCAAA
Lce1f-F	GATCTTGTGCTGTCCACAGTCT
Lce1f-R	CTTTGGAGTCTGGCACTTGG
Mcpt8-F	CTGGTGGCAAGAGACATCGT
Mcpt8-R	GAGCTTTGCGTTCCAGCTTC
Polr2j-F	ACTAAGGTTCCCAACGCCTG
Polr2j-R	GGGTGAGGGACTTTGTAGCC
Prl2c3-F	CAACTCCAGTAAAGCATCTTCCC
Prl2c3-R	CGACCATTCCTCATTGCACAC
Spr1a-F	GAGAACCTGCTCTTCTCTGAGT
Spr1a-R	GCTGAGGAGGTACAGTGCAG
Ubac1-F	GATGTTCTGTCAGGAGGAGAAGAT
Ubac1-R	TGGGTCACATTCTTGGGGTCT

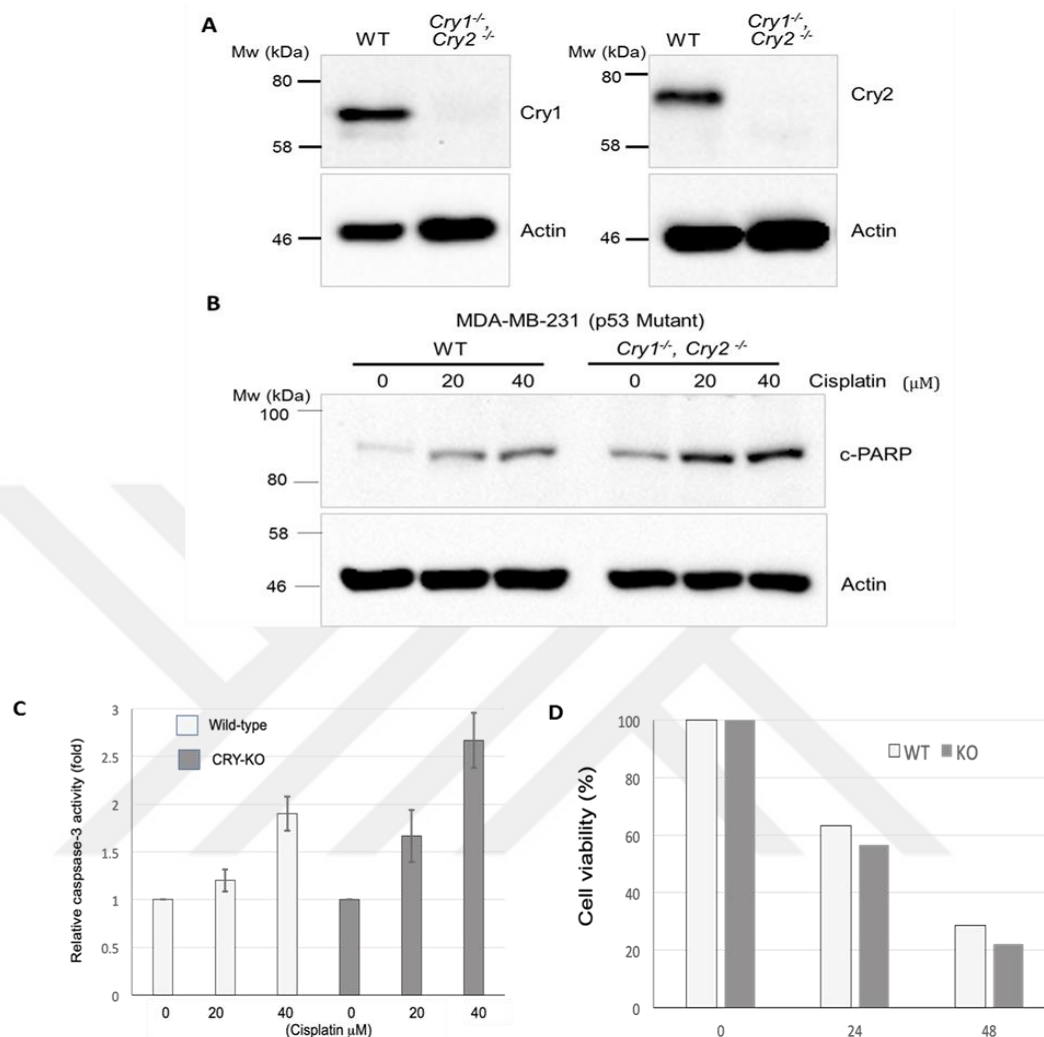
p21-F 5'-TGA GCC GCG ACT GTG ATG-3'
p21-R 5'-GTC TCG GTG ACA AAG TCG AAG TT-3'
Egr1-F 5'CTTCAACCCTCAGGC GGACA3'
Egr1-R 5'GGAAAAGCGGCCAGTATAGGT3'
Casp4-F 5'AGCGTTGGGTTTTTGTAGATGC3'
Casp4-R 5'CCTTGTGAACTCTTCAGGGGA3'
Jun-F 5'GTGTGGGACGACGATCAAAAG3'
Jun-R 5'TGACCACTAACAGGGAAGGAC3'
Mmp9-F 5'GCAGAGGCATACTTGTACCG3'
Mmp9-R 5'TGATGTTATGATGGTCCCACTTG3'
Platr5-F 5' GCTTTGAAACCCTCCTGC3'
Platr5-R 5'CAGCACACACACCTTTGG3'
Platr17-F 5'ATTTGCCAGGGTCCC3'
Platr17-R 5'CACAACCATCCTTAATG3'
Sirt1-F 5'TGATTGGCACCGATCCTCG3'
Sirt1-R 5'CCACAGCGTCATATCATCCAG3'
Casp1-F 5'TAGGGGAAAGTGCGAGTTTCA3'
Casp1-R 5'GGGCAATCCAGCATTTACCT3'
Tnfaip3-F 5'GCGGCCACAGAAAACACTC3'
Tnfaip3-R 5'CTCCCAATGGTCAAGGCATC3'
Myc-F 5'CCGCTCAAGTTGCTCGAAAAG3'
Myc-R 5'TCTCCTTGTAAGACATTGCTGAC3'

These experiments were carried out by Tuba Korkmaz and Assoc. Prof. Nuri Öztürk in Gebze Technical University;

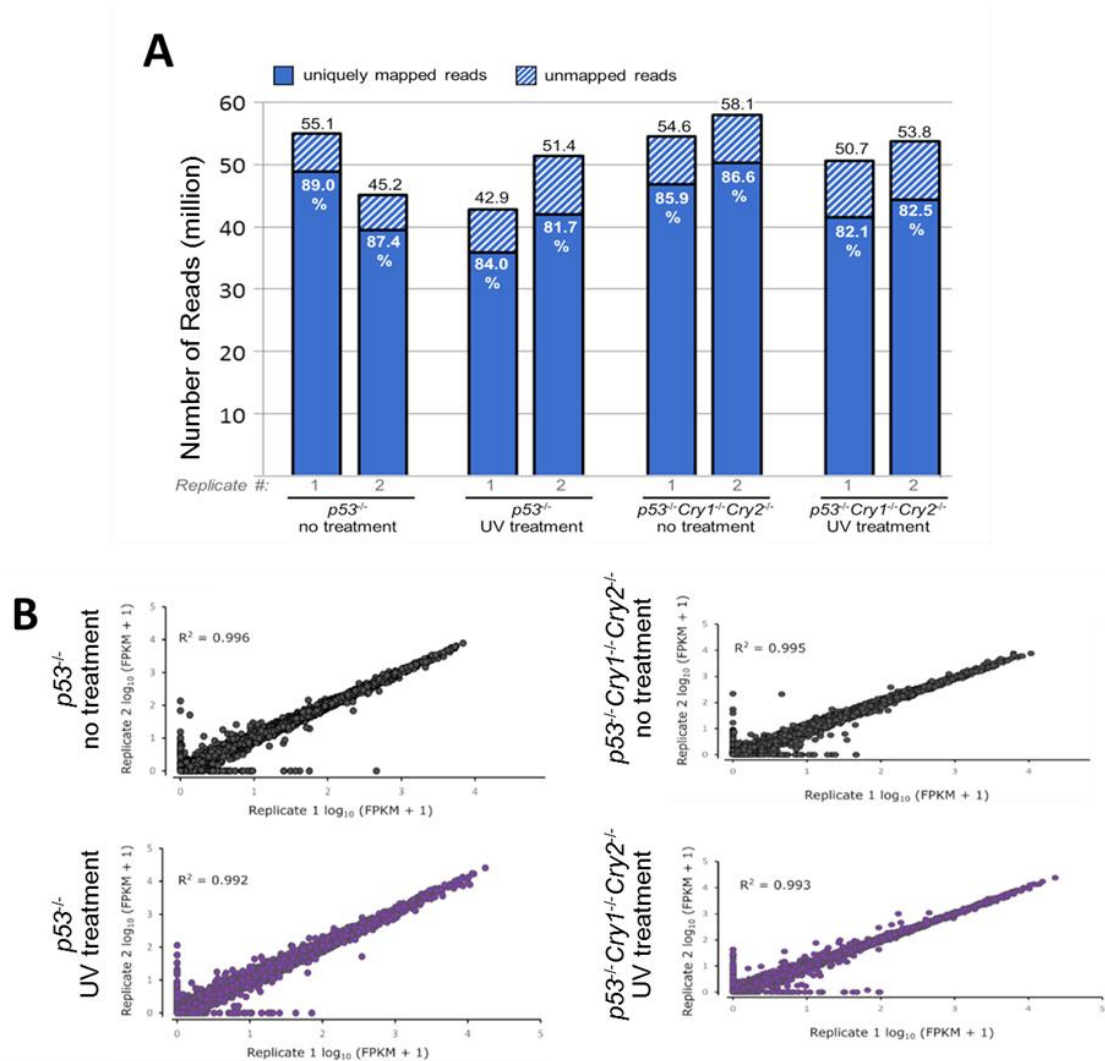


Appendix Figure D.1: Confirmation of the effect of cryptochrome deletion on apoptosis induced by UV-mimetic cisplatin in mouse skin fibroblasts. In order to confirm the authenticity of the fibroblasts used in this study, *p53*^{-/-} or *p53*^{-/-}*Cry1*^{-/-}*Cry2*^{-/-} skin fibroblasts were treated with oxaliplatin (at 0, 10 or 20 μM doses), and cell lysates were probed for cleaved PARP (c-PARP) by immunoblotting. Actin served as a loading control. These experiments were carried out by Tuba Korkmaz and Assoc. Prof. Nuri Öztürk in Gebze Technical University. (Cavga et al. 2019)

These experiments were carried out by Tuba Korkmaz and Assoc. Prof. Nuri Öztürk in Gebze Technical University;

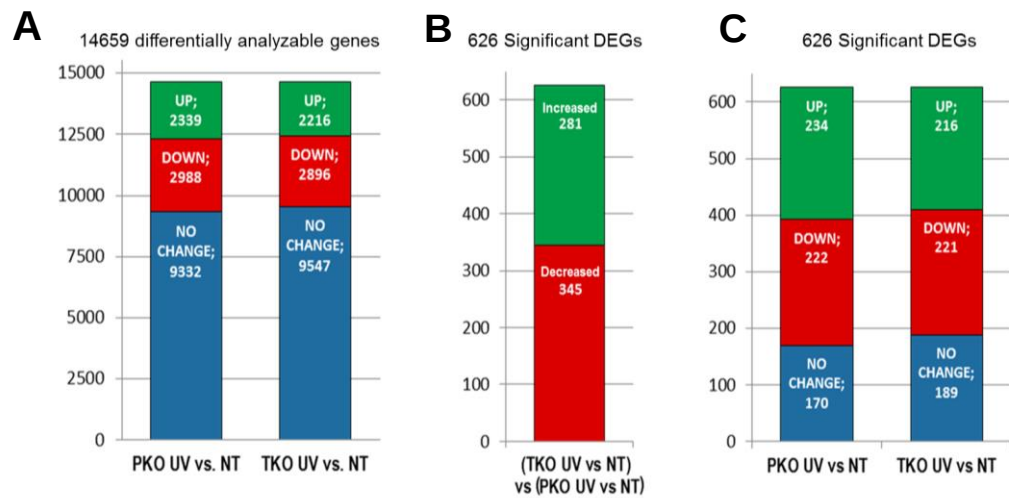


Appendix Figure D.2: Effect of cryptochrome deletion on apoptosis induced by cisplatin in breast cancer cell line MDA-MB-231. **A)** Confirmation of Cryptochrome deletion by immunoblotting. Immunoblot of Cry1, and Cry2 in MDA-MB-231 derived clone confirmed the knockout. Actin was blotted as a loading control. **B)** Effect of Cryptochrome deletion on apoptosis. Cells were treated with 0, 20, and 40 μ M cisplatin as UV-mimetic for 16 h. Cell lysates were probed for cleaved PARP (c-PARP) by immunoblotting. Actin was blotted as a loading control. **C)** Measurement of Caspase-3 activity in WT and CRY-KO MDA-MB-231 cells treated with Cisplatin. Cells were treated with 0, 20, 40 μ M Cisplatin for 18 hours. Caspase-3 activity in total cell lysates was measured using the caspase-3 substrate, Z-DEVD-R110. The activity was normalized to total protein concentration and presented as fold activity to 0 mM treated samples. **D)** Cell viability assay where cells were treated with 40 μ M Cisplatin for 24 or 48 hours. These experiments were carried out by Tuba Korkmaz and Assoc. Prof. Nuri Öztürk in Gebze Technical University. (Cavga et al. 2019)



Appendix Figure D.3: RNA sequencing quality and content assessment. A)

Mapping and coverage of RNA-seq reads. B) Expression comparison between 2 biological replicates for all 4 experimental groups. Correlation plots for transcript-level $\log_{10}$ (FPKM+1) values between replicates of $p53^{-/-}$ no treatment (PKO_NT), $p53^{-/-}$ UV treatment (PKO_UV), $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ no treatment (TKO_NT) and $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ UV treatment (TKO_UV) cells. Pearson R^2 values for the correlation are included as an insert in the plot.



Appendix Figure D.4: A) Up- and down-regulation status of the 14659 differentially analyzable genes, in $p53^{-/-}$ and $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ cells. A ‘change’ is considered as above 2 fold, and significant according to Cufflinks. Genes below these cutoffs are labeled ‘no change’. B) The increased or decreased response status of the 626 selected significant DEGs in terms of their DEG Score= $\log_2[(TKO_UV/TKO_NT)/(PKO_UV/PKO_NT)]$. C) The up or down-regulation status of the 626 selected significant DEGs in terms of their fold changes as PKO_UV/PKO_NT and TKO_UV/TKO_NT

Table D.2: STRING input list of 487 genes.

A4galt	Cd1d1	ENSMUSG0	Hic2	Mamdc2	Peg12	Selrc1	Tob2
AA986860	Cd274	0000073374	Hist1h2bl	Map3k8	Pfkm	Sema6c	Top2a
Abca4	Cd74	ENSMUSG0	Hist1h2bn	Mapk13	Pik3ip1	Sema7a	Tprn
Abi3bp	Cd81	0000090622	Hist1h3e	Mast4	Pip5kl1	Sept4	Trappc13
Ablim1	Cdk1	Ephb2	Hist1h3h	Mblac2	Pla1a	Serpinb8	Trib1
Acap1	Cdk5r1	Epm2aip1	Hist1h3i	Mcf2l	Plagl1	Serpinb9b	Trim12c
Acot1	Cdkl1	Eps15	Hist1h4f	mCG_14706	Plekha4	Serping1	Trim47
Adamts1	Cdkn1a	Esam	Hist1h4m	8	Plet1	Sertad3	Trim6
Adamts2	Ceacam1	Etv3	Hist2h2be	Mcpt8	Plk2	Sh2d1b1	Trmt10c
Adamtsl1	Cemip	F8a	Hist2h3c1	Mdfi	Pmp22	Sh2d3c	Trmt12
Adck2	Cend1	Fam110b	Hlf	Meaf6	Pnck	Shroom3	Trp53
Adm	Cercam	Fam129a	Hoxa1	Mfap2	Podxl	Sipar	Tspan12
Adrb1	Ch25h	Fam19a5	Hoxa5	Mfsd9	Ppan	Sirt1	Tspan33
Agpat9	Chrd	Fam219a	Hoxa9	Mgp	Ppp1r3b	Slc13a4	Tspyl5
Agt	Chst10	Fam46a	Hoxd8	Mitf	Prkg2	Slc16a11	Tst
AI429214	Cish	Fam69c	Hoxd9	Mktn3	Prl14l	Slc16a3	Tct25
Akap6	Clec11a	Fam84b	Hp	Mmp17	Prl2c2	Slc1a6	Ttc5
Akna	Cntfr	Fcer1g	Hs3st6	Mmp9	Prl2c3	Slc22a4	Ttll11
Aldh1a3	Col12a1	Fgf21	Hsf2bp	Mob3c	Prl7a1	Slc24a3	Uap111
Aldoc	Col22a1	Fgfr3	Hspa1l	Mocs3	Prorsd1	Slc25a29	Ubac1
Alox8	Col27a1	Fhl1	Htr1d	Mpped2	Prr5	Slc39a1	Ubxn11
Amotl2	Col3a1	Fhl4	Icosl	Msh4	Prss16	Slc4a1ap	Vgf
Ampd1	Colec11	Fibcd1	Ifit3	Mtm1	Prss22	Slpi	Vsir
Angptl4	Commmd4	Fkbp1	Ifit3b	Mtmr14	Ptch2	Smg8	Wisp2
Ankrd24	Cpne9	Flrt3	Igfbp2	mt-Nd6	Ptma	Smug1	Wnt2
Ankrd49	Cpt2	Folr1	Iigp1	Mx1	Ptplad1	Socs1	Wnt9a
Ankrd66	Csgalnact1	Foxq1	Il15ra	Myadm	Ptpn3	Sod3	Xaf1
Ankrd9	Csrmp1	Fpgt	Il1rn	Myc	Ptprv	Sorcs2	Xdh
Anks1b	Cst6	Frat2	Isl1	Mycbp	Ptprz1	Sp110	Zbtb3
Apol6	Ctf1	Fscn1	Itgb6	Mycl1	Pygm	Sp7	Zbtb39
Apol9b	Ctsf	Fzd8	Itgb1l	Myo1g	Rab11b	Spata2	Zbtb43
Apold1	Cyb5rl	Gadd45g	Ivl	Myom2	Rab7b	Spib	Zbtb45
Apoo-ps	D0H4S114	Gapdh	Jph2	Naif1	Ralgds	Spin4	Zbtb6
Areg	Dbndd2	Gatc	Jrkl	Nap1l2	Ramp1	Spock1	Zbtb9
Arhgap27	Dcun1d3	Gdf7	Jun	Ncf2	Ranbp6	Spon2	Zcchc3
Arhgef15	Ddx60	Gdnf	Kbtbd8	Ndnf	Rap2b	Sprr1a	Zfp109
Arhgef37	Dhcr7	Glce	Kcng1	Nfatc4	Ras11a	Sprr2k	Zfp11
Arl4d	Dhx16	Gldc	Kcnn4	Nfe2	Rassf2	Sprtn	Zfp113
Arl5c	Dlx2	Gm10263	Kctd21	Nfe2l3	Rassf6	Spry1	Zfp119b
Arntl	Dmrt2	Gm12216	Kif22	Ngef	Rbm12b1	Sptlc2	Zfp128
Asb5	Dpcr1	Gm4076	Kin	Nipal3	Rcan1	Stmn4	Zfp383
Asphd2	Drd4	Gm5468	Klf4	Nkain4	Retn	Ston2	Zfp39
Atf3	Dusp6	Gm5884	Klrd1	Nog	Rgs16	Stxbp6	Zfp408
Atp2b4	Dusp8	Gm8420	Krt18	Nol4l	Rhob	Sugct	Zfp418
AW494830	Dynap	Gm9892	Krt7	Npr3	Ribc1	Susd1	Zfp518b
Bbs10	Edn1	Gm9900	Krtap1-5	Npy1r	Rnf113a1	Syng3	Zfp583
BC025920	Efnb1	Gm9938	Lama5	Nr2f1	Rnf207	Sytl1	Zfp738
BC030307	Egr1	Gm9958	Laptm4b	Nr4a3	Rnf39	Taf9	Zfp747
Bdkrb1	Egr3	Gpr137c	Lce1f	Nrn1	Rora	Tbkbp1	Zfp830
Bet1l	Ehd1	Gpr64	Lce1h	Nrp2	Rpusd2	Tbr1	Zfp865
Bhlhe22	Ehd2	Gramd2	Loxl2	Ntn4	Rsl1	Tbxa2r	Zfp867
Bmp4	Eid2	Gtsf1	Lpar6	Nynrin	Rtkn	Tcea3	Zic1
Bves	Enpp2	H2-Q6	Lrfn3	Olfr2	Rwdd2a	Tec	Zkscan14
C1qtnf3	ENSMUSG0	H3f3a	Lrp2	Olfr456	S100a5	Tef	Zxdb
C1qtnf6	0000018543	Hacl1	Lrrc29	ORF63	S1pr1	Tgfb1	
Cabp4	ENSMUSG0	Haghl	Lrrc8e	Paqr6	Saa3	Thap1	
Casp1	0000044854	Has1	Ltbp2	Pcca	Sat1	Tktl1	
Casp4	ENSMUSG0	Hck	Lyn	Pcdhb16	Satb2	Tlr3	
Cbr2	0000045506	Hdh3	Lypd6	Pcdhb22	Scd3	Tmem115	
Ccdc120	ENSMUSG0	Heatr7b1	Lysmd4	Pcdhb5	Scml2	Tmem151a	
Ccl20	0000066178	Hexim2	Madd	Pcdhb9	Scn2a1	Tnfaip3	
Ccl8	ENSMUSG0	Hhex	Mageb3	Pde4b	Scube3	Tnfaip8l1	
Ccnyl1	0000070997	Hhip1l	Magi2	Pdgfa	Sde2	Tob1	

Appendix Table D.3: Additional genes obtained by fixing zero errors.

Gene ID	p53KO_NT FPKM	p53KO_UV FPKM	TKO_NT FPKM	TKO_UV FPKM	DEG Score	Name	Feature Type
Gm12858	0.89	0.00	0.00	0.53	25.49	predicted gene 12858	pseudogene
Gm26168	9.71	247.54	0.00	706.18	18.08	predicted gene, 26168	unclassified non-coding RNA gene
Gm26427	0.00	283.34	17.33	310.59	-17.27	predicted gene, 26427	snRNA gene
Rny1	9.71	446.19	0.00	612.82	17.03	RNA, Y1 small cytoplasmic, Ro-associated	scRNA gene
Gm23119	0.00	36.77	3.94	28.52	-15.63	predicted gene, 23119	snoRNA gene
Gm24407	0.00	22.66	2.04	12.05	-15.23	predicted gene, 24407	snRNA gene
Gm16056	1.42	16.30	0.00	27.88	14.57	predicted gene 16056	pseudogene
Rnu11	1.85	21.90	0.00	14.90	13.62	U11 small nuclear RNA	snRNA gene
Gm14863	0.34	0.00	0.12	0.31	13.12	predicted gene 14863	pseudogene
Gm9174	0.79	0.00	0.72	0.44	12.24	predicted pseudogene 9174	pseudogene
Rpl36-ps10	0.00	0.88	0.28	0.75	-11.70	ribosomal protein L36, pseudogene 10	pseudogene
Gm11843	0.00	0.71	0.13	0.47	-10.97	predicted gene 11843	pseudogene
Gm15344	0.84	0.00	0.61	0.13	10.77	predicted gene 15344	unclassified non-coding RNA gene
Cbx3-ps4	0.19	0.00	0.34	0.28	10.66	chromobox 3, pseudogene 4	pseudogene
Gm20594	0.00	2.21	0.15	2.22	-10.58	predicted gene, 20594	protein coding gene
Rpl26-ps2	0.00	0.66	0.13	0.62	-10.39	ribosomal protein L26, pseudogene 2	pseudogene
Rpl21-ps12	0.00	0.32	0.23	0.56	-10.37	ribosomal protein L21, pseudogene 12	pseudogene
Gm8261	0.00	0.16	0.10	0.12	-10.35	predicted gene 8261	pseudogene
Fam180a	0.18	0.00	1.16	0.70	10.09	family with sequence similarity 180, member A	protein coding gene
Gtpbp4-ps5	0.09	0.29	0.00	0.36	10.07	GTP binding protein 4, pseudogene 5	pseudogene
Adig	0.09	0.38	0.00	0.44	10.07	adipogenin	protein coding gene
Gm8805	0.00	0.36	0.18	0.61	-9.99	predicted gene 8805	pseudogene
Gm5384	0.00	0.23	0.09	0.21	-9.92	predicted gene 5384	pseudogene
Hist3h2bb-ps	0.12	0.19	0.00	0.16	9.87	histone cluster 3, H2bb, pseudogene	pseudogene
Platr5	0.00	0.33	0.35	1.29	-9.84	pluripotency associated transcript 5	lincRNA gene
Rn7s1	0.00	1.93	0.13	3.81	-9.33	7S RNA 1	SRP RNA gene
E130102H24Rik	0.14	0.00	0.18	0.09	9.32	RIKEN cDNA E130102H24 gene	lincRNA gene
Gm10766	0.00	0.56	0.11	0.95	-9.32	predicted gene 10766	protein coding gene
Mogat1	0.06	0.79	0.00	0.80	9.30	monoacylglycerol O-acyltransferase 1	protein coding gene
Gm5866	0.19	0.00	0.18	0.06	9.14	predicted gene 5866	pseudogene
1700039M10Rik	0.00	0.20	0.05	0.17	-9.13	RIKEN cDNA 1700039M10 gene	bidirectional promoter lncRNA gene
Gm8216	0.17	0.00	0.12	0.04	9.00	predicted gene 8216	pseudogene
1700031A10Rik	0.00	0.34	0.05	0.36	-8.97	RIKEN cDNA 1700031A10 gene	antisense lncRNA gene
Gm26035	0.00	2.56	0.06	3.06	-8.95	predicted gene, 26035	unclassified non-coding RNA gene
Gm13602	0.00	0.27	0.05	0.29	-8.95	predicted gene 13602	pseudogene
1700040D17Rik	0.07	0.23	0.00	0.16	8.93	RIKEN cDNA 1700040D17 gene	lincRNA gene
Gm13880	0.16	0.00	0.17	0.05	8.89	predicted gene 13880	pseudogene
Gm12279	0.00	1.08	0.07	1.61	-8.83	predicted gene 12279	antisense lncRNA gene
Spr2b	0.07	0.24	0.00	0.15	8.79	small proline-rich protein 2B	protein coding gene
Platr17	0.00	0.20	0.04	0.20	-8.71	pluripotency associated transcript 17	lincRNA gene
Nkain2	0.08	1.32	0.00	0.66	8.64	Na ⁺ /K ⁺ transporting ATPase interacting 2	protein coding gene
Tlrc2	0.23	0.05	0.17	0.00	-8.48	T cell receptor beta, constant 2	gene segment
Gm9027	0.26	0.04	0.26	0.00	-8.45	predicted gene 9027	pseudogene
Gm13935	0.00	0.20	0.04	0.25	-8.33	predicted gene 13935	pseudogene
Cd1642	0.04	0.23	0.00	0.18	8.27	CD164 sialomucin-like 2	protein coding gene
Eva1c	0.37	0.00	3.19	0.26	8.25	eva-1 homolog C (C. elegans)	protein coding gene
Slco4a1	0.19	0.00	0.64	0.10	8.25	solute carrier organic anion transporter family, member 4a1	protein coding gene
Spsb4	0.24	0.00	0.35	0.04	8.04	splA/ryanodine receptor domain and SOCS box containing 4	protein coding gene
1700010B08Rik	0.03	0.18	0.00	0.17	7.99	RIKEN cDNA 1700010B08 gene	protein coding gene
Gm26587	0.42	0.02	0.42	0.00	-7.74	predicted gene, 26587	pseudogene
Fam83a	0.00	0.40	0.01	0.13	-7.71	family with sequence similarity 83, member A	protein coding gene
Ifna13	0.00	0.25	0.01	0.17	-7.62	interferon alpha 13	protein coding gene
Gm5278	0.00	0.14	0.02	0.16	-7.60	predicted pseudogene 5278	pseudogene
Hypm	0.00	0.33	0.02	0.38	-7.59	huntingtin interacting protein M	protein coding gene
Gm13836	0.02	0.67	0.00	0.58	7.55	predicted gene 13836	pseudogene
Gm21738	0.02	0.39	0.00	0.33	7.42	predicted gene, 21738	protein coding gene
Cd59b	0.04	0.48	0.00	0.18	7.27	CD59b antigen	protein coding gene
Chst13	0.17	0.00	0.63	0.05	7.15	carbohydrate (chondroitin 4) sulfotransferase 13	protein coding gene
Gm11563	0.01	0.43	0.00	0.49	7.13	predicted gene 11563	protein coding gene
Gm14115	0.00	0.17	0.03	0.39	-6.90	predicted gene 14115	pseudogene
Jsrp1	0.04	0.74	0.00	0.18	6.76	junctional sarcoplasmic reticulum protein 1	protein coding gene
Gm2383	0.00	0.14	0.01	0.20	-6.56	predicted gene 2383	pseudogene
Gm9097	0.00	0.15	0.01	0.20	-6.50	predicted gene 9097	pseudogene
Efcc1	0.00	0.31	0.01	0.41	-6.29	EF hand and coiled-coil domain containing 1	protein coding gene
Fignl2	0.19	0.00	0.69	0.02	6.01	fidgetin-like 2	protein coding gene
Gm22314	1.30	0.00	0.32	0.00	2.03	predicted gene, 22314	unclassified non-coding RNA gene
Fam229a	0.00	0.59	0.00	0.16	-1.92	family with sequence similarity 229, member A	protein coding gene
Fam71a	0.00	0.43	0.00	0.17	-1.32	family with sequence similarity 71, member A	protein coding gene
Mir1983	0.00	10.54	0.00	4.48	-1.23	microRNA 1983	miRNA gene
Gm25072	0.00	0.44	0.00	1.00	1.19	predicted gene, 25072	unclassified non-coding RNA gene

Appendix Table D.4: All 626 DEGs; FPKM values per sample, DEG Score, Gene name, type and locus information.

Gene ID	PKO_NT	PKO_UV	TKO_NT	TKO_UV	PKO_UV PKO_NT	TKO_UV TKO_NT	$\frac{TKO_UV}{TKO_NT}$	$\log_2\left(\frac{PKO_UV}{PKO_NT}\right)$	$\log_2\left(\frac{TKO_UV}{TKO_NT}\right)$	$\log_2\left(\frac{TKO_UV}{PKO_UV}\right)$	Response	Gene Name	Feature Type	Locus	Absolute DEG Score
1	Ptch2	0.32	0.04	0.06	0.450.12	7.62	62.23	3.03	2.93	5.96	increased	patched 1	protein coding gene	4:117096074-117116101	5.96
2	Gm1848	19.77	0.35	0.43	0.360.02	0.83	47.04	2.22	0.27	5.56	increased	predicted gene 1848	pseudogene	X:21146842-21147640	5.56
3	Zbtb9	4.28	0.80	3.09	0.020.19	0.01	0.04	2.42	2.04	4.78	Decreased	zinc finger and BTB domain containing 9	protein coding gene	17:26973166-27036378	4.78
4	Bdkrb1	2.21	0.38	0.11	0.450.17	4.12	24.26	2.56	2.04	6.60	increased	bradykinin receptor, beta 1	protein coding gene	12:105603084-105605428	4.6
5	Gm5560	1.69	3.18	78.66	6.141.88	0.08	0.04	0.91	3.68	4.59	Decreased	predicted pseudogene 5560	pseudogene	5:97264827-97265719	4.59
6	Rnf39	0.04	0.49	0.42	0.2712.73	0.64	0.05	1.67	0.65	4.32	Decreased	ring finger protein 39	protein coding gene	17:36942917-36947986	4.32
7	Fgf21	0.55	2.40	0.05	2.464.37	54.25	12.40	3.13	1.06	6.63	increased	fibroblast growth factor 21	protein coding gene	7:45613906-45615490	3.63
8	Msh4	0.00	0.13	0.04	0.1241.09	3.43	0.08	0.96	0.78	3.58	Decreased	mutS homolog 4	protein coding gene	3:153857146-153906138	3.58
9	Hist1h4f	0.05	4.07	0.42	3.1988.04	7.59	0.09	0.96	2.92	3.54	Decreased	histone cluster 1, H4c	protein coding gene	13:23551257-23551648	3.54
10	No14l	0.18	1.35	1.39	0.917.45	0.65	0.09	2.90	0.62	3.52	Decreased	nucleolar protein 4-like	protein coding gene	2:153407461-153529971	3.52
11	Hp	1.34	0.19	0.23	0.350.14	1.55	11.07	2.84	0.63	3.47	increased	haptoglobin	protein coding gene	8:109575129-109579172	3.47
12	Rpl31-ps12	0.04	2.26	0.33	1.7454.76	5.35	0.10	0.99	0.42	3.36	Decreased	ribosomal protein L31, pseudogene 1 2	pseudogene	16:16753015-16829456	3.36
13	Itgb6	0.46	0.80	0.14	2.301.71	16.98	9.93	0.77	0.09	3.31	increased	integrin beta 6	protein coding gene	2:605894172-28623290	3.31
14	Egr3	3.04	6.16	0.25	4.922.02	19.72	9.75	0.02	1.30	3.29	increased	early growth response 3	protein coding gene	14:70077444-70080157	3.29
15	Ndnf	1.83	0.20	0.16	0.180.11	1.08	9.61	3.16	0.11	3.26	increased	neuron-derived neurotrophic factor	protein coding gene	6:65671589-65706930	3.26
16	Lce1f	1.23	5.25	0.05	1.894.28	41.03	9.60	3.10	1.06	3.26	increased	late cornified envelope 1F	protein coding gene	3:92718684-92720350	3.26
17	Bmp4	0.10	0.20	11.40	2.572.11	0.23	0.11	1.08	2.15	3.23	Decreased	bone morphogenetic protein 4	protein coding gene	14:46379413-46390669	3.23
18	Colec11	0.10	1.47	0.16	0.2515.13	1.62	0.11	0.92	0.70	3.22	Decreased	collectin sub-family member 11	protein coding gene	12:28594172-28623290	3.22
19	Nog	0.08	0.20	0.55	0.162.44	0.30	0.12	1.29	1.75	3.04	Decreased	noggin	protein coding gene	11:89300637-89302559	3.04
20	Ribc1	0.18	1.42	0.70	0.687.67	0.97	0.13	2.94	0.04	2.98	Decreased	RIB43A domain with coiled-coils 1	protein coding gene	X:152004581-152016295	2.98
21	Cntfr	0.31	0.18	0.24	0.020.57	0.08	0.13	0.81	0.72	2.91	Decreased	ciliary neurotrophic factor receptor	protein coding gene	4:41657497-41697089	2.91
22	Fam69c	1.26	0.04	0.38	0.090.03	0.23	7.49	0.08	2.14	2.90	increased	family with sequence similarity 69, member C	protein coding gene	18:84719003-84740436	2.90
23	Ephb2	1.73	0.61	0.15	0.010.35	0.05	0.13	1.50	2.40	2.90	Decreased	Eph receptor B2	protein coding gene	4:136647538-136835988	2.90
24	Nrep	4.65	0.09	4.25	0.600.02	0.14	7.10	0.05	2.82	2.83	increased	neuronal regeneration related protein	protein coding gene	18:33437018-33464029	2.83
25	Agt	1.22	0.58	0.58	1.960.48	3.36	7.06	1.07	0.75	2.82	increased	angiotensinogen (serpin peptidase inhibitor, clade A, member	protein coding gene	8:124556533-124569706	2.82
26	Trim6	0.24	0.34	1.16	0.241.44	0.21	0.14	1.53	2.27	2.80	Decreased	tripartite motif-containing 6	protein coding gene	7:104218792-104235152	2.80
27	Pagr6	0.02	0.63	0.08	0.4941.16	6.15	0.15	0.96	2.62	2.74	Decreased	progesterin and adipoQ receptor family member VI	protein coding gene	3:88364583-88368541	2.74
28	Adrb1	0.43	0.09	0.69	0.020.20	0.03	0.15	2.29	2.04	2.74	Decreased	adrenergic receptor, beta 1	protein coding gene	19:5672371-56724862	2.74
29	Rnf207	0.25	0.17	0.03	0.130.68	4.49	6.64	0.57	2.17	2.73	increased	ring finger protein 207	protein coding gene	4:152297226-152318993	2.73
30	Slc16a3	0.44	0.56	4.02	0.791.26	0.20	0.15	1.34	2.36	2.69	Decreased	solute carrier family 16 (monocarboxylic acid transporters), me	protein coding gene	11:120948479-120960868	2.69
31	Apol6	1.27	0.12	0.56	0.350.10	0.63	6.44	0.36	0.67	2.69	increased	apolipoprotein L 6	protein coding gene	15:77015488-77057107	2.69
32	Ifit3	0.77	0.51	1.97	8.340.67	4.24	6.33	0.58	2.08	2.66	increased	interferon-induced protein with tetratricopeptide repeats 3	protein coding gene	19:34583530-34588731	2.66
33	Mid1-ps1	0.11	0.20	0.68	0.201.86	0.29	0.16	0.89	1.76	2.66	Decreased	midline 1, pseudogene 1	pseudogene	Y:90755656-90763485	2.66
34	Gm12216	0.31	0.24	0.27	1.330.80	4.96	6.24	0.33	0.31	2.64	increased	predicted gene 12216	protein coding gene	11:53783417-53859256	2.64
35	Hsf2bp	0.29	0.03	0.24	0.170.12	0.71	5.95	3.07	0.49	2.57	increased	heat shock transcription factor 2 binding protein	protein coding gene	17:31944768-32034508	2.57
36	Arl4d	2.20	5.26	6.66	2.742.39	0.41	0.17	1.26	1.28	2.54	Decreased	ADP-ribosylation factor-like 4D	protein coding gene	11:10166540-101667832	2.54
37	Zcchc3	0.17	0.62	9.59	6.213.73	0.65	0.17	0.90	0.63	2.53	Decreased	zinc finger, CCHC domain containing 3	protein coding gene	2:152411955-152415044	2.53
38	Akna	0.71	0.11	0.57	0.020.15	0.03	0.17	2.70	3.12	2.52	Decreased	AT-hook transcription factor	protein coding gene	4:63367124-63403354	2.52
39	Ddx60	0.02	0.35	0.27	0.8017.03	2.97	0.17	0.09	1.57	2.52	Decreased	DEAD (Asp-Glu-Ala-Asp) box polypeptide 60	protein coding gene	8:61928086-62037701	2.52
40	Nipa3	0.13	0.11	0.24	0.030.81	0.14	0.18	0.30	2.81	2.50	Decreased	NIPA-like domain containing 3	protein coding gene	4:135445419-135537803	2.50
41	Pde4b	3.23	1.07	4.52	8.260.33	1.83	5.55	1.60	0.87	2.47	increased	phosphodiesterase 4B, cAMP specific	protein coding gene	4:102087542-102607259	2.47
42	Edn1	1.15	3.01	8.78	4.232.62	0.48	0.18	1.39	1.05	2.44	Decreased	endothelin 1	protein coding gene	13:42301269-42307989	2.44
43	Tbkbp1	1.05	0.83	0.60	0.090.79	0.15	0.18	0.34	2.78	2.44	Decreased	TBK1 binding protein 1	protein coding gene	11:97136170-97151495	2.44
44	Icosl	0.65	0.07	0.18	0.10	0.11	0.59	3.19	0.77	2.42	increased	icos ligand	protein coding gene	10:78069359-78079525	2.42
45	Gm10827	0.02	0.46	0.02	2.67	23.26	124.23	0.54	0.00	2.42	increased	predicted gene 10827	lincRNA gene	10:18235052-18238060	2.42
46	Foxq1	0.08	0.25	1.39	0.77	2.93	0.55	0.55	0.85	2.41	Decreased	forkhead box Q1	protein coding gene	13:31556133-31560970	2.41
47	Spin4	0.51	1.83	0.53	0.36	3.59	0.69	0.85	0.54	2.39	Decreased	serine (or cysteine) peptidase inhibitor, clade E, member 2	protein coding gene	X:95022506-95026682	2.39
48	2610528A11Rik	0.30	1.11	1.25	0.89	3.72	0.71	0.90	0.49	2.39	Decreased	RIKEN cDNA 2610528A11 gene	protein coding gene	14:37102139-37110101	2.39
49	Ceacam1	0.23	0.07	0.19	0.01	0.29	0.06	1.77	4.15	2.38	Decreased	carcinoembryonic antigen-related cell adhesion molecule 1	protein coding gene	7:25461701-25477625	2.38
50	Pr12c3	42.56	42.21	2.31	11.82	0.99	5.11	0.01	0.35	2.36	increased	prolactin family 2, subfamily c, member 3	protein coding gene	13:12790821-12800078	2.36
51	Rubcnl	0.61	0.01	0.22	0.03	0.02	0.12	0.45	3.06	2.36	increased	RUN and cysteine rich domain containing beclin 1 interacting p	protein coding gene	14:75016026-75052532	2.36
52	Il1rn	7.00	1.52	0.54	0.60	0.22	1.11	2.20	0.16	2.36	increased	interleukin 1 receptor antagonist	protein coding gene	2:24336852-24351494	2.36
53	Krt7	0.14	1.04	2.12	3.13	7.43	1.48	2.89	0.56	2.33	Decreased	keratin 7	protein coding gene	15:101411042-101430313	2.33
54	Tspyl5	0.14	0.23	0.17	0.06	1.63	0.32	0.70	1.63	2.33	Decreased	testis-specific protein, Y-encoded-like 5	protein coding gene	15:33683875-33687883	2.33
55	Klf4	1.24	3.28	12.79	6.78	2.63	0.53	0.40	0.92	2.31	Decreased	Kruppel-like factor 4 (gut)	protein coding gene	4:55527142-55532466	2.31

56	Stxbp6	0.22	0.46	2.98	1.27	2.12	0.43	0.20	1.09	1.22	2.31	Decreased	syntaxin binding protein 6 (amisyn)	protein coding gene	12:44852487-45074493	2.31
57	Ch25h	0.21	0.35	4.98	1.67	1.65	0.34	0.20	1.72	1.58	2.30	Decreased	cholesterol 25-hydroxylase	protein coding gene	19:34473785-34481546	2.30
58	Map3k7cl	1.05	2.90	0.70	0.39	2.76	0.56	0.20	1.47	0.83	2.29	Decreased	Map3k7 C-terminal like	protein coding gene	16:87553329-87595336	2.29
59	Hist1h4m	0.11	6.04	0.29	3.32	55.88	11.56	0.21	1.01	0.53	2.27	Decreased	histone cluster 1, H4d	protein coding gene	13:21811745-21812150	2.27
60	H2-Q6	0.09	0.16	0.03	0.27	1.80	8.57	4.76	1.85	1.10	2.25	Increased	histocompatibility 2, Q region locus 6	protein coding gene	17:35424849-35430055	2.25
61	Mfsd9	0.93	0.10	0.16	0.08	0.11	0.52	4.72	1.17	0.93	2.24	Increased	major facilitator superfamily domain containing 9	protein coding gene	1:40772037-40790687	2.24
62	Stmn4	1.37	3.17	2.14	1.06	2.31	0.49	0.21	1.21	1.02	2.22	Decreased	stathmin-like 4	protein coding gene	14:66344295-66361680	2.22
63	Podxl	0.22	0.48	2.12	0.98	2.16	0.46	0.21	1.11	1.11	2.22	Decreased	podocalyxin-like	protein coding gene	6:31519487-31563981	2.22
64	Spock1	0.07	0.33	0.01	0.18	5.12	23.81	4.65	1.36	1.57	2.22	Increased	sparc/osteonectin, cwcv and kazal-like domains proteoglycan 1	protein coding gene	13:57429294-57907587	2.22
65	Gm11221	0.34	0.17	0.88	2.04	0.50	2.31	4.61	0.99	1.21	2.20	Increased	predicted gene 11221	pseudogene	4:69691977-69692574	2.20
66	Serping1	0.02	0.35	0.06	0.25	19.89	4.37	0.22	1.31	2.13	2.19	Decreased	serine (or cysteine) peptidase inhibitor, clade G, member 1	protein coding gene	2:84765386-84775444	2.19
67	Flrt3	0.59	0.56	6.84	1.43	0.95	0.21	0.22	0.07	2.25	2.18	Decreased	fibronectin leucine rich transmembrane protein 3	protein coding gene	2:140395308-142392966	2.18
68	B230369F24Rik	0.05	0.69	0.10	0.31	13.35	2.94	0.22	1.74	1.56	2.18	Decreased	RIKEN cDNA B230369F24 gene	antisense lncRNA gene	1:178313327-178318549	2.18
69	Gm20627	0.28	4.59	0.02	1.61	16.61	74.48	4.48	1.05	1.76	2.17	Increased	predicted gene 20627	unclassified non-coding RNA gene	3:96254618-96256277	2.17
70	Slpi	0.01	0.17	5.22	16.31	13.98	3.12	0.22	1.81	1.64	2.16	Decreased	secretory leukocyte peptidase inhibitor	protein coding gene	2:164354069-164389095	2.16
71	Zfp583	0.17	0.13	0.17	0.03	0.76	0.17	0.22	1.39	2.55	2.16	Decreased	zinc finger protein 583	protein coding gene	7:6315659-6331285	2.16
72	Mmp17	0.21	0.91	0.73	0.71	4.27	0.96	0.22	1.09	0.06	2.15	Decreased	matrix metalloproteinase 17	protein coding gene	5:129584154-129608875	2.15
73	Mageb3	0.02	0.57	0.05	0.33	30.26	6.80	0.22	1.92	2.77	2.15	Decreased	melanoma antigen, family B, 3	protein coding gene	2:121953751-121955360	2.15
74	Lpar6	2.55	15.03	5.28	7.00	5.90	1.33	0.22	1.56	0.41	2.15	Decreased	lysophosphatidic acid receptor 6	protein coding gene	14:73123036-73325822	2.15
75	Rpl15-ps1	0.11	1.34	0.76	2.07	12.07	2.71	0.22	1.59	1.44	2.15	Decreased	ribosomal protein L15, pseudogene 1	pseudogene	11:21953139-21953703	2.15
76	Serpinb8	0.73	0.17	0.46	0.49	0.24	1.05	4.43	2.07	0.08	2.15	Increased	serine (or cysteine) peptidase inhibitor, clade B, member 8	protein coding gene	1:107590005-107608909	2.15
77	Hist2h3c1	0.27	1.65	0.27	0.37	6.04	1.38	0.23	1.60	0.46	2.14	Decreased	histone cluster 2, H3c1	protein coding gene	3:96243549-96248509	2.14
78	Myo1g	0.01	0.70	0.02	0.22	57.54	13.11	0.23	1.05	1.71	2.13	Decreased	myosin IG	protein coding gene	11:6506547-6520965	2.13
79	Il15ra	1.71	0.30	0.80	0.62	0.18	0.77	4.34	1.50	0.38	2.12	Increased	interleukin 15 receptor, alpha chain	protein coding gene	2:11705286-11733986	2.12
80	Fhl4	0.16	0.64	0.12	0.11	3.87	0.89	0.23	1.95	0.16	2.11	Decreased	four and a half LIM domains 4	protein coding gene	10:85097018-85102495	2.11
81	1700001P01Rik	0.08	2.73	0.15	1.21	34.73	8.05	0.23	1.12	1.01	2.11	Decreased	RIKEN cDNA 1700001P01 gene	protein coding gene	11:97771480-97775918	2.11
82	Ifit3b	0.15	0.13	0.51	1.90	0.87	3.76	4.31	0.20	1.91	2.11	Increased	interferon-induced protein with tetratricopeptide repeats 3B	protein coding gene	19:34607969-34613401	2.11
83	Fam46a	0.38	0.79	0.53	0.26	2.08	0.49	0.23	1.06	1.04	2.10	Decreased	family with sequence similarity 46, member A	protein coding gene	9:85320438-85327124	2.10
84	Slc1a6	1.07	12.98	0.15	0.44	12.13	2.89	0.24	1.60	1.53	2.07	Decreased	solute carrier family 1 (high affinity aspartate/glutamate transpor	protein coding gene	10:78780495-78814825	2.07
85	Cpne9	0.03	0.87	0.04	0.27	27.54	6.57	0.24	1.78	2.72	2.07	Decreased	copine V	protein coding gene	6:113282306-113305627	2.07
86	Gm26698	0.11	0.18	0.10	0.65	1.62	6.71	4.15	1.69	1.75	2.05	Increased	predicted gene, 26698	lincRNA gene	12:86891403-86892567	2.05
87	1700123M08Ri	0.73	0.15	0.72	0.60	0.20	0.83	4.14	1.32	0.28	2.05	Increased	RIKEN cDNA 1700123M08 gene	unclassified non-coding RNA gene	4:11966564-11994280	2.05
88	Cercam	0.92	0.98	0.54	0.14	1.07	0.26	0.24	1.10	1.95	2.04	Decreased	cerebral endothelial cell adhesion molecule	protein coding gene	2:29868725-29882840	2.04
89	Mir99ahg	0.31	3.02	1.84	4.30	9.63	2.34	0.24	1.27	1.23	2.04	Decreased	Mir99a and Mirlet7c-1 host gene (non-protein coding)	unclassified non-coding RNA gene	16:77236316-77691403	2.04
90	Sprr1a	####	####	4.74	50.68	2.60	10.69	4.11	1.38	1.42	2.04	Increased	small proline-rich protein 1B	protein coding gene	3:92483951-92485895	2.04
91	Ranbp6	0.65	3.50	0.87	1.15	5.38	1.33	0.25	1.43	0.41	2.01	Decreased	RAN binding protein 6	protein coding gene	19:29808107-29812974	2.01
92	Ptpn3	0.88	0.14	0.69	0.46	0.16	0.66	4.04	1.60	0.59	2.01	Increased	protein tyrosine phosphatase, non-receptor type 3	protein coding gene	4:57190840-57307305	2.01
93	Ntn4	1.29	0.74	3.53	0.51	0.58	0.14	0.25	0.79	2.80	2.01	Decreased	myeloid leukemia factor 2	protein coding gene	10:93640680-93747207	2.01
94	Nr4a3	0.05	0.14	0.03	0.33	3.10	12.41	4.00	1.63	1.63	2.00	Increased	nuclear receptor subfamily 4, group A, member 3	protein coding gene	4:48045152-48086447	2.00
95	Uqcrr-ps1	1.46	1.11	1.12	3.43	0.76	3.05	4.00	0.39	1.61	2.00	Increased	ubiquinol-cytochrome c reductase hinge protein, pseudogene 1	pseudogene	15:99898286-99898556	2.00
96	Fkbp1	4.46	2.81	5.21	0.83	0.63	0.16	0.25	0.66	2.65	1.99	Decreased	FK506 binding protein-like	protein coding gene	17:34644763-34646324	1.99
97	Dusp8	1.44	2.14	2.35	0.89	1.49	0.38	0.25	1.57	1.40	1.98	Decreased	dual specificity phosphatase 8	protein coding gene	7:142079489-142095843	1.98
98	Cemp	0.81	1.00	1.53	0.49	1.23	0.32	0.26	1.30	1.66	1.96	Decreased	cell migration inducing protein, hyaluronan binding	protein coding gene	7:83932856-84086502	1.96
99	Rps15a-ps8	0.40	1.03	0.23	2.32	2.57	9.96	3.87	1.36	1.32	1.95	Increased	ribosomal protein S15A, pseudogene 8	pseudogene	4:53791069-53791456	1.95
100	Hoxa1	2.54	8.89	1.73	1.56	3.50	0.90	0.26	1.81	0.15	1.95	Decreased	homeobox A1	protein coding gene	6:52155589-52162119	1.95
101	Olfra456	0.10	0.50	0.73	0.94	4.98	1.29	0.26	1.32	0.37	1.95	Decreased	olfactory receptor 456	protein coding gene	6:42486162-42487214	1.95
102	Plekha4	1.10	0.98	0.13	0.45	0.89	3.43	3.86	0.17	1.78	1.95	Increased	pleckstrin homology domain containing, family A (phosphoinosit	protein coding gene	7:45526329-45554229	1.95
103	Asb5	0.09	0.35	0.22	0.22	3.75	0.98	0.26	1.91	0.03	1.94	Decreased	ankyrin repeat and SOCs box-containing 5	protein coding gene	8:54550330-54610098	1.94
104	Gm5468	0.18	1.01	0.11	2.34	5.74	21.98	3.83	1.52	1.46	1.94	Increased	predicted gene 5468	lincRNA gene	15:25414191-25414844	1.94
105	Hoxaas3	1.20	24.87	0.67	3.66	20.79	5.43	0.26	1.38	1.44	1.94	Decreased	Hoxa cluster antisense RNA 3	antisense lncRNA gene	6:52162350-52221854	1.94
106	Gpat3	0.53	0.06	0.40	0.18	0.12	0.46	3.83	1.06	1.13	1.94	Increased	coiled-coil-helix-coiled-coil-helix domain containing 2	protein coding gene	5:100840796-100899102	1.94
107	Gm10433	0.03	0.51	0.04	0.19	17.28	4.52	0.26	1.11	0.18	1.93	Decreased	predicted gene 10433	unclassified non-coding RNA gene	12:100187967-100193540	1.93
108	Cish	0.50	0.35	1.04	0.20	0.72	0.19	0.26	0.48	2.41	1.93	Decreased	cytokine inducible SH2-containing protein	protein coding gene	9:107292434-107302784	1.93
109	H3f3a	####	54.46	####	####	0.25	0.93	3.78	2.02	0.10	1.92	Increased	H3 histone, family 3A	protein coding gene	1:180800831-180816859	1.92
110	Scube3	0.06	0.14	0.42	0.24	2.12	0.56	0.27	1.08	0.83	1.91	Decreased	signal peptide, CUB domain, EGF-like 3	protein coding gene	17:28142315-28174852	1.91

111Ctf1	3.30	0.51	1.94	1.13	0.16	0.58	3.75	2.69	0.78	1.91	increased	cardiotrophin 1	protein coding gene	7:127712735-127725616	1.91
112S100a5	2.12	2.26	0.59	2.36	1.07	3.99	3.75	1.09	1.00	1.91	increased	S100 calcium binding protein A5	protein coding gene	3:90608522-90611780	1.91
113Madd	0.73	0.60	1.35	0.30	0.83	0.22	0.27	0.27	2.18	1.90	Decreased	electron transferring flavoprotein, alpha polypeptide	protein coding gene	2:91137359-91183837	1.90
114Hist2h2be	0.49	2.01	0.82	0.89	4.07	1.09	0.27	1.03	1.13	1.90	Decreased	histone cluster 1, H2bp	protein coding gene	3:96221120-96223738	1.90
115Saa3	0.08	1.49	0.40	2.14	19.75	5.31	0.27	1.30	1.41	1.89	Decreased	serum amyloid A 1	protein coding gene	7:46711997-46715676	1.89
116Stk19-ps1	0.71	0.76	0.52	2.04	1.06	3.95	3.71	1.09	1.98	1.89	increased	serine/threonine kinase 19, pseudogene 1	pseudogene	17:34744752-34747154	1.89
117Adm	4.52	5.71	1.13	0.38	1.26	0.34	0.27	0.34	1.55	1.89	Decreased	adrenomedullin	protein coding gene	7:110627668-110629819	1.89
118Ttc25	0.03	1.28	0.03	0.42	46.63	12.61	0.27	1.53	1.66	1.89	Decreased	tetralatricopeptide repeat domain 25	protein coding gene	11:100545606-100572568	1.89
119Isl1	0.95	0.14	0.22	0.12	0.15	0.55	3.68	2.74	0.86	1.88	increased	ISL1 transcription factor, LIM/homeodomain	protein coding gene	13:116298280-116309688	1.88
120Sertad3	8.71	2.26	4.54	0.32	0.26	0.07	0.27	1.95	1.83	1.88	Decreased	SERTA domain containing 3	protein coding gene	7:2473767-2477359	1.88
121Arhgef15	0.52	0.14	0.31	0.31	0.27	0.99	3.67	1.89	0.01	1.88	increased	Rho guanine nucleotide exchange factor (GEF) 15	protein coding gene	11:68943154-68957480	1.88
122Ivl	0.02	0.23	0.06	0.23	13.83	3.79	0.27	1.79	1.92	1.87	Decreased	involucrin	protein coding gene	3:92570911-92573715	1.87
123Pip5k1l	0.08	0.58	0.23	0.45	7.02	1.95	0.28	1.81	1.96	1.85	Decreased	phosphatidylinositol-4-phosphate 5-kinase-like 1	protein coding gene	2:32574183-32583782	1.85
124AI429214	0.30	0.92	0.40	0.34	3.09	0.86	0.28	1.63	0.22	1.85	Decreased	expressed sequence AI429214	protein coding gene	8:36993574-36995531	1.85
125Hist1h2bn	0.62	1.15	0.36	2.41	1.86	6.70	3.61	0.89	0.74	1.85	increased	histone cluster 1, H2bm	protein coding gene	13:21754122-21754503	1.85
126Prr14l	1.79	1.25	1.92	0.37	0.70	0.19	0.28	0.51	2.36	1.85	Decreased	proline rich 14-like	protein coding gene	5:32736300-32854256	1.85
127Zfp109	0.51	0.10	0.17	0.01	0.20	0.06	0.28	2.30	1.15	1.84	Decreased	zinc finger protein 109	protein coding gene	7:24228047-24236632	1.84
128Plet1	0.13	0.15	0.06	0.26	1.17	4.19	3.59	1.22	1.07	1.84	increased	placenta expressed transcript 1	protein coding gene	9:50494524-50505639	1.84
129Angptl4	6.05	0.80	5.09	2.42	0.13	0.48	3.59	2.91	1.07	1.84	increased	angiotensin-like 4	protein coding gene	17:33773749-33781575	1.84
130Prss16	0.83	0.30	0.71	0.90	0.36	1.27	3.54	1.48	0.34	1.82	increased	protease, serine 16 (thymus)	protein coding gene	13:22002175-22009742	1.82
131Col27a1	0.24	0.08	2.14	2.48	0.33	1.16	3.53	1.61	0.21	1.82	increased	collagen, type XXVII, alpha 1	protein coding gene	4:63214003-63334991	1.82
132Gm12372	0.15	0.99	0.39	0.74	6.69	1.91	0.29	1.74	1.93	1.81	Decreased	predicted gene 12372	pseudogene	4:37139106-37139874	1.81
133Htr1d	0.06	0.46	0.03	0.94	8.30	29.02	3.50	1.05	1.86	1.81	increased	5-hydroxytryptamine (serotonin) receptor 1D	protein coding gene	4:136423523-136444398	1.81
134Apoo-ps	0.32	0.27	0.32	0.95	0.86	2.99	3.49	0.22	1.58	1.80	increased	apolipoprotein O, pseudogene	pseudogene	13:107413864-107414767	1.80
135Apol9b	0.80	3.84	12.95	17.75	4.78	1.37	0.29	2.26	0.45	1.80	Decreased	apolipoprotein L 9b	protein coding gene	15:77729120-77736381	1.80
136493048318Rik	0.12	2.82	0.16	1.03	22.71	6.52	0.29	1.51	1.70	1.80	Decreased	RIKEN cDNA 4930483J18 gene	unclassified non-coding RNA gene	15:81190773-81192763	1.80
137Gm3940	0.07	1.06	0.32	1.39	14.82	4.27	0.29	1.89	1.09	1.80	Decreased	predicted gene 3940	pseudogene	15:1597147-152107188	1.80
138Adamts1	6.63	2.39	13.00	1.35	0.36	0.10	0.29	1.47	0.26	1.79	Decreased	a disintegrin-like and metalloprotease (reprolysin type) with thr	protein coding gene	16:85793826-85803113	1.79
139Myom2	0.10	2.80	0.61	4.96	27.85	8.09	0.29	1.90	1.02	1.78	Decreased	myomesin 2	protein coding gene	8:15057652-15133541	1.78
140Plk2	27.82	31.81	87.26	29.02	1.14	0.33	0.29	1.19	1.59	1.78	Decreased	polo like kinase 2	protein coding gene	13:110395043-110400844	1.78
141Pik3ip1	2.92	0.47	2.01	1.11	0.16	0.55	3.43	2.63	0.85	1.78	increased	phosphoinositide-3-kinase interacting protein 1	protein coding gene	11:3288873-3342971	1.78
142Jrkl	1.19	4.89	1.19	1.43	4.11	1.20	0.29	1.04	1.26	1.78	Decreased	Jrk-like	protein coding gene	9:13242789-13245741	1.78
143Tnfrsf81l	1.24	0.45	1.19	0.12	0.36	0.10	0.29	1.48	0.25	1.78	Decreased	tumor necrosis factor, alpha-induced protein 8-like 1	protein coding gene	17:56162476-56173955	1.78
144Hist1h3i	0.35	7.92	0.84	5.57	22.40	6.62	0.30	1.49	0.73	1.76	Decreased	histone cluster 1, H3e	protein coding gene	13:21782914-21783397	1.76
145Akap6	0.03	0.17	0.16	0.24	5.24	1.55	0.30	1.39	1.63	1.75	Decreased	A kinase (PRKA) anchor protein 6	protein coding gene	12:5269382-53151015	1.75
146Mcpt8	0.26	0.53	0.49	3.42	2.07	6.98	3.37	1.05	1.80	1.75	increased	mast cell protease 8	protein coding gene	14:56082165-56085216	1.75
147Zfp418	0.34	0.05	0.25	0.12	0.14	0.49	3.36	2.79	1.04	1.75	increased	zinc finger protein 418	protein coding gene	7:7171329-7183562	1.75
148C920006O11Rik	0.50	3.32	0.96	1.90	6.61	1.97	0.30	1.73	1.98	1.75	Decreased	RIKEN cDNA C920006O11 gene	lincRNA gene	9:78175913-78178879	1.75
149Zfp950	3.67	2.98	1.52	4.14	0.81	2.71	3.35	0.30	1.44	1.74	increased	zinc finger protein 950	protein coding gene	19:60889748-61140840	1.74
150Gm9903	0.06	0.78	0.07	0.25	11.96	3.57	0.30	1.58	1.84	1.74	Decreased	predicted gene 9903	protein coding gene	5:33629924-33632092	1.74
151Itgbl1	1.14	0.68	0.09	0.19	0.60	2.02	3.34	0.73	1.01	1.74	increased	integrin, beta-like 1	protein coding gene	14:123659970-123975618	1.74
152Ttl11	0.36	0.77	0.51	0.33	2.17	0.65	0.30	1.12	0.62	1.74	Decreased	tubulin tyrosine ligase-like family, member 11	protein coding gene	2:35751240-35979913	1.74
153Abca4	0.10	0.42	0.25	0.30	4.03	1.21	0.30	1.01	1.27	1.74	Decreased	ATP-binding cassette, sub-family A (ABC1), member 4	protein coding gene	3:122044442-122180061	1.74
154Areg	1.71	0.56	0.59	0.64	0.33	1.09	3.33	1.61	0.13	1.74	increased	amphiregulin	protein coding gene	5:91139614-91148432	1.74
155Nfe2	0.03	0.31	0.13	0.37	9.37	2.82	0.30	1.23	1.50	1.73	Decreased	nuclear factor, erythroid derived 2	protein coding gene	15:103248211-103258403	1.73
156Cdkn1a	4.21	6.67	0.81	4.23	1.58	5.22	3.29	0.66	1.38	1.72	increased	cyclin-dependent kinase inhibitor 1A (P21)	protein coding gene	17:29090978-29100722	1.72
157Slc24a3	0.41	0.53	4.81	1.90	1.30	0.40	0.30	1.38	1.34	1.72	Decreased	solute carrier family 24 (sodium/potassium/calcium exchanger),	protein coding gene	2:145167753-145642166	1.72
158Gm9800	27.07	6.33	17.66	13.59	0.23	0.77	3.29	2.10	0.38	1.72	increased	predicted gene 9800	pseudogene	14:8799588-8800504	1.72
159Zfp867	0.52	0.67	0.60	0.24	1.30	0.40	0.31	1.37	1.33	1.71	Decreased	zinc finger protein 867	protein coding gene	11:59461196-59472474	1.71
160Ankrd66	0.42	0.16	0.58	0.70	0.37	1.21	3.26	1.43	0.28	1.71	increased	ankyrin repeat domain 66	protein coding gene	17:43534173-43543639	1.71
161Anks1b	0.64	0.30	0.34	0.05	0.46	0.14	0.31	1.11	2.81	1.70	Decreased	ankyrin repeat and sterile alpha motif domain containing 1B	protein coding gene	10:89873508-90973300	1.70
162Ankrd9	3.15	0.11	4.79	0.52	0.03	0.11	3.24	1.51	0.23	1.70	increased	ankyrin repeat domain 9	protein coding gene	12:110975352-110979040	1.70
163Gm3272	0.17	0.32	0.12	0.76	1.92	6.23	3.24	0.94	0.64	1.70	increased	predicted pseudogene 3272	pseudogene	8:92715746-92716975	1.70
164Ilgp1	0.14	0.18	0.29	1.20	1.28	4.13	3.24	0.35	0.05	1.70	increased	interferon inducible GTPase 1	protein coding gene	18:60376028-60392627	1.70
1651700055D18Rik	0.12	0.38	0.16	0.16	0.20	0.99	0.31	1.68	0.02	1.69	Decreased	RIKEN cDNA 1700055D18 gene	antisense lncRNA gene	4:45012829-45014399	1.69
166Ampd1	0.21	0.75	0.56	0.63	3.61	1.12	0.31	1.85	1.16	1.69	Decreased	adenosine monophosphate deaminase 1	protein coding gene	3:103074013-103099714	1.69
167Chrd	0.06	0.26	0.26	0.39	4.80	1.49	0.31	1.26	1.58	1.69	Decreased	chordin	protein coding gene	16:20725157-20742384	1.69
168Egr1	37.08	97.99	3.83	32.58	2.64	8.50	3.22	1.40	1.09	1.68	increased	early growth response 1	protein coding gene	18:34859822-34864984	1.68

169Gmpr	4.99	0.58	25.17	9.41	0.12	0.37	3.20	3.10	1.42	1.68	increased	guanosine monophosphate reductase	protein coding gene	13:45507443-45546382	1.68
170Hlf	0.18	0.03	0.23	0.13	0.17	0.54	3.20	2.55	0.88	1.68	increased	hepatic leukemia factor	protein coding gene	11:90336535-90390895	1.68
171Hhpl1	1.26	0.59	0.53	0.08	0.47	0.15	0.31	1.09	2.76	1.67	Decreased	hedghog interacting protein-like 1	protein coding gene	12:108306269-108328300	1.67
172Hs3st6	0.88	0.08	0.15	0.05	0.10	0.30	3.18	3.38	1.71	1.67	increased	heparan sulfate (glucosamine) 3-O-sulfotransferase 6	protein coding gene	17:24753002-24758683	1.67
173Bbs10	0.32	0.32	0.48	0.15	0.98	0.31	0.32	0.02	1.69	1.66	Decreased	Bardet-Biedl syndrome 10 (human)	protein coding gene	10:111298678-111301736	1.66
174Gm15947	0.00	0.15	0.03	0.31	35.35	11.16	0.32	1.14	1.48	1.66	Decreased	predicted gene 15947	antisense lncRNA gene	17:23836784-23859776	1.66
175Fibcd1	0.16	0.19	0.03	0.13	1.19	3.75	3.15	0.26	1.91	1.65	increased	fibrinogen C domain containing 1	protein coding gene	2:31813289-31846005	1.65
176Pygm	0.53	2.20	1.04	1.37	4.14	1.32	0.32	2.05	1.40	1.65	Decreased	brain glycogen phosphorylase	protein coding gene	19:6384398-6398459	1.65
177Atf3	2.03	20.83	2.77	9.10	10.27	3.28	0.32	1.36	1.71	1.65	Decreased	activating transcription factor 3	protein coding gene	1:191170295-191183340	1.65
178Drd4	0.15	0.17	0.49	0.18	1.13	0.36	0.32	0.17	1.47	1.65	Decreased	dopamine receptor D4	protein coding gene	7:141292005-141294999	1.65
179Hhex	0.18	0.01	0.68	0.12	0.06	0.18	3.12	1.09	2.45	1.64	increased	hematopoietically expressed homeobox	protein coding gene	19:37432886-37440728	1.64
180Dhcr7	10.25	0.71	8.98	1.93	0.07	0.21	3.12	3.86	2.22	1.64	increased	7-dehydrocholesterol reductase	protein coding gene	7:143823144-143848410	1.64
181Hoxd8	3.96	1.59	0.86	1.07	0.40	1.25	3.12	1.32	1.32	1.64	increased	homeobox D8	protein coding gene	2:74703247-74748274	1.64
182Gtsf1	0.22	0.21	0.25	0.71	0.93	2.88	3.09	0.10	1.53	1.63	increased	gametocyte specific factor 1	protein coding gene	15:103402404-103430470	1.63
183Gm27019	0.30	1.67	0.11	1.83	5.48	16.96	3.09	2.46	1.08	1.63	increased	predicted gene, 27019	lincRNA gene	6:31063016-31063495	1.63
184Ifi207	0.15	0.18	0.14	0.53	1.19	3.67	3.08	0.25	1.88	1.62	increased	interferon activated gene 207	protein coding gene	1:173723429-173741809	1.62
185Cbr2	2.69	2.46	0.79	2.22	0.92	2.82	3.08	0.13	1.50	1.62	increased	carbonyl reductase 2	protein coding gene	11:120729488-120732114	1.62
186Has1	0.44	0.61	0.61	2.59	1.38	4.26	3.08	1.47	1.09	1.62	increased	hyaluronan synthase 1	protein coding gene	17:17843327-17855188	1.62
187Gm13203	0.29	0.40	0.22	0.93	1.38	4.23	3.07	2.46	1.08	1.62	increased	predicted gene 13203	lincRNA gene	4:148713702-148716581	1.62
188Lrrc29	0.42	0.50	0.51	0.20	1.19	0.39	0.33	0.26	1.36	1.61	Decreased	leucine rich repeat containing 29	protein coding gene	8:105312340-105326276	1.61
189Gm9900	0.19	0.22	0.11	0.41	1.20	3.66	3.06	0.26	1.87	1.61	increased	predicted gene 9900	protein coding gene	11:58327720-58330339	1.61
190Slc22a4	0.78	0.23	1.71	1.54	0.29	0.90	3.06	1.77	0.16	1.61	increased	solute carrier family 22 (organic cation transporter), member 5	protein coding gene	11:53983122-54028090	1.61
191Cpt2	6.90	0.69	10.90	3.32	0.10	0.30	3.06	1.33	1.72	1.61	increased	carnitine palmitoyltransferase 2	protein coding gene	4:107903980-107923610	1.61
192Pnck	0.04	0.29	0.05	0.12	7.04	2.31	0.33	1.82	1.21	1.61	Decreased	pregnancy upregulated non-ubiquitously expressed CaM kinase protein coding gene	protein coding gene	X:73655993-73660117	1.61
193Krt18	0.13	0.09	1.45	3.01	0.68	2.08	3.05	0.55	1.06	1.61	increased	keratin 18	protein coding gene	15:102028215-102032026	1.61
194Casp1	0.63	0.19	0.27	0.25	0.30	0.93	3.05	1.72	0.11	1.61	increased	caspase 1	protein coding gene	9:5298516-5307265	1.61
195Plagl1	1.15	0.75	0.12	0.03	0.66	0.22	0.33	0.61	2.22	1.61	Decreased	pleomorphic adenoma gene-like 1	protein coding gene	10:13090690-13131694	1.61
196Gm4617	27.42	6.32	15.36	10.77	0.23	0.70	3.04	2.12	0.51	1.61	increased	predicted pseudogene 4617	pseudogene	3:124385509-124385998	1.61
197Klrtd1	0.10	1.31	0.04	0.19	12.89	4.24	0.33	1.69	1.08	1.60	Decreased	killer cell lectin-like receptor, subfamily D, member 1	protein coding gene	6:129591781-129598775	1.60
198Sp110	0.11	0.15	0.11	0.45	1.37	4.16	3.04	1.45	1.06	1.60	increased	peripheral myelin protein 22	protein coding gene	1:85575675-85598817	1.60
1991500015A07Rik	0.83	9.83	1.26	4.92	11.85	3.90	0.33	0.57	1.96	1.60	Decreased	RIKEN cDNA 1500015A07 gene	antisense lncRNA gene	18:61726389-61786702	1.60
200Kbtbd8	0.45	0.47	0.40	0.14	1.04	0.35	0.33	0.06	1.54	1.59	Decreased	kelch repeat and BTB (POZ) domain containing 8	protein coding gene	6:95114016-95129718	1.59
201Rgs16	13.83	16.93	16.04	6.51	1.22	0.41	0.33	0.29	1.30	1.59	Decreased	regulator of G-protein signaling 16	protein coding gene	1:153740348-153745468	1.59
202Sugct	0.13	0.82	0.15	0.32	6.41	2.13	0.33	1.68	1.09	1.59	Decreased	succinyl-CoA glutarate-CoA transferase	protein coding gene	13:16857854-17694732	1.59
203Dusp6	20.42	15.92	25.04	6.51	0.78	0.26	0.33	0.36	1.94	1.58	Decreased	dual specificity phosphatase 6	protein coding gene	10:99263230-99267488	1.58
2044933421010Rik	0.45	6.76	0.54	2.74	15.14	5.07	0.33	1.92	1.34	1.58	Decreased	RIKEN cDNA 4933421010 gene	unclassified non-coding RNA gene	4:33027141-33031323	1.58
205Fam84b	0.79	0.62	2.40	0.63	0.79	0.26	0.33	0.34	1.92	1.58	Decreased	family with sequence similarity 84, member B	protein coding gene	15:60818995-60831793	1.58
206Gm11351	0.37	0.30	1.86	4.47	0.81	2.40	2.98	0.31	1.27	1.57	increased	predicted gene 11351	lincRNA gene	13:25921032-25930941	1.57
207Ptpnz1	0.06	0.28	0.99	1.57	4.73	1.59	0.34	2.24	0.67	1.57	Decreased	protein tyrosine phosphatase, receptor type Z, polypeptide 1	protein coding gene	6:22875501-23052916	1.57
208Hspa1l	0.98	1.44	1.01	0.50	1.48	0.50	0.34	0.56	1.01	1.57	Decreased	heat shock protein 1-like	protein coding gene	17:34972702-34981788	1.57
209Retn	0.09	1.21	0.02	0.93	12.84	38.15	2.97	1.68	1.25	1.57	increased	resistin	protein coding gene	8:3655769-3660110	1.57
210H2-BI	0.16	0.56	0.05	0.49	3.45	10.23	2.97	1.79	1.35	1.57	increased	histocompatibility 2, blastocyst	polymorphic pseudogene	17:36080114-36084223	1.57
211Sept4	0.08	0.46	0.36	0.68	5.69	1.92	0.34	2.51	1.94	1.57	Decreased	septin 4	protein coding gene	11:87566652-87590539	1.57
212Col22a1	0.62	1.44	0.05	0.34	2.31	6.84	2.96	1.21	1.77	1.57	increased	collagen, type XXII, alpha 1	protein coding gene	15:71795794-72034227	1.57
213Gm2423	0.28	1.21	0.42	0.61	4.29	1.45	0.34	1.10	0.54	1.56	Decreased	predicted gene 2423	pseudogene	13:13232157-13232895	1.56
214Hoxa9	28.96	6.19	0.14	0.01	0.21	0.07	0.34	2.23	1.78	1.56	Decreased	homeobox A7	protein coding gene	6:52223096-52231089	1.56
215Hac1	1.83	0.42	1.15	0.77	0.23	0.67	2.92	2.12	0.57	1.55	increased	keratin associated protein 11-1	protein coding gene	14:31523081-31668579	1.55
2166720427107Rik	0.93	0.51	1.41	0.26	0.55	0.19	0.34	0.87	1.42	1.55	Decreased	RIKEN cDNA 6720427107 gene	lincRNA gene	13:48535522-48537822	1.55
217Clec11a	0.15	0.82	0.69	1.28	5.42	1.86	0.34	2.44	1.89	1.55	Decreased	C-type lectin domain family 11, member a	protein coding gene	7:44303769-44306959	1.55
218Gdnf	0.54	0.17	0.24	0.22	0.32	0.93	2.92	1.65	0.10	1.55	increased	glial cell line derived neurotrophic factor	protein coding gene	15:7811010-7837575	1.55
219lgfbp2	3.29	24.40	0.78	1.98	7.42	2.54	0.34	2.89	1.35	1.54	Decreased	insulin-like growth factor binding protein 2	protein coding gene	1:72824502-728512474	1.54
220Sptlc2	20.78	4.76	34.42	22.99	0.23	0.67	2.92	2.13	0.58	1.54	increased	serine palmitoyltransferase, long chain base subunit 2	protein coding gene	12:87305057-87388355	1.54
221BC025920	0.29	0.90	0.31	0.33	3.08	1.06	0.34	1.62	1.08	1.54	Decreased	cDNA sequence BC025920	protein coding gene	10:81606269-81620703	1.54
222Dynap	0.13	0.19	0.20	0.85	1.42	4.14	2.91	1.51	1.05	1.54	increased	dynactin associated protein	protein coding gene	18:70240428-70244584	1.54
223Syngr3	0.39	0.49	0.47	0.20	1.27	0.44	0.34	1.35	1.19	1.54	Decreased	synaptogrin 3	protein coding gene	17:24685091-24689955	1.54
224Zfp113	1.08	1.04	1.40	0.46	0.96	0.33	0.34	0.05	1.59	1.54	Decreased	zinc finger protein 113	protein coding gene	5:138139701-138155744	1.54
225Sp7	1.25	0.13	0.23	0.07	0.11	0.31	2.90	1.23	1.69	1.54	increased	Sp7 transcription factor 7	protein coding gene	15:102356608-102366325	1.54
226A330069E16Rik	0.60	15.00	0.98	8.51	25.03	8.65	0.35	1.05	1.11	1.53	Decreased	RIKEN cDNA A330069E16 gene	lincRNA gene	2:91237145-91238352	1.53

227Xdh	4.07	2.90	2.55	5.24	0.71	2.05	2.89	0.49	1.04	1.53	increased	xanthine dehydrogenase	protein coding gene	17:73883907-73950182	1.53
228Map3k8	0.52	0.06	0.19	0.06	0.11	0.33	2.89	3.13	1.61	1.53	increased	mitogen-activated protein kinase kinase kinase 8	protein coding gene	18:4331326-4353015	1.53
229Psmb9	1.03	2.48	0.34	2.37	2.42	6.98	2.89	1.27	1.80	1.53	increased	proteasome (prosome, macropain) subunit, beta type 9 (large m	protein coding gene	17:34181987-34197225	1.53
230Mitf	1.02	0.12	0.23	0.08	0.11	0.33	2.88	3.13	1.60	1.53	increased	melanogenesis associated transcription factor	protein coding gene	6:97807057-98021349	1.53
231Spata2	5.25	2.06	5.70	0.78	0.39	0.14	0.35	1.35	2.87	1.52	Decreased	spermatogenesis associated 2	protein coding gene	2:167481132-167516173	1.52
232Zfp830	4.84	3.44	4.09	1.01	0.71	0.25	0.35	0.49	2.01	1.52	Decreased	zinc finger protein 830	protein coding gene	11:82764344-82767622	1.52
233Gm26710	0.17	0.33	0.05	0.27	1.94	5.55	2.86	0.96	2.47	1.51	increased	predicted gene, 26710	lincRNA gene	10:80578646-80580677	1.51
234Prl2c2	14.39	16.52	4.20	13.70	1.15	3.26	2.84	0.20	1.71	1.51	increased	prolactin family 2, subfamily c, member 2	protein coding gene	13:12996131-13005383	1.51
235Wisp2	22.88	5.56	3.35	2.32	0.24	0.69	2.84	2.04	0.53	1.51	increased	WNT1 inducible signaling pathway protein 2	protein coding gene	2:163820860-163833146	1.51
236Dpcr1	0.19	0.45	0.27	0.22	2.35	0.83	0.35	1.23	0.27	1.50	Decreased	diffuse panbronchiolitis critical region 1 (human)	protein coding gene	17:35601548-35605071	1.50
237Fam220a	12.82	6.56	10.96	1.98	0.51	0.18	0.35	0.97	2.47	1.50	Decreased	family with sequence similarity 220, member A	protein coding gene	5:143548705-143564531	1.50
238Gm9958	0.35	2.28	0.42	0.95	6.46	2.28	0.35	2.69	1.19	1.50	Decreased	predicted gene 9958	unclassified gene	5:90367203-90368487	1.50
239Slc16a11	0.70	0.28	0.21	0.25	0.41	1.15	2.83	1.30	1.20	1.50	increased	solute carrier family 16 (monocarboxylic acid transporters), mem	protein coding gene	11:70212751-70216417	1.50
240Mblac2	0.41	0.03	0.36	0.07	0.07	0.21	2.83	3.78	2.28	1.50	increased	metallo-beta-lactamase domain containing 2	protein coding gene	13:81711416-81753275	1.50
241Ptprv	0.68	0.80	0.21	0.69	1.17	3.32	2.83	0.23	0.73	1.50	increased	protein tyrosine phosphatase, receptor type, V	protein coding gene	1:135108496-135132594	1.50
242B930036N10Ri	0.05	1.63	0.09	1.10	32.66	11.57	0.35	0.03	0.53	1.50	Decreased	RIKEN cDNA B930036N10 gene	antisense lincRNA gene	1:171500567-171504750	1.50
2431700086P04Rik	0.18	0.89	0.16	0.28	4.82	1.71	0.35	2.27	0.77	1.50	Decreased	RIKEN cDNA 1700086P04 gene	lincRNA gene	4:129050096-129057983	1.50
244Bhlhe22	4.61	1.10	0.14	0.10	0.24	0.68	2.82	2.05	0.56	1.50	increased	basic helix-loop-helix family, member e22	protein coding gene	3:18054324-18057509	1.50
245Tktl1	0.05	0.61	0.05	0.24	12.93	4.59	0.36	1.69	0.20	1.49	Decreased	transketolase-like 1	protein coding gene	X:74177258-74208500	1.49
246A930033H14Ri	1.02	1.01	0.49	0.17	0.98	0.35	0.36	0.02	0.52	1.49	Decreased	RIKEN cDNA A930033H14 gene	protein coding gene	10:69151433-69291791	1.49
247Ubac1	26.01	8.01	37.80	32.74	0.31	0.87	2.81	1.70	0.21	1.49	increased	ubiquitin associated domain containing 1	protein coding gene	2:25998542-26021747	1.49
248Hdh3	1.91	0.07	2.43	0.25	0.04	0.10	2.81	4.78	0.30	1.49	increased	haloacid dehalogenase-like hydrolase domain containing 3	protein coding gene	4:62499007-62503467	1.49
249Cst6	0.28	0.43	0.38	1.66	1.58	4.43	2.80	1.66	0.15	1.49	increased	cystatin B	protein coding gene	19:5344704-5349574	1.49
250Laptm4b	17.67	4.00	19.27	12.23	0.23	0.63	2.80	2.14	0.66	1.49	increased	lysosomal-associated protein transmembrane 4B	protein coding gene	15:34238025-34284295	1.49
251Wnt2	1.33	0.49	0.52	0.54	0.37	1.03	2.80	1.44	0.05	1.48	increased	wingless-type MMTV integration site family, member 2	protein coding gene	6:17988939-18030585	1.48
252Mtn1	0.43	0.58	0.42	0.21	1.36	0.49	0.36	0.44	1.04	1.48	Decreased	X-linked myotubular myopathy gene 1	protein coding gene	X:71210766-71315691	1.48
253Tnfaip3	4.52	2.14	6.78	1.15	0.47	0.17	0.36	1.08	2.56	1.48	Decreased	tumor necrosis factor, alpha-induced protein 3	protein coding gene	10:19000909-19015657	1.48
254Gm20554	0.43	0.19	0.60	0.10	0.44	0.16	0.36	1.17	0.65	1.48	Decreased	predicted gene, 20554	lincRNA gene	13:72626173-72627013	1.48
255B230208H11Rik	0.35	0.32	1.14	0.38	0.92	0.33	0.36	0.12	1.60	1.48	Decreased	RIKEN cDNA B230208H11 gene	lincRNA gene	10:12916647-12923100	1.48
256Acot1	0.58	0.13	3.25	2.03	0.22	0.63	2.78	2.15	0.68	1.48	increased	acyl-CoA thioesterase 1	protein coding gene	12:84009501-84017670	1.48
257Kin	7.51	7.29	12.74	4.45	0.97	0.35	0.36	0.04	1.52	1.48	Decreased	Kin17 DNA and RNA binding protein	protein coding gene	2:10080592-10092806	1.48
258Ptma	####	####	####	####	0.23	0.65	2.78	2.09	0.62	1.48	increased	prothymosin alpha	protein coding gene	1:86526735-86530698	1.48
259Gm10031	0.32	0.09	0.21	0.17	0.29	0.81	2.78	1.78	0.31	1.47	increased	predicted pseudogene 10031	protein coding gene	1:156524091-156526659	1.47
260Zic1	12.48	2.64	0.20	0.02	0.21	0.08	0.36	2.24	0.71	1.46	Decreased	zinc finger protein of the cerebellum 1	protein coding gene	9:91358057-91389348	1.46
261Rpl39l	####	####	0.78	0.77	2.72	0.99	0.36	1.44	0.02	1.46	Decreased	ribosomal protein L39-like	protein coding gene	16:10170225-10174911	1.46
262Sat1	62.05	27.37	47.57	57.60	0.44	1.21	2.75	1.18	0.28	1.46	increased	spermidine/spermine N1-acetyl transferase 1	protein coding gene	X:155213131-155216449	1.46
263Ppp1r3b	0.38	0.01	0.73	0.07	0.04	0.10	2.74	4.51	0.35	1.46	increased	protein phosphatase 1, regulatory (inhibitor) subunit 3B	protein coding gene	8:35375740-35388124	1.46
264Sema6c	0.55	0.18	0.83	0.10	0.34	0.12	0.37	1.58	0.03	1.45	Decreased	sema domain, transmembrane domain (TM), and cytoplasmic d	protein coding gene	3:95160456-95174024	1.45
265C1qtnf3	4.97	18.69	0.09	0.12	3.76	1.38	0.37	0.91	0.46	1.45	Decreased	C1q and tumor necrosis factor related protein 3	protein coding gene	15:10952331-10980150	1.45
266E130201H02Rik	0.38	0.89	0.40	0.34	2.36	0.87	0.37	1.24	0.21	1.45	Decreased	RIKEN cDNA E130201H02 gene	pseudogene	7:120597624-120598475	1.45
267Zfp39	0.61	0.59	0.58	0.20	0.97	0.36	0.37	0.05	1.49	1.45	Decreased	zinc finger protein 39	protein coding gene	11:58888152-58904225	1.45
268Gm5388	1.98	0.59	1.65	1.35	0.30	0.82	2.72	1.73	0.29	1.44	increased	predicted gene 5388	pseudogene	X:51218876-51219993	1.44
269Mcf2l	4.15	0.65	0.63	0.04	0.16	0.06	0.37	2.67	0.11	1.44	Decreased	mcf2 transforming sequence-like	protein coding gene	8:12757013-13020905	1.44
270Cd274	0.28	0.07	0.34	0.24	0.26	0.71	2.71	1.94	0.50	1.44	increased	CD274 antigen	protein coding gene	19:29367454-29388095	1.44
271Zfp119b	0.36	2.81	0.51	1.47	7.84	2.90	0.37	0.97	0.54	1.44	Decreased	zinc finger protein 119b	protein coding gene	17:55938552-55945174	1.44
272Pcdhb9	0.17	0.02	0.13	0.04	0.11	0.30	2.70	3.18	1.74	1.43	increased	protocadherin beta 9	protein coding gene	18:37400854-37403909	1.43
273Ngef	1.52	0.73	0.85	1.11	0.48	1.30	2.70	1.05	0.38	1.43	increased	neuronal guanine nucleotide exchange factor	protein coding gene	1:87476833-87597845	1.43
274Thap1	12.66	####	9.52	####	57.03	21.19	0.37	0.93	0.40	1.43	Decreased	THAP domain containing, apoptosis associated protein 1	protein coding gene	8:26158211-26162940	1.43
275Gm10263	0.73	6.64	2.57	8.64	9.05	3.36	0.37	1.18	0.75	1.43	Decreased	predicted gene 10263	pseudogene	7:75366054-75366264	1.43
276Gm16174	1.03	2.87	0.80	5.97	2.78	7.47	2.69	1.47	0.90	1.43	increased	predicted gene 16174	antisense lincRNA gene	7:19808461-19822770	1.43
277Gadd45g	78.23	####	####	####	3.94	1.47	0.37	0.98	1.55	1.43	Decreased	growth arrest and DNA-damage-inducible 45 gamma	protein coding gene	13:51846677-51848468	1.43
278Gdf7	0.60	0.45	0.61	0.17	0.74	0.28	0.37	0.43	0.86	1.43	Decreased	growth differentiation factor 7	protein coding gene	12:8297933-8301954	1.43
279Nfatc4	5.31	1.42	1.61	1.15	0.27	0.72	2.69	1.91	0.48	1.43	increased	nuclear factor of activated T cells, cytoplasmic, calcineurin depe	protein coding gene	14:55824794-55833943	1.43
280Gm12350	2.59	2.19	2.56	5.79	0.84	2.26	2.68	0.25	1.18	1.42	increased	predicted gene 12350	pseudogene	4:34485831-34486293	1.42
281Adck2	4.56	0.79	3.27	1.52	0.17	0.47	2.68	2.53	1.10	1.42	increased	aarf domain containing kinase 2	protein coding gene	6:39573872-39725463	1.42
2824933406P04Rik	0.16	6.03	0.13	1.76	36.95	13.77	0.37	0.33	0.78	1.42	Decreased	RIKEN cDNA 4933406P04 gene	lincRNA gene	10:20310505-20312294	1.42
283Sema7a	0.89	0.32	0.14	0.13	0.35	0.95	2.68	1.50	0.07	1.42	increased	sema domain, immunoglobulin domain (Ig), and GPI membrane	protein coding gene	9:57940112-57962864	1.42
284Mapk13	0.62	2.08	0.04	0.34	3.35	8.97	2.68	1.74	0.17	1.42	increased	mitogen-activated protein kinase 13	protein coding gene	17:28769306-28778698	1.42

285Trib1	4.12	2.25	9.37	1.91	0.55	0.20	0.37	0.87	2.29	1.42	Decreased	tribbles pseudokinase 1	protein coding gene	15:59648349-59656550	1.42
286Csgalnact1	1.39	0.26	0.22	0.11	0.19	0.50	2.67	2.42	1.01	1.42	Increased	chondroitin sulfate N-acetylglucosaminyltransferase 1	protein coding gene	8:68356780-68735146	1.42
287Adamts1	1.94	4.97	0.26	0.25	2.56	0.96	0.37	1.36	0.06	1.42	Decreased	ADAMTS-like 1	protein coding gene	4:86053914-86428385	1.42
288Gm26798	0.12	0.39	0.11	0.13	3.17	1.19	0.38	1.66	0.25	1.41	Decreased	predicted gene, 26798	lincRNA gene	15:89211558-89215012	1.41
289Gm10101	0.09	2.91	0.12	1.53	33.27	12.48	0.38	0.96	0.64	1.41	Decreased	predicted gene 10101	lincRNA gene	14:46829108-46832027	1.41
290Kif22	37.88	6.92	37.21	18.08	0.18	0.49	2.66	2.45	1.04	1.41	Increased	kinesin family member 22	protein coding gene	7:127027731-127042420	1.41
291Lyn	0.47	0.31	1.35	0.34	0.68	0.25	0.38	0.57	1.98	1.41	Decreased	LYN proto-oncogene, Src family tyrosine kinase	protein coding gene	4:3678114-3813122	1.41
292Ncf2	0.04	0.23	0.10	0.19	5.28	1.99	0.38	2.40	1.99	1.41	Decreased	neutrophil cytosolic factor 2	protein coding gene	1:152807889-152836990	1.41
293Pcca	8.41	10.48	7.23	23.83	1.24	3.30	2.65	1.12	0.72	1.41	Increased	propionyl-Coenzyme A carboxylase, alpha polypeptide	protein coding gene	14:122534323-122984035	1.41
294Gm12396	0.91	1.00	0.58	1.68	1.10	2.92	2.65	1.14	0.55	1.41	Increased	predicted gene 12396	pseudogene	4:40775132-40775725	1.41
295Gm10364	0.18	0.61	0.29	0.38	3.44	1.30	0.38	1.78	0.38	1.40	Decreased	predicted pseudogene 10364	pseudogene	14:55038461-55042979	1.40
296A4galt	0.59	0.06	0.13	0.03	0.10	0.26	2.64	3.35	1.94	1.40	Increased	alpha 1,4-galactosyltransferase	protein coding gene	15:83226725-83251774	1.40
297Arl6ip1	40.13	7.96	39.20	20.53	0.20	0.52	2.64	2.33	0.93	1.40	Increased	ADP-ribosylation factor-like 6 interacting protein 1	protein coding gene	7:118118890-118129625	1.40
298Sirt1	4.33	4.03	1.86	0.66	0.93	0.35	0.38	0.10	1.50	1.39	Decreased	sirtuin 1	protein coding gene	10:63319004-63381704	1.39
299Satb2	2.29	0.55	1.22	0.77	0.24	0.63	2.63	2.06	0.67	1.39	Increased	special AT-rich sequence binding protein 2	protein coding gene	1:56793985-56978650	1.39
300Gm12504	1.14	0.43	1.09	1.07	0.37	0.98	2.63	1.42	0.03	1.39	Increased	predicted pseudogene 12504	pseudogene	4:44121892-44124123	1.39
301Gm15361	0.51	1.82	0.93	1.27	3.58	1.37	0.38	1.84	0.45	1.39	Decreased	predicted pseudogene 15361	pseudogene	X:74494378-74494822	1.39
302Hoxa5	3.47	1.03	0.29	0.23	0.30	0.77	2.62	1.76	0.37	1.39	Increased	homeobox A5	protein coding gene	5:52162350-52221854	1.39
303Lph-ps1	0.68	3.22	1.61	2.92	4.74	1.81	0.38	1.24	0.86	1.39	Decreased	LLP homolog, pseudogene 1	pseudogene	4:116200562-116200952	1.39
304Arhgap27	0.78	2.33	0.74	0.84	2.98	1.14	0.38	1.58	0.19	1.39	Decreased	Rho GTPase activating protein 27	protein coding gene	11:103331496-103412687	1.39
305Gm5776	1.15	1.70	0.77	2.98	1.48	3.85	2.60	1.56	0.94	1.38	Increased	predicted gene 5776	pseudogene	7:54411908-54412519	1.38
306Erg28	38.18	7.86	39.46	21.12	0.21	0.54	2.60	2.28	0.90	1.38	Increased	ergosterol biosynthesis 28	protein coding gene	12:85815447-85824550	1.38
307Gm12537	0.16	0.19	0.25	0.76	1.19	3.07	2.59	1.24	0.62	1.38	Increased	predicted gene 12537	pseudogene	4:57370952-57371954	1.38
308Ccdc120	3.09	0.36	4.06	1.24	0.12	0.31	2.59	3.08	1.71	1.37	Increased	coiled-coil domain containing 120	protein coding gene	X:7731713-7750905	1.37
309Ttc41	0.18	0.19	0.22	0.09	1.06	0.41	0.39	1.08	1.29	1.37	Decreased	tetratricopeptide repeat domain 41	protein coding gene	10:86705810-86776843	1.37
310Ras11a	8.37	2.74	13.29	1.68	0.33	0.13	0.39	1.61	2.99	1.37	Decreased	RAS-like, family 11, member A	protein coding gene	5:146845070-146847726	1.37
311Zbtb6	2.44	2.68	3.00	1.28	1.10	0.42	0.39	1.14	1.24	1.37	Decreased	zinc finger and BTB domain containing 6	protein coding gene	2:37425499-37443171	1.37
312Zfp518b	2.10	1.55	2.05	0.59	0.74	0.29	0.39	0.44	1.81	1.37	Decreased	zinc finger protein 518B	protein coding gene	5:38668483-38684826	1.37
313Gm9938	1.01	0.33	0.56	0.07	0.33	0.13	0.39	1.59	0.95	1.36	Decreased	predicted gene 9938	protein coding gene	19:23723278-23725034	1.36
314Sprtn	4.11	2.97	2.50	0.71	0.72	0.28	0.39	0.47	1.83	1.36	Decreased	SprT-like N-terminal domain	protein coding gene	8:124897885-124903813	1.36
315Jph2	0.09	0.23	0.15	0.14	2.44	0.95	0.39	1.29	0.07	1.36	Decreased	junctophilin 2	protein coding gene	2:163336241-163397993	1.36
316Zfp420	0.76	0.30	0.46	0.46	0.40	1.01	2.56	1.34	1.02	1.36	Increased	zinc finger protein 420	protein coding gene	7:29859978-29877292	1.36
317Sh2d3c	0.19	0.02	0.34	0.02	0.13	0.05	0.39	2.98	0.34	1.35	Decreased	SH2 domain containing 3C	protein coding gene	2:32721054-32755512	1.35
318Arhgef37	0.19	0.08	0.16	0.17	0.42	1.06	2.56	1.27	0.09	1.35	Increased	Rho guanine nucleotide exchange factor (GEF) 37	protein coding gene	18:61493793-61536536	1.35
319Gm16589	0.22	0.52	0.31	0.29	2.42	0.95	0.39	1.28	0.08	1.35	Decreased	predicted gene 16589	pseudogene	8:4087542-40889900	1.35
320Gm7670	0.23	0.34	0.36	1.36	1.47	3.74	2.55	1.55	0.90	1.35	Increased	predicted gene 7670	pseudogene	18:16839316-16839957	1.35
321Nrp2	3.51	0.48	5.27	1.83	0.14	0.35	2.55	2.88	1.53	1.35	Increased	neuropilin 2	protein coding gene	1:62703284-62818695	1.35
322Lama5	0.45	2.54	41.35	92.32	5.66	2.23	0.39	2.50	0.16	1.34	Decreased	laminin, alpha 5	protein coding gene	2:180176372-180225859	1.34
323Polrmt	21.16	5.17	22.84	14.13	0.24	0.62	2.53	2.03	0.69	1.34	Increased	polymerase (RNA) mitochondrial (DNA directed)	protein coding gene	10:79736122-79746581	1.34
324Fzd8	4.20	2.68	2.33	0.59	0.64	0.25	0.39	0.65	1.99	1.34	Decreased	frizzled class receptor 8	protein coding gene	18:9212855-9216201	1.34
325Gm15766	0.24	2.76	0.90	4.14	11.71	4.62	0.39	1.55	1.21	1.34	Decreased	predicted gene 15766	pseudogene	6:147271836-147272649	1.34
326Tob1	6.98	25.79	5.29	7.74	3.70	1.46	0.40	1.89	1.55	1.34	Decreased	transducer of ErbB-2.1	protein coding gene	11:94211453-94215495	1.34
327Commd4	32.12	14.90	36.77	43.07	0.46	1.17	2.52	1.11	0.23	1.34	Increased	COMM domain containing 4	protein coding gene	9:57155040-57158299	1.34
328Gm20628	0.05	0.73	0.06	0.32	13.41	5.32	0.40	1.75	0.41	1.33	Decreased	predicted gene 20628	antisense lincRNA gene	3:96267074-96279001	1.33
329Gm27042	0.60	0.57	0.59	0.22	0.94	0.37	0.40	0.09	1.42	1.33	Decreased	predicted gene, 27042	lincRNA gene	19:40589434-40594862	1.33
330Gm4473	0.59	1.61	0.35	0.38	2.72	1.08	0.40	1.44	1.12	1.33	Decreased	predicted gene 4473	bidirectional promoter lincRNA gene	10:12089156-120901888	1.33
331Gm12643	0.90	1.01	0.86	2.41	1.12	2.80	2.50	1.16	0.49	1.32	Increased	predicted gene 12643	pseudogene	4:91037899-91038273	1.32
332Prorsd1	6.20	22.33	7.77	11.20	3.60	1.44	0.40	1.85	0.53	1.32	Decreased	prolyl-4RNA synthetase domain containing 1	protein coding gene	11:29511756-29515033	1.32
333Gm10540	2.04	2.19	1.44	3.87	1.07	2.69	2.50	1.10	0.43	1.32	Increased	predicted gene 10540	pseudogene	18:45268859-45896509	1.32
334Gm11191	0.11	0.25	0.16	0.85	2.18	5.45	2.50	1.12	0.45	1.32	Increased	predicted gene 11191	pseudogene	11:77747040-77748041	1.32
335Mdfl	1.48	0.42	0.59	0.07	0.28	0.11	0.40	1.82	0.14	1.32	Decreased	MyoD family inhibitor	protein coding gene	17:47815327-47834691	1.32
336Rwd2a	0.52	1.52	0.37	0.44	2.94	1.18	0.40	1.56	0.24	1.32	Decreased	RWD domain containing 2A	protein coding gene	9:86572052-86574896	1.32
337Ctsf	1.04	1.74	0.61	2.54	1.67	4.18	2.50	0.74	0.06	1.32	Increased	cathepsin F	protein coding gene	19:4855128-4860912	1.32
338Ankrd49	3.97	1.57	4.65	0.74	0.40	0.16	0.40	1.34	0.65	1.32	Decreased	ankyrin repeat domain 49	protein coding gene	9:14780208-14782964	1.32
339Dmrt2	0.19	0.03	0.94	0.07	0.18	0.07	0.40	2.49	0.81	1.32	Decreased	doublesex and mab-3 related transcription factor 2	protein coding gene	19:25672419-25679010	1.32
340Gm4963	1.01	1.35	0.89	2.97	1.34	3.33	2.49	1.42	0.74	1.32	Increased	predicted gene 4963	pseudogene	5:137993046-137993424	1.32
341Ramp1	0.09	0.76	0.16	0.54	8.27	3.33	0.40	1.05	0.74	1.31	Decreased	receptor (calcitonin) activity modifying protein 1	protein coding gene	1:91179821-91225196	1.31
342Cabp4	0.09	0.17	0.08	0.40	1.98	4.93	2.48	1.99	0.30	1.31	Increased	calcium binding protein 4	protein coding gene	19:4135424-4139609	1.31

343Gm11450	3.18	2.88	4.66	10.48	0.91	2.25	2.48	0.14	0.17	1.31	increased	predicted gene 11450	pseudogene	2:162805589-162805922	1.31
344Rbm12b1	1.00	5.40	1.32	2.86	5.38	2.17	0.40	2.43	0.12	1.31	Decreased	RNA binding motif protein 12 B1	protein coding gene	4:12140263-12146731	1.31
345Gm14017	1.17	4.80	1.60	2.65	4.09	1.65	0.40	2.03	0.72	1.31	Decreased	predicted gene 14017	pseudogene	2:121481128-121481514	1.31
346Zbtb45	7.08	0.50	6.65	1.16	0.07	0.17	2.48	3.83	2.52	1.31	increased	zinc finger and BTB domain containing 45	protein coding gene	7:13004935-13009822	1.31
347Pmp22	33.50	3.53	25.10	6.56	0.11	0.26	2.48	3.25	1.93	1.31	increased	peripheral myelin protein 22	protein coding gene	11:63128981-63159547	1.31
348Dlx2	15.13	8.08	22.78	4.91	0.53	0.22	0.40	0.91	2.21	1.31	Decreased	distal-less homeobox 2	protein coding gene	2:71543408-71546754	1.31
349Gm10524	0.58	6.82	0.78	3.66	11.67	4.72	0.40	1.54	2.24	1.31	Decreased	predicted gene 10524	unclassified gene	18:82692948-82706174	1.31
350Ccl20	0.24	1.60	0.01	0.23	6.66	16.46	2.47	2.74	0.04	1.31	increased	chemokine (C-C motif) ligand 20	protein coding gene	1:83116765-83119167	1.31
351Mmp9	2.32	9.72	0.67	1.13	4.19	1.69	0.40	2.07	0.76	1.31	Decreased	matrix metalloproteinase 9	protein coding gene	2:164940779-164955850	1.31
3524930465K10Rik	0.08	1.13	0.07	0.41	13.85	5.61	0.41	1.79	0.49	1.30	Decreased	RIKEN cDNA 4930465K10 gene	unclassified non-coding RNA gene	18:77714209-77715449	1.30
353Prss22	1.58	0.25	0.25	0.10	0.16	0.40	2.47	2.63	1.33	1.30	increased	protease, serine 22	protein coding gene	17:23993534-23998100	1.30
354Uap11l	8.90	10.06	20.64	57.52	1.13	2.79	2.47	0.18	2.48	1.30	increased	UDP-N-acteylglucosamine pyrophosphorylase 1-like 1	protein coding gene	2:25359888-25365682	1.30
355Gm11408	2.11	2.47	2.10	6.07	1.17	2.89	2.46	0.23	0.53	1.30	increased	predicted gene 11408	pseudogene	4:80003395-80003914	1.30
356Rpl7a-ps8	0.16	0.17	0.25	0.65	1.08	2.66	2.46	0.11	0.41	1.30	increased	ribosomal protein L7a_pseudogene 8	pseudogene	7:19953782-19954583	1.30
357A930002H24Ri	1.24	9.49	0.77	2.39	7.66	3.11	0.41	0.94	0.64	1.30	Decreased	RIKEN cDNA A930002H24 gene	unclassified gene	17:63863299-63863791	1.30
358Tspan12	0.28	0.07	1.04	0.61	0.24	0.59	2.46	2.07	0.77	1.30	increased	tetraspanin 12	protein coding gene	6:21771394-21852515	1.30
359Tbr1	0.02	0.21	0.03	0.13	10.91	4.44	0.41	1.45	0.15	1.30	Decreased	T-box brain gene 1	protein coding gene	2:61802929-61814114	1.30
360Hexim2	2.05	1.45	1.92	0.55	0.71	0.29	0.41	0.50	1.80	1.30	Decreased	hexamethylene bis-acetamide inducible 2	protein coding gene	11:103132428-103139876	1.30
361Naif1	1.99	1.43	1.95	0.57	0.72	0.29	0.41	0.48	0.77	1.29	Decreased	nuclear apoptosis inducing factor 1	protein coding gene	2:32414486-32456953	1.29
362Ralgds	18.35	0.55	3.55	0.26	0.03	0.07	2.45	0.05	0.76	1.29	increased	ral guanine nucleotide dissociation stimulator	protein coding gene	2:28513124-28553081	1.29
3631700056E22Rik	0.41	6.66	0.64	4.24	16.31	6.66	0.41	0.03	0.73	1.29	Decreased	RIKEN cDNA 1700056E22 gene	antisense lncRNA gene	1:184013301-184075636	1.29
364Cdk1	####	25.35	####	61.63	0.09	0.23	2.45	3.44	2.15	1.29	increased	cyclin-dependent kinase 1	protein coding gene	10:69336634-69352938	1.29
365Llph-ps2	1.84	6.96	4.27	6.60	3.78	1.55	0.41	1.92	0.22	1.29	Decreased	LLP homolog_pseudogene 2	pseudogene	X:13217922-13218650	1.29
366Rpusd2	1.65	0.22	2.73	0.15	0.14	0.06	0.41	2.88	0.12	1.29	Decreased	RNA pseudouridylyl synthase domain containing 2	protein coding gene	2:119034789-119039769	1.29
367Arl5c	9.87	####	7.72	####	32.81	13.43	0.41	0.04	0.75	1.29	Decreased	ADP-ribosylation factor-like 5C	protein coding gene	11:97989577-97996181	1.29
368Nfe2l3	0.20	0.52	0.90	0.98	2.65	1.09	0.41	1.41	0.12	1.29	Decreased	nuclear factor, erythroid derived 2, like 3	protein coding gene	6:51432669-51458768	1.29
369Gm17501	2.18	14.80	2.59	7.22	6.79	2.78	0.41	2.76	0.48	1.29	Decreased	predicted gene, 17501	lincRNA gene	3:145650311-145652093	1.29
370Tmtt12	1.26	0.76	1.73	0.42	0.60	0.25	0.41	0.74	2.02	1.28	Decreased	tRNA methyltransferase 12	protein coding gene	15:58872648-58876781	1.28
371Gm26734	0.06	0.40	0.05	0.16	6.96	2.86	0.41	0.80	0.51	1.28	Decreased	predicted gene, 26734	lincRNA gene	17:94834905-94836642	1.28
372Fgfr3	0.21	0.22	0.38	0.16	1.02	0.42	0.41	1.03	1.25	1.28	Decreased	fibroblast growth factor receptor 3	protein coding gene	5:33721723-33737068	1.28
373Hist1h3h	0.34	1.55	1.20	2.27	4.60	1.89	0.41	2.20	0.92	1.28	Decreased	histone cluster 1, H3b	protein coding gene	13:21717627-21718115	1.28
374Mob3c	1.41	0.51	2.44	0.37	0.36	0.15	0.41	1.46	2.74	1.28	Decreased	MOB kinase activator 3C	protein coding gene	4:115828091-115836185	1.28
375Frat2	0.87	5.47	1.02	2.64	6.28	2.59	0.41	2.65	0.37	1.28	Decreased	frequently rearranged in advanced T cell lymphomas 2	protein coding gene	19:41845977-41848132	1.28
376Tbxa2r	0.19	0.95	0.15	0.29	4.87	2.01	0.41	2.29	0.01	1.28	Decreased	thromboxane A2 receptor	protein coding gene	10:81328731-81335171	1.28
377Rnf113a1	3.19	3.29	3.73	1.59	1.03	0.43	0.41	1.05	1.23	1.28	Decreased	ring finger protein 113A1	protein coding gene	X:37191244-37192462	1.28
378Hoxd9	1.43	0.19	0.30	0.10	0.13	0.32	2.42	2.01	1.63	1.27	increased	homeobox D9	protein coding gene	2:74697679-74700208	1.27
379Gm6564	0.48	0.42	0.77	1.62	0.86	2.09	2.42	0.21	0.06	1.27	increased	predicted pseudogene 6564	pseudogene	5:22931389-22932034	1.27
380Magi2	0.13	0.37	0.11	0.13	2.90	1.20	0.41	1.53	0.27	1.27	Decreased	membrane associated guanylate kinase, WW and PDZ domain	protein coding gene	5:19227045-20706049	1.27
3812610037D02Rik	0.06	1.21	0.07	0.59	19.45	8.07	0.42	2.28	0.01	1.27	Decreased	RIKEN cDNA 2610037D02 gene	lincRNA gene	15:96134818-96284495	1.27
382Fam19a5	5.72	1.18	0.47	0.04	0.21	0.09	0.42	2.28	0.55	1.27	Decreased	family with sequence similarity 19, member A5	protein coding gene	15:87625229-87759336	1.27
383Aldoc	0.77	1.04	0.35	1.13	1.34	3.21	2.40	0.42	0.68	1.26	increased	aldolase C, fructose-bisphosphate	protein coding gene	11:78322967-78327781	1.26
384Arntl	3.63	2.83	0.88	0.29	0.78	0.33	0.42	0.36	1.61	1.26	Decreased	aryl hydrocarbon receptor nuclear translocator-like	protein coding gene	7:113207464-113341422	1.26
385Rps8-ps2	0.23	1.57	0.57	1.63	6.87	2.87	0.42	2.78	0.52	1.26	Decreased	ribosomal protein S8_pseudogene 2	pseudogene	2:175804982-175805609	1.26
386C030037D09Rik	0.31	0.81	0.10	0.66	2.63	6.27	2.39	1.39	0.65	1.26	increased	RIKEN cDNA C030037D09 gene	lincRNA gene	11:88718644-88728893	1.26
387Coa7	16.15	1.79	36.59	9.64	0.11	0.26	2.38	3.18	1.92	1.25	increased	cytochrome c oxidase assembly factor 7	protein coding gene	4:108328139-108341544	1.25
388Gm5471	0.78	2.16	0.39	2.58	2.79	6.64	2.38	1.48	0.73	1.25	increased	predicted pseudogene 5471	pseudogene	15:44971793-44972135	1.25
389Sprr2k	7.39	6.45	1.83	3.81	0.87	2.08	2.38	0.20	0.06	1.25	increased	small proline-rich protein 2K	protein coding gene	3:92432581-92433925	1.25
390Alox8	0.09	15.69	0.07	4.66	170.16	71.53	0.42	0.02	0.08	1.25	Decreased	arachidonate 8-lipoxygenase	protein coding gene	11:69183931-69197843	1.25
391Mgp	58.08	####	9.35	39.50	1.78	4.22	2.38	0.83	0.08	1.25	increased	matrix Gla protein	protein coding gene	6:136872435-136875805	1.25
392Folr1	0.10	0.62	0.63	1.70	6.42	2.70	0.42	2.68	1.44	1.25	Decreased	folate receptor 1 (adult)	protein coding gene	7:101858330-101870788	1.25
393Enpp2	2.27	2.91	0.73	2.22	1.28	3.04	2.37	0.36	0.61	1.25	increased	ectonucleotide pyrophosphatase/phosphodiesterase 2	protein coding gene	15:54838897-54920146	1.25
394Npy1r	0.46	0.46	0.76	0.33	1.02	0.43	0.42	0.02	1.22	1.25	Decreased	neuropeptide Y receptor Y1	protein coding gene	8:66697421-66706798	1.25
395Sde2	7.02	6.30	6.67	2.53	0.90	0.38	0.42	0.16	1.40	1.25	Decreased	SDE2 telomere maintenance homolog (S. pombe)	protein coding gene	1:180851150-180868097	1.25
396Fam110b	5.35	1.75	4.63	0.64	0.33	0.14	0.42	1.61	2.86	1.25	Decreased	family with sequence similarity 110, member B	protein coding gene	4:5644089-6108223	1.25
3971700056N10Rik	0.20	1.29	0.22	0.60	6.58	2.78	0.42	2.72	1.47	1.24	Decreased	RIKEN cDNA 1700056N10 gene	lincRNA gene	16:17047519-17049111	1.24
398Sod3	3.11	2.90	5.04	11.14	0.93	2.21	2.37	0.10	0.15	1.24	increased	superoxide dismutase 3, extracellular	protein coding gene	5:52363945-52371405	1.24
399Xaf1	1.82	15.72	15.95	58.18	8.64	3.65	0.42	0.11	0.87	1.24	Decreased	XIAP associated factor 1	protein coding gene	11:72301628-72313733	1.24
400AA986860	0.33	0.35	1.08	0.49	1.06	0.45	0.42	0.09	0.15	1.24	Decreased	expressed sequence AA986860	protein coding gene	1:130731975-130744621	1.24

401Gldc	0.45	0.13	0.23	0.16	0.29	0.68	2.36	1.80	0.56	1.24	increased	glycine decarboxylase	protein coding gene	19:30098448-30175418	1.24
402Gm11814	0.40	0.47	0.29	0.80	1.18	2.79	2.36	1.24	1.48	1.24	increased	predicted gene 11814	pseudogene	4:10438814-10439743	1.24
403Dcun1d3	2.97	1.03	1.44	0.21	0.35	0.15	0.42	1.53	2.77	1.24	Decreased	DCN1, defective in cullin neddylation 1, domain containing 3 (S	protein coding gene	7:119853162-119916756	1.24
404Epm2aip1	3.72	2.27	3.01	0.78	0.61	0.28	0.42	0.71	1.95	1.24	Decreased	EPM2A (laforin) interacting protein 1	protein coding gene	9:111271844-111279089	1.24
405AV099323	0.72	1.35	0.53	2.32	1.87	4.40	2.35	0.90	0.14	1.24	increased	expressed sequence AV099323	antisense lncRNA gene	2:132253356-132261333	1.24
406Tgfb1	25.02	9.07	5.77	4.93	0.36	0.85	2.35	1.46	0.23	1.23	increased	transforming growth factor, beta induced	protein coding gene	13:56609602-56639339	1.23
407Zkscan14	1.00	2.46	0.96	1.00	2.47	1.05	0.43	1.30	0.07	1.23	Decreased	zinc finger with KRAB and SCAN domains 14	protein coding gene	5:145194945-145201868	1.23
408Casp4	0.81	1.05	2.09	6.35	1.30	3.04	2.35	0.37	1.60	1.23	increased	caspase 1	protein coding gene	9:5308827-5336783	1.23
409Rab7b	5.65	0.41	5.61	0.96	0.07	0.17	2.34	0.78	0.35	1.23	increased	RAB7B, member RAS oncogene family	protein coding gene	1:131688694-131713542	1.23
410Gm5778	0.31	1.37	0.47	0.88	4.41	1.88	0.43	1.14	0.91	1.23	Decreased	predicted gene 5778	pseudogene	10:73099341-74646314	1.23
411Atp2b4	0.31	0.17	0.77	0.18	0.54	0.23	0.43	0.89	0.12	1.23	Decreased	ATPase, Ca++ transporting, plasma membrane 4	protein coding gene	1:133702673-133801041	1.23
412Fscn1	####	77.69	####	####	0.28	0.65	2.34	1.85	0.62	1.22	increased	fascin actin-bundling protein 1	protein coding gene	5:142960342-142973185	1.22
413Gm9892	1.39	9.99	2.08	6.39	7.16	3.07	0.43	0.84	1.62	1.22	Decreased	predicted gene 9892	pseudogene	8:52195779-52197160	1.22
414Mroh2a	1.13	0.90	1.40	0.48	0.80	0.34	0.43	0.33	1.55	1.22	Decreased	maestro heat-like repeat family member 2A	protein coding gene	1:88226985-88262289	1.22
4152610307P16Rik	1.97	7.27	1.69	2.68	3.69	1.58	0.43	1.88	0.66	1.22	Decreased	RIKEN cDNA 2610307P16 gene	unclassified non-coding RNA gene	13:28460033-28885620	1.22
416Gm26707	0.04	0.54	0.03	0.18	15.05	6.46	0.43	0.91	0.69	1.22	Decreased	predicted gene, 26707	lincRNA gene	7:25802474-25820654	1.22
417Rsl1	0.72	0.40	1.05	0.25	0.56	0.24	0.43	0.84	2.06	1.22	Decreased	zinc finger protein 429	protein coding gene	13:67173206-67183126	1.22
418Tspan33	0.08	0.30	0.35	0.55	3.67	1.58	0.43	1.88	1.66	1.22	Decreased	tetraspanin 33	protein coding gene	6:29694221-29718559	1.22
419Sh2d1b1	0.76	0.10	0.24	0.08	0.14	0.32	2.32	0.86	1.65	1.21	increased	SH2 domain containing 1B1	protein coding gene	1:170277375-170285662	1.21
420Mpped2	0.30	0.05	0.25	0.10	0.17	0.39	2.32	0.56	1.34	1.21	increased	metallophosphoesterase domain containing 2	protein coding gene	2:106693268-106868356	1.21
421Vgf	0.82	2.18	0.97	1.12	2.67	1.15	0.43	1.42	0.20	1.21	Decreased	VEGF nerve growth factor inducible	protein coding gene	5:137030294-137033351	1.21
422Gm4419	0.12	0.53	0.11	0.19	4.29	1.85	0.43	1.10	0.89	1.21	Decreased	predicted gene 4419	antisense lncRNA gene	12:21417910-21419803	1.21
423Spon2	8.94	1.90	10.04	4.93	0.21	0.49	2.31	2.24	1.03	1.21	increased	spondin 2, extracellular matrix protein	protein coding gene	5:33213517-33218238	1.21
424Lysmd4	1.22	0.77	0.72	0.20	0.63	0.27	0.43	0.67	1.88	1.21	Decreased	LysM, putative peptidoglycan-binding, domain containing 4	protein coding gene	7:67222543-67228468	1.21
425Gm8982	0.21	1.13	0.58	1.32	5.30	2.29	0.43	0.41	0.20	1.21	Decreased	predicted gene 8982	pseudogene	7:106123173-106124730	1.21
426Rap2b	7.36	7.61	6.03	2.70	1.03	0.45	0.43	1.05	1.16	1.21	Decreased	RAP2B, member of RAS oncogene family	protein coding gene	3:61362249-61368703	1.21
427F8a	0.55	0.48	1.46	0.55	0.87	0.38	0.43	0.20	1.41	1.21	Decreased	factor 8-associated gene A	protein coding gene	4:73107634-73292976	1.21
428Dhx16	28.23	13.07	34.39	36.73	0.46	1.07	2.31	1.11	1.10	1.21	increased	DEAH (Asp-Glu-Ala-His) box polypeptide 16	protein coding gene	17:35879818-35892670	1.21
429Pcdhb22	1.11	0.15	0.61	0.19	0.13	0.30	2.31	0.52	1.72	1.21	increased	protocadherin beta 22	protein coding gene	18:37518352-37521419	1.21
430Cd74	0.59	1.07	0.74	3.04	1.79	4.13	2.30	0.84	0.05	1.20	increased	CD74 antigen (invariant polypeptide of major histocompatibility	protein coding gene	18:60803847-60812646	1.20
431Amotl2	42.40	11.65	50.86	6.07	0.27	0.12	0.43	1.86	0.07	1.20	Decreased	angiotensin-like 2	protein coding gene	9:102716671-102733418	1.20
432Rtkn	21.32	3.04	12.91	4.24	0.14	0.33	2.30	0.81	1.61	1.20	increased	rhotein	protein coding gene	6:83135462-83156397	1.20
433Bet1l	11.93	4.28	7.52	1.17	0.36	0.16	0.43	1.48	0.68	1.20	Decreased	Bet1 golgi vesicular membrane trafficking protein like	protein coding gene	7:140853383-140856383	1.20
434Mfap2	12.09	3.70	1.72	1.21	0.31	0.70	2.30	1.71	0.51	1.20	increased	microfilament-associated protein 2	protein coding gene	4:141010417-141015984	1.20
435Mx1	0.06	0.14	0.03	0.15	2.13	4.89	2.30	1.09	0.29	1.20	increased	MX dynamin-like GTPase 1	protein coding gene	16:97447034-97462907	1.20
436Tmem151a	1.60	3.44	19.96	18.72	2.15	0.94	0.44	1.11	0.09	1.20	Decreased	transmembrane protein 151A	protein coding gene	19:5070858-5085531	1.20
437Zbtb43	2.34	0.51	3.01	0.29	0.22	0.10	0.44	2.20	0.39	1.20	Decreased	zinc finger and BTB domain containing 43	protein coding gene	2:33450286-33468559	1.20
438C1qtnf6	11.43	1.52	11.13	3.39	0.13	0.30	2.29	0.91	0.71	1.20	increased	C1q and tumor necrosis factor related protein 6	protein coding gene	15:78523345-78529625	1.20
439Csrnp1	4.88	4.29	3.50	1.34	0.88	0.38	0.44	0.19	1.38	1.19	Decreased	cysteine-serine-rich nuclear protein 1	protein coding gene	9:119894894-120068283	1.19
440Gm2308	0.24	0.26	0.28	0.67	1.05	2.41	2.29	1.08	0.27	1.19	increased	predicted gene 2308	pseudogene	7:42116470-42195776	1.19
441Gm13443	1.01	1.09	0.92	2.24	1.07	2.45	2.28	0.10	0.29	1.19	increased	predicted gene 13443	pseudogene	2:35340360-35341316	1.19
442Rnu3a	0.19	27.41	0.90	56.75	144.44	63.29	0.44	1.11	1.11	1.19	Decreased	U3A small nuclear RNA	snRNA gene	10:40258466-40258680	1.19
443Pdgra	10.53	1.15	6.38	1.58	0.11	0.25	2.28	0.20	0.01	1.19	increased	platelet derived growth factor, alpha	protein coding gene	5:138976013-139000576	1.19
444Spry1	4.45	1.08	6.13	0.66	0.24	0.11	0.44	2.04	0.23	1.18	Decreased	sprouty RTK signaling antagonist 1	protein coding gene	3:37639946-37644598	1.18
445C030034I22Rik	3.38	5.22	0.46	2.76	13.71	6.03	0.44	1.78	0.59	1.18	Decreased	RIKEN cDNA C030034I22 gene	lincRNA gene	17:69416659-69439304	1.18
446Scd3	0.25	0.08	0.14	0.10	0.31	0.70	2.27	1.70	0.52	1.18	increased	stearyl-coenzyme A desaturase 3	protein coding gene	19:44203287-44244016	1.18
447Ston2	0.50	1.70	0.62	0.92	3.41	1.50	0.44	1.77	0.59	1.18	Decreased	stonin 2	protein coding gene	12:91638541-91788387	1.18
448Tef	14.00	2.35	22.21	8.47	0.17	0.38	2.27	0.57	1.39	1.18	increased	thyrotroph embryonic factor	protein coding gene	15:81800790-81826861	1.18
449Nlr5	0.76	0.32	0.23	0.22	0.43	0.96	2.27	1.23	0.05	1.18	increased	NLR family, CARD domain containing 5	protein coding gene	8:94472762-94527272	1.18
450Gm17388	0.23	0.11	0.36	0.07	0.46	0.20	0.44	1.12	0.30	1.18	Decreased	predicted gene, 17388	lincRNA gene	4:136355773-136359440	1.18
451Adgrg2	0.29	0.10	0.22	1.78	0.35	0.80	2.27	1.49	0.31	1.18	increased	adhesion G protein-coupled receptor G2	protein coding gene	X:160390689-160499870	1.18
452Ccl8	2.29	7.74	1.60	12.24	3.38	7.66	2.26	1.76	0.94	1.18	increased	chemokine (C-C motif) ligand 8	protein coding gene	11:82115184-82116799	1.18
453Gm21769	0.07	0.64	0.06	0.25	9.34	4.12	0.44	0.22	0.04	1.18	Decreased	predicted gene, 21769	pseudogene	8:20122364-20126624	1.18
454Fpgt	0.98	2.20	1.31	1.31	2.26	1.00	0.44	1.18	0.00	1.18	Decreased	fucose-1-phosphate guanylyltransferase	protein coding gene	3:155084920-155093378	1.18
455Gm11934	0.05	2.99	0.08	1.83	54.46	24.07	0.44	0.77	0.59	1.18	Decreased	predicted gene 11934	unclassified non-coding RNA gene	4:33303568-33310137	1.18
456Gm42743	0.33	6.69	0.37	3.30	20.11	8.89	0.44	1.33	0.15	1.18	Decreased	predicted gene 42743	lincRNA gene	3:96267074-96279001	1.18
457Dbndd2	3.12	21.67	36.90	####	6.94	3.07	0.44	0.80	0.62	1.18	Decreased	dysbindin (dystrobrein binding protein 1) domain containing 2	protein coding gene	2:164456963-164493319	1.18
458Gm12583	0.10	0.99	0.19	0.83	9.88	4.37	0.44	1.31	0.13	1.18	Decreased	predicted gene 12583	pseudogene	3:37063526-37125959	1.18

459	6030445D17Rik	0.22	0.74	0.18	0.28	3.41	1.51	0.44	1.77	0.59	1.18	Decreased	RIKEN cDNA 6030445D17 gene	unclassified gene	4:136462249-136463977	1.18
460	Zfp738	0.71	0.26	1.27	1.04	0.36	0.82	2.26	1.46	0.28	1.17	Increased	zinc finger protein 738	protein coding gene	13:67667436-67687071	1.17
461	Scn2a	0.25	0.13	0.28	0.06	0.52	0.23	0.44	0.96	2.13	1.17	Decreased	sodium channel, voltage-gated, type II, alpha	protein coding gene	2:65620856-65767447	1.17
462	Lypd6	0.34	0.22	1.81	0.52	0.65	0.29	0.44	0.61	1.79	1.17	Decreased	LYS/PLAUR domain containing 6	protein coding gene	2:50066428-50193569	1.17
463	Cyb5rl	1.84	0.88	2.38	2.55	0.48	1.07	2.25	1.07	0.10	1.17	Increased	cytochrome b5 reductase-like	protein coding gene	4:107066987-107088268	1.17
464	Acap1	0.09	0.44	0.08	0.92	5.03	11.33	2.25	2.33	0.50	1.17	Increased	ArfGAP with coiled-coil, ankyrin repeat and PH domains 1	protein coding gene	11:69881566-69895539	1.17
465	Zfp11	0.69	0.24	0.62	0.09	0.34	0.15	0.44	1.55	2.73	1.17	Decreased	zinc finger protein 11	protein coding gene	5:129654592-129670099	1.17
466	Serpib9b	0.45	1.65	8.86	14.55	3.69	1.64	0.44	1.88	1.72	1.17	Decreased	serine (or cysteine) peptidase inhibitor, clade B, member 9b	protein coding gene	13:33027415-33041884	1.17
467	Krtap1-5	2.69	1.00	1.08	0.90	0.37	0.84	2.25	1.42	0.26	1.17	Increased	keratin associated protein 1-5	protein coding gene	11:99579976-99581016	1.17
468	Bves	0.16	0.60	1.32	2.17	3.70	1.65	0.45	1.89	0.72	1.17	Decreased	blood vessel epicardial substance	protein coding gene	10:45335761-45369708	1.17
469	Tmem115	21.22	12.39	15.93	4.14	0.58	0.26	0.45	0.78	1.94	1.17	Decreased	transmembrane protein 115	protein coding gene	9:107533944-107538656	1.17
470	Gm20109	0.23	0.93	0.09	0.78	4.06	9.12	2.24	2.02	0.19	1.17	Increased	predicted gene, 20109	lincRNA gene	17:28522380-28523148	1.17
471	Chst10	0.33	0.02	0.91	0.15	0.07	0.17	2.24	1.74	2.57	1.17	Increased	carbohydrate sulfotransferase 10	protein coding gene	1:38863876-38898160	1.17
472	Rcan1	7.13	20.28	4.44	5.63	2.84	1.27	0.45	1.51	0.34	1.17	Decreased	regulator of calcineurin 1	protein coding gene	16:92391952-92466146	1.17
473	Syt11	0.08	0.44	0.07	0.16	5.16	2.32	0.45	1.37	0.21	1.15	Decreased	synaptotagmin-like 1	protein coding gene	4:133253089-133263113	1.15
474	Gm3809	0.19	0.30	0.19	0.65	1.58	3.51	2.22	1.66	0.81	1.15	Increased	predicted gene 3809	pseudogene	1:186960507-186961513	1.15
475	Mocs3	3.78	1.88	5.47	1.23	0.50	0.22	0.45	1.01	2.16	1.15	Decreased	molybdenum cofactor synthesis 3	protein coding gene	2:168230621-168232302	1.15
476	Trim12c	0.18	0.17	0.35	0.72	0.93	2.07	2.22	0.10	0.05	1.15	Increased	tripartite motif-containing 12C	protein coding gene	7:104244456-104353362	1.15
477	Tlr3	0.48	0.18	0.56	0.46	0.37	0.82	2.21	1.43	0.29	1.14	Increased	toll-like receptor 3	protein coding gene	8:45395664-45410608	1.14
478	Gm11836	5.12	8.12	5.13	17.97	1.59	3.50	2.21	0.67	0.81	1.14	Increased	predicted gene 11836	pseudogene	4:14433890-14434337	1.14
479	Ubxn11	0.67	1.35	0.11	0.47	2.01	4.43	2.21	1.01	0.15	1.14	Increased	UBX domain protein 11	protein coding gene	4:134102569-134128789	1.14
480	Mamdc2	2.24	0.59	0.29	0.17	0.26	0.58	2.20	1.92	0.78	1.14	Increased	MAM domain containing 2	protein coding gene	19:23302608-23448322	1.14
481	Kcnn4	2.00	0.37	4.84	1.96	0.18	0.40	2.20	2.45	1.31	1.14	Increased	potassium intermediate/small conductance calcium-activated ch protein coding gene	protein coding gene	7:24370262-24385209	1.14
482	Gm13415	0.72	0.25	0.37	0.06	0.35	0.16	0.45	1.50	0.64	1.14	Decreased	predicted gene 13415	unclassified non-coding RNA gene	2:24420550-24486855	1.14
483	Zfp128	0.58	0.09	0.39	0.13	0.15	0.32	2.20	2.78	1.64	1.14	Increased	zinc finger protein 128	protein coding gene	7:12881176-12893422	1.14
484	Col12a1	36.45	####	22.92	30.41	2.92	1.33	0.45	1.54	0.41	1.14	Decreased	collagen, type XII, alpha 1	protein coding gene	7:79598990-79718831	1.14
485	Tmt10c	6.73	14.24	6.62	6.38	2.12	0.96	0.45	1.08	0.06	1.14	Decreased	tRNA methyltransferase 10C	protein coding gene	16:56032608-56037819	1.14
486	Cbx8	5.45	1.52	4.48	0.57	0.28	0.13	0.46	1.84	0.98	1.14	Decreased	chromobox 8	protein coding gene	11:119036304-119040969	1.14
487	Gm6740	0.63	0.20	0.46	0.07	0.32	0.15	0.46	1.63	0.76	1.13	Decreased	predicted gene 6740	pseudogene	14:52861766-52862630	1.13
488	Slc11a1	0.21	0.83	0.05	0.41	3.91	8.55	2.19	1.97	0.10	1.13	Increased	solute carrier family 11 (proton-coupled divalent metal ion trans	protein coding gene	1:74375202-74386062	1.13
489	Nkain4	0.03	0.17	5.14	12.50	5.32	2.43	0.46	2.41	0.28	1.13	Decreased	Na+/K+ transporting ATPase interacting 4	protein coding gene	2:180934771-180966812	1.13
490	Fam120aos	0.94	5.80	1.17	3.29	6.15	2.81	0.46	1.62	1.49	1.13	Decreased	family with sequence similarity 120A, opposite strand	lincRNA gene	13:48968286-48972031	1.13
491	Kctd21	0.93	0.05	0.52	0.06	0.06	0.12	2.19	0.17	0.04	1.13	Increased	potassium channel tetramerisation domain containing 21	protein coding gene	7:97332326-97350213	1.13
492	Gm11914	0.50	0.60	0.66	1.73	1.20	2.62	2.18	0.26	0.39	1.13	Increased	predicted gene 11914	pseudogene	4:27867148-27867655	1.13
493	Rassf2	1.59	0.18	0.59	0.14	0.11	0.24	2.18	3.16	2.04	1.13	Increased	Ras association (RalGDS/AF-6) domain family member 2	protein coding gene	2:131989414-132030258	1.13
494	Pcdhb5	0.52	0.13	0.14	0.07	0.24	0.52	2.18	2.06	0.94	1.13	Increased	protocadherin beta 5	protein coding gene	18:37320380-37323913	1.13
495	Eid2	8.05	2.27	0.65	0.08	0.28	0.13	0.46	1.83	0.95	1.12	Decreased	EP300 interacting inhibitor of differentiation 2	protein coding gene	7:28267880-28269165	1.12
496	Kcng1	0.16	0.03	0.20	0.08	0.18	0.39	2.18	2.49	1.37	1.12	Increased	potassium voltage-gated channel, subfamily G, member 1	protein coding gene	2:168260116-168281736	1.12
497	Gm12059	0.12	0.21	0.09	0.35	1.70	3.70	2.18	0.77	0.89	1.12	Increased	predicted gene 12059	unclassified non-coding RNA gene	11:23067997-23075574	1.12
498	Oas1a	0.74	6.47	4.52	18.14	8.74	4.01	0.46	1.13	0.00	1.12	Decreased	2'-5' oligoadenylate synthetase 1G	protein coding gene	5:120896255-120907521	1.12
499	Mtmr14	13.90	2.75	15.57	6.71	0.20	0.43	2.18	2.34	1.21	1.12	Increased	myotubularin related protein 14	protein coding gene	6:113237842-113281387	1.12
500	Vsir	0.19	0.03	0.47	0.17	0.17	0.36	2.18	2.58	1.46	1.12	Increased	V-set immunoregulatory receptor	protein coding gene	10:60302747-60696490	1.12
501	Hic2	0.34	0.08	0.34	0.04	0.23	0.11	0.46	2.09	0.21	1.12	Decreased	hypermethylated in cancer 2	protein coding gene	16:17233571-17263430	1.12
502	Adams2	6.65	16.60	2.17	2.49	2.49	1.15	0.46	1.32	0.20	1.12	Decreased	a disintegrin-like and metallopeptidase (reprolysin type) with thr	protein coding gene	11:50599297-50807573	1.12
503	Zfp747	2.31	1.41	1.77	0.50	0.61	0.28	0.46	0.72	1.84	1.12	Decreased	zinc finger protein 747	protein coding gene	7:127372563-127376050	1.12
504	Gm9522	0.25	0.26	0.29	0.68	1.07	2.32	2.17	0.09	0.21	1.12	Increased	predicted gene 9522	pseudogene	15:39180510-39181441	1.12
505	Fhl1	12.11	64.86	59.53	####	5.36	2.47	0.46	2.42	0.30	1.12	Decreased	complement component factor h	protein coding gene	X:56731786-56793346	1.12
506	Abli1	1.88	0.78	7.44	6.68	0.41	0.90	2.17	1.27	0.16	1.12	Increased	actin-binding LIM protein 1	protein coding gene	19:57032732-57134919	1.12
507	Gpr137c	0.07	1.43	0.05	0.45	20.99	9.69	0.46	0.39	0.28	1.12	Decreased	G protein-coupled receptor 137C	protein coding gene	14:45134447-45281230	1.12
508	Abi3bp	0.62	1.29	0.74	0.71	2.09	0.96	0.46	1.06	0.05	1.12	Decreased	ABI gene family, member 3 (NESH) binding protein	protein coding gene	16:56477845-56690128	1.12
509	Smug1	3.50	1.37	3.95	0.71	0.39	0.18	0.46	1.35	2.47	1.12	Decreased	single-strand selective monofunctional uracil DNA glycosylase	protein coding gene	15:103153306-103158002	1.12
510	Cd81	90.45	23.50	####	58.25	0.26	0.56	2.16	1.94	0.83	1.11	Increased	CD81 antigen	protein coding gene	7:143021783-143067934	1.11
511	Gm3170	0.11	0.56	0.34	0.80	5.02	2.32	0.46	2.33	0.21	1.11	Decreased	predicted gene 3170	pseudogene	14:5607619-5807982	1.11
512	Gm3348	0.11	0.56	0.34	0.80	5.02	2.32	0.46	2.33	0.21	1.11	Decreased	predicted gene 3348	pseudogene	14:5365793-5389049	1.11
513	Trim47	28.22	9.37	21.56	15.47	0.33	0.72	2.16	1.59	0.48	1.11	Increased	tripartite motif-containing 47	protein coding gene	11:116105751-116131128	1.11
514	Scml2	0.70	0.24	0.15	0.11	0.35	0.75	2.16	1.52	0.41	1.11	Increased	Scm polcomb group protein like 2	protein coding gene	X:161117192-161258213	1.11
515	Tec	0.33	0.04	0.96	0.23	0.11	0.24	2.16	0.15	0.04	1.11	Increased	tec protein tyrosine kinase	protein coding gene	5:72755717-72868483	1.11
516	Atp5k-ps2	2.57	3.45	3.16	9.18	1.34	2.90	2.16	0.43	0.54	1.11	Increased	ATP synthase, H+ transporting, mitochondrial F1F0 complex, su	pseudogene	2:173984481-173984691	1.11

517Oas1b	0.39	2.00	1.32	3.17	5.17	2.40	0.46	2.37	1.26	1.11	Decreased	Z'-5' oligoadenylate synthetase 1B	polymorphic pseudogene	5:120812634-120824163	1.11
518Chchd2-ps	0.69	0.85	0.87	2.32	1.23	2.65	2.15	1.30	1.41	1.11	Increased	coiled-coil-helix-coiled-coil-helix domain containing 2, pseudogene	pseudogene	4:148068030-148068492	1.11
519Lrrc8e	3.00	0.95	1.77	0.26	0.32	0.15	0.47	1.66	2.76	1.10	Decreased	leucine rich repeat containing 8 family, member E	protein coding gene	8:4226826-4237470	1.10
520Gm11224	0.16	0.91	0.20	0.54	5.79	2.70	0.47	2.53	1.43	1.10	Decreased	predicted gene 11224	pseudogene	4:69954165-69955276	1.10
521Tcea3	0.98	0.96	0.40	0.84	0.98	2.10	2.14	0.03	1.07	1.10	Increased	transcription elongation factor A (SII), 3	protein coding gene	4:136247728-136274898	1.10
522Rora	0.43	1.85	0.07	0.14	4.33	2.02	0.47	2.11	1.01	1.10	Decreased	RAR-related orphan receptor alpha	protein coding gene	9:68653785-69388246	1.10
523Sorcs2	0.80	2.86	0.43	0.71	3.56	1.66	0.47	1.83	0.73	1.10	Decreased	sortilin-related VPS10 domain containing receptor 2	protein coding gene	5:36017179-36398139	1.10
524Etv3	2.71	1.41	2.35	0.57	0.52	0.24	0.47	0.94	2.04	1.10	Decreased	ets variant 3	protein coding gene	3:87525406-87540156	1.10
525Gramd2	1.23	0.35	0.41	0.25	0.29	0.61	2.14	1.81	0.71	1.10	Increased	GRAM domain containing 2	protein coding gene	9:59680143-59718874	1.10
526Polr2j	####	74.87	####	####	0.41	0.89	2.14	1.27	0.18	1.10	Increased	polymerase (RNA) II (DNA directed) polypeptide J	protein coding gene	5:136116630-136136074	1.10
527Yod1	0.43	1.68	0.37	0.67	3.89	1.82	0.47	1.96	1.86	1.10	Decreased	YOD1 deubiquitinase	protein coding gene	1:130689042-130729253	1.10
528Zxdb	0.82	0.35	0.80	0.16	0.43	0.20	0.47	1.22	2.32	1.09	Decreased	zinc finger, X-linked, duplicated B	protein coding gene	X:94724568-94730191	1.09
529Peg12	5.00	0.66	0.36	0.10	0.13	0.28	2.13	2.93	1.83	1.09	Increased	frequently rearranged in advanced T cell lymphomas	protein coding gene	7:62461870-62464510	1.09
530Gm8864	0.29	0.35	0.28	0.72	1.23	2.62	2.13	1.30	1.39	1.09	Increased	predicted gene 8864	pseudogene	X:88572050-88573060	1.09
531Mrps36-ps1	0.49	1.02	0.39	1.75	2.08	4.43	2.13	1.05	1.15	1.09	Increased	mitochondrial ribosomal protein S36, pseudogene 1	pseudogene	17:53866416-53867041	1.09
532Prkg2	14.00	6.78	7.82	8.07	0.48	1.03	2.13	1.05	1.05	1.09	Increased	protein kinase, cGMP-dependent, type II	protein coding gene	5:98929772-99037351	1.09
533Fam129a	2.74	5.61	5.64	5.42	2.05	0.96	0.47	1.03	0.06	1.09	Decreased	family with sequence similarity 129, member A	protein coding gene	1:151571185-151721939	1.09
534Efna1	10.10	1.94	8.62	0.78	0.19	0.09	0.47	2.38	1.47	1.09	Decreased	ephrin B1	protein coding gene	X:99136129-99148991	1.09
535Fam219a	4.05	1.23	8.36	1.19	0.30	0.14	0.47	1.72	2.81	1.09	Decreased	family with sequence similarity 219, member A	protein coding gene	4:41517436-41569538	1.09
536Apolld1	0.35	0.02	0.18	0.03	0.07	0.14	2.12	1.88	5.79	1.09	Increased	apolipoprotein L domain containing 1	protein coding gene	6:134982000-134986836	1.09
537Mkrn3	0.30	0.16	0.13	0.03	0.53	0.25	0.47	0.92	2.00	1.09	Decreased	makorin, ring finger protein, 3	protein coding gene	7:62417592-62420139	1.09
538Pri7a1	0.09	0.37	0.45	0.90	4.26	2.01	0.47	2.09	1.00	1.09	Decreased	prolactin family 7, subfamily a, member 1	protein coding gene	13:27633379-27642474	1.09
539Slc13a4	0.02	1.21	0.02	0.45	55.92	26.36	0.47	1.1	1.72	1.08	Decreased	solute carrier family 13 (sodium/sulfate symporters), member 4	protein coding gene	6:35267956-35308131	1.08
540Prr5	11.48	3.14	10.25	1.32	0.27	0.13	0.47	1.87	2.96	1.08	Decreased	proline rich 5 (renal)	protein coding gene	15:84669619-84703673	1.08
541Trappc13	1.94	5.88	2.88	4.12	3.03	1.43	0.47	1.60	0.52	1.08	Decreased	trafficking protein particle complex 13	protein coding gene	13:104142148-104178469	1.08
542Ltpb2	40.79	####	7.27	9.32	2.71	1.28	0.47	1.44	1.36	1.08	Decreased	latent transforming growth factor beta binding protein 2	protein coding gene	12:84783211-84878132	1.08
543Pla1a	0.25	0.25	0.21	0.43	0.98	2.08	2.11	0.02	1.06	1.08	Increased	lysophospholipase 1	protein coding gene	16:38396118-38433145	1.08
544Pfkfb	19.14	9.11	16.38	16.47	0.48	1.01	2.11	1.07	1.01	1.08	Increased	phosphofructokinase, muscle	protein coding gene	15:98038743-98132447	1.08
545A930029G22Ri	0.19	1.77	0.18	0.82	9.37	4.44	0.47	1.23	1.15	1.08	Decreased	RIKEN cDNA A930029G22 gene	lincRNA gene	17:69416659-69439304	1.08
546Gm7927	0.11	0.73	0.29	0.93	6.78	3.21	0.47	2.76	1.68	1.08	Decreased	predicted gene 7927	pseudogene	X:55845758-55846638	1.08
547Gm14111	0.47	0.68	0.44	1.32	1.43	3.02	2.11	1.52	1.59	1.08	Increased	predicted gene 14111	pseudogene	2:146754706-146755705	1.08
548Slc4a1ap	12.37	5.17	13.97	12.32	0.42	0.88	2.11	1.26	0.18	1.08	Increased	solute carrier family 4 (anion exchanger), member 1, adaptor pr	protein coding gene	5:31514568-31554042	1.08
549Myc	85.67	49.86	88.34	24.38	0.58	0.28	0.47	0.78	1.86	1.08	Decreased	myelocytomatosis oncogene	protein coding gene	15:61985390-61990374	1.08
550S1pr1	2.65	1.15	0.70	0.15	0.45	0.21	0.47	1.14	2.22	1.08	Decreased	sphingosine-1-phosphate receptor 1	protein coding gene	3:115710432-115715055	1.08
551A430110L20Rik	0.18	0.08	0.19	0.04	0.46	0.22	0.47	1.12	2.20	1.08	Decreased	RIKEN cDNA A430110L20 gene	lincRNA gene	1:181225985-181228498	1.08
552Gm15892	0.37	3.18	1.14	4.61	8.52	4.05	0.47	1.09	1.02	1.08	Decreased	predicted gene 15892	lincRNA gene	11:88059065-88067476	1.08
553Tob2	16.39	5.82	10.81	1.82	0.36	0.17	0.48	1.49	2.57	1.07	Decreased	transducer of ERBB2, 2	protein coding gene	15:81848269-81858796	1.07
554Smg8	11.66	6.22	8.88	2.25	0.53	0.25	0.48	0.91	1.98	1.07	Decreased	smg-8 homolog, nonsense mediated mRNA decay factor (C. ele	protein coding gene	11:87077731-87086774	1.07
555Lrfn3	0.76	0.05	0.32	0.04	0.06	0.13	2.10	1.02	1.95	1.07	Increased	leucine rich repeat and fibronectin type III domain containing 3	protein coding gene	7:30355513-30362772	1.07
556Col3a1	98.12	44.26	76.92	72.95	0.45	0.95	2.10	1.15	0.08	1.07	Increased	collagen, type III, alpha 1	protein coding gene	1:45311537-45349706	1.07
557Hck	0.99	2.31	0.27	0.30	2.34	1.11	0.48	1.23	1.15	1.07	Decreased	hemopoietic cell kinase	protein coding gene	2:153108467-153151441	1.07
558Mast4	5.37	2.41	4.73	4.46	0.45	0.94	2.10	1.15	0.08	1.07	Increased	microtubule associated serine/threonine kinase family member	protein coding gene	13:102693557-103334497	1.07
559Spib	0.01	0.15	0.02	0.12	12.90	6.15	0.48	1.89	1.62	1.07	Decreased	Spi-B transcription factor (Spi-1/Pu.1 related)	protein coding gene	7:44525994-44532071	1.07
560Gm8464	2.61	0.44	2.33	0.83	0.17	0.36	2.10	2.56	1.49	1.07	Increased	predicted pseudogene 8464	pseudogene	X:152489660-152491860	1.07
561Hist1h2bl	0.16	0.33	0.36	1.56	2.08	4.36	2.10	1.06	1.13	1.07	Increased	histone cluster 1, H2b2	protein coding gene	13:21715762-21716143	1.07
562Hacd3	19.91	7.10	27.87	20.84	0.36	0.75	2.10	1.49	0.42	1.07	Increased	3-hydroxyacyl-CoA dehydratase 3	protein coding gene	9:64986984-65021717	1.07
563Haghl	10.31	3.78	10.65	8.18	0.37	0.77	2.09	1.45	0.38	1.07	Increased	hydroxyacylglutathione hydrolase-like	protein coding gene	17:25773775-25785673	1.07
564Tubb4b-ps1	1.29	0.51	2.53	2.11	0.40	0.83	2.09	1.33	0.26	1.06	Increased	tubulin, beta 4B class IVB, pseudogene 1	pseudogene	5:7179364-7180699	1.06
565Nr2f1	4.48	0.53	0.65	0.16	0.12	0.25	2.09	3.08	2.01	1.06	Increased	nuclear receptor subfamily 2, group F, member 1	protein coding gene	13:78188972-78279391	1.06
566Nap1l2	0.20	0.17	2.16	0.88	0.86	0.41	0.48	0.22	1.29	1.06	Decreased	nucleosome assembly protein 1-like 2	protein coding gene	X:103182970-103311782	1.06
567Gm12466	0.12	0.79	0.27	0.86	6.75	3.23	0.48	1.75	1.69	1.06	Decreased	predicted gene 12466	pseudogene	4:52240902-52241522	1.06
568Cdk5r1	0.84	0.45	0.79	0.20	0.54	0.26	0.48	0.90	1.96	1.06	Decreased	cyclin-dependent kinase 5, regulatory subunit 1 (p35)	protein coding gene	11:80477045-80481179	1.06
569Myadm	94.08	17.11	60.91	23.10	0.18	0.38	2.09	2.46	1.40	1.06	Increased	myeloid-associated differentiation marker	protein coding gene	7:3289037-3299345	1.06
570Hist1h3e	0.27	1.97	0.58	2.00	7.16	3.44	0.48	1.84	1.78	1.06	Decreased	histone cluster 1, H3e	protein coding gene	13:23561533-23562369	1.06
571Glce	5.25	1.26	3.18	0.37	0.24	0.12	0.48	2.05	1.11	1.06	Decreased	glucuronyl C5-epimerase	protein coding gene	9:62052472-62070606	1.06
572Pcdhb16	0.40	0.08	0.41	0.14	0.16	0.33	2.07	2.64	1.58	1.05	Increased	protocadherin beta 16	protein coding gene	18:37477767-37483035	1.05
573Socs1	8.85	1.39	3.46	1.12	0.16	0.32	2.07	2.67	1.62	1.05	Increased	suppressor of cytokine signaling 1	protein coding gene	16:10783807-10785536	1.05
574Ccnyl1	8.68	5.92	6.10	2.01	0.68	0.33	0.48	0.55	1.60	1.05	Decreased	cyclin Y-like 1	protein coding gene	1:64679870-64725644	1.05

575Eps15	7.64	3.08	10.29	8.60	0.40	0.84	2.07	1.31	0.26	1.05	increased	epidermal growth factor receptor pathway substrate 15	protein coding gene	4:109280267-109387817	1.05
576Shroom3	0.57	0.11	7.04	2.76	0.19	0.39	2.07	2.40	1.35	1.05	increased	shroom family member 3	protein coding gene	5:92683434-92965318	1.05
577Zfp983	0.68	0.36	0.31	0.08	0.53	0.26	0.48	0.90	1.95	1.05	Decreased	zinc finger protein 983	protein coding gene	17:21657581-21662540	1.05
578Zbtb3	1.18	0.14	1.06	0.26	0.12	0.24	2.07	3.08	2.03	1.05	increased	zinc finger and BTB domain containing 3	protein coding gene	19:8802529-8804854	1.05
579Mycbp	7.74	34.75	8.72	18.93	4.49	2.17	0.48	2.17	1.12	1.05	Decreased	MYC binding protein	protein coding gene	4:123898632-123917269	1.05
580Tigd5	1.52	0.41	1.72	0.23	0.27	0.13	0.48	1.89	2.93	1.05	Decreased	tigger transposable element derived 5	protein coding gene	15:75909759-75912148	1.05
581Lce1h	0.74	6.30	0.82	3.38	8.56	4.14	0.48	1.10	2.05	1.05	Decreased	late comifed envelope 1H	protein coding gene	3:92763214-92765065	1.05
582Tprn	7.83	1.80	12.18	5.79	0.23	0.48	2.07	2.12	1.07	1.05	increased	taperin	protein coding gene	2:25262617-25269885	1.05
583Slc39a1	50.76	17.80	36.20	26.18	0.35	0.72	2.06	1.51	0.47	1.04	increased	solute carrier family 40 (iron-regulated transporter), member 1	protein coding gene	3:90248171-90253612	1.04
584Rasf6	0.36	1.48	0.35	0.70	4.07	1.97	0.49	1.02	1.98	1.04	Decreased	Ras association (RalGDS/AF-6) domain family member 6	protein coding gene	5:90603075-90640487	1.04
585Sowahc	1.61	1.30	1.03	0.41	0.81	0.39	0.49	0.31	1.35	1.04	Decreased	sosondowah ankryrin repeat domain family member C	protein coding gene	10:59221952-59226434	1.04
586Loxl2	3.72	7.80	7.57	7.71	2.10	1.02	0.49	1.07	1.02	1.04	Decreased	propionyl Coenzyme A carboxylase, beta polypeptide	protein coding gene	14:69555423-69693698	1.04
587Bicdl1	3.21	0.37	2.15	0.12	0.12	0.06	0.49	3.11	3.14	1.04	Decreased	BICD family like cargo adaptor 1	protein coding gene	5:115648174-115731621	1.04
5889330104G04Rik	0.30	0.10	0.20	0.13	0.33	0.67	2.05	1.60	0.57	1.03	increased	RIKEN cDNA 9330104G04 gene	lincRNA gene	7:16871353-16874886	1.03
589Mycl	0.28	0.11	1.56	0.29	0.38	0.19	0.49	1.39	2.43	1.03	Decreased	v-myc avian myelocytomatosis viral oncogene lung carcinoma d protein coding gene	protein coding gene	4:122995651-123002485	1.03
590Ankrd24	6.32	2.90	5.55	5.21	0.46	0.94	2.04	1.12	0.09	1.03	increased	ankryrin repeat domain 24	protein coding gene	10:81628539-81647610	1.03
591Gm11946	0.08	1.90	0.02	1.19	24.41	49.86	2.04	1.61	0.64	1.03	increased	predicted gene 11946	unclassified non-coding RNA gene	11:3517522-3540612	1.03
592Gm8420	0.30	0.80	0.17	0.95	2.66	5.43	2.04	1.41	2.44	1.03	increased	predicted gene 8420	pseudogene	1:51940649-51941261	1.03
593Ehd2	38.85	9.77	28.68	14.73	0.25	0.51	2.04	1.99	0.96	1.03	increased	EH-domain containing 3	protein coding gene	7:15946976-15967567	1.03
594Zbtb39	2.33	1.53	1.64	0.53	0.66	0.32	0.49	0.60	1.63	1.03	Decreased	zinc finger and BTB domain containing 39	protein coding gene	10:127739537-127751732	1.03
595Ehd1	78.65	26.99	82.63	57.84	0.34	0.70	2.04	1.54	0.51	1.03	increased	EH-domain containing 1	protein coding gene	19:6276724-6300096	1.03
596Gatc	16.71	5.04	16.67	10.26	0.30	0.62	2.04	1.73	0.70	1.03	increased	glutamy-tRNA(Gln) amidotransferase, subunit C	protein coding gene	5:115333241-115341178	1.03
597Slc25a29	0.63	0.31	0.96	0.23	0.49	0.24	0.49	1.02	2.05	1.03	Decreased	solute carrier family 25 (mitochondrial carrier, palmitoylcarnitine protein coding gene	protein coding gene	12:108825872-108835883	1.03
598Fcer1g	0.26	0.78	0.48	0.71	3.01	1.48	0.49	1.59	1.56	1.03	Decreased	Fc receptor, IgE, high affinity I, gamma polypeptide	protein coding gene	1:171229571-171234365	1.03
599Gm5884	0.10	0.25	0.15	0.18	2.52	1.24	0.49	1.33	1.31	1.03	Decreased	predicted pseudogene 5884	pseudogene	6:128644837-128647096	1.03
600Zfp408	3.51	2.19	4.97	1.52	0.62	0.31	0.49	0.68	1.71	1.03	Decreased	zinc finger protein 408	protein coding gene	2:91643668-91649791	1.03
601Lrp2	0.37	0.91	1.40	1.70	2.48	1.22	0.49	1.31	1.28	1.02	Decreased	low density lipoprotein receptor-related protein 2	protein coding gene	2:69424339-69586065	1.02
602Cdk1	0.35	0.08	0.18	0.09	0.24	0.48	2.03	2.08	1.05	1.02	increased	cyclin-dependent kinase-like 1 (CDC2-related kinase)	protein coding gene	12:69746847-69794511	1.02
603Sud1	0.14	0.74	0.50	1.27	5.21	2.56	0.49	1.38	1.36	1.02	Decreased	sushi domain containing 1	protein coding gene	4:59314682-59438633	1.02
604Zfp865	5.19	2.83	3.36	0.90	0.54	0.27	0.49	0.88	1.90	1.02	Decreased	zinc finger protein 865	protein coding gene	7:5020375-5033138	1.02
605Gm4076	2.81	4.56	1.78	5.85	1.62	3.29	2.03	0.70	1.72	1.02	increased	predicted gene 4076	protein coding gene	13:85126998-85127514	1.02
606Asphd2	4.26	0.88	1.47	0.62	0.21	0.42	2.03	2.27	1.25	1.02	increased	aspartate beta-hydroxylase domain containing 2	protein coding gene	5:112385447-112392213	1.02
607Top2a	69.16	24.22	93.78	66.58	0.35	0.71	2.03	1.51	0.49	1.02	increased	topoisomerase (DNA) II alpha	protein coding gene	11:98992942-99024189	1.02
608Cend1	2.09	0.34	0.32	0.10	0.16	0.33	2.03	2.63	1.61	1.02	increased	cell cycle exit and neuronal differentiation 1	protein coding gene	7:141426445-141429491	1.02
609Esam	0.27	0.40	0.15	0.47	1.52	3.08	2.03	0.60	1.62	1.02	increased	endothelial cell-specific adhesion molecule	protein coding gene	9:37528077-37538319	1.02
610Jun	22.72	64.91	18.60	26.26	2.86	1.41	0.49	1.51	1.50	1.02	Decreased	jun proto-oncogene	protein coding gene	4:95049033-95167318	1.02
611Ttc5	7.53	1.20	10.11	3.25	0.16	0.32	2.02	2.65	1.64	1.02	increased	tetratricopeptide repeat domain 5	protein coding gene	14:50765414-50785518	1.02
612Rab11b	67.47	8.09	65.80	15.95	0.12	0.24	2.02	3.06	2.04	1.02	increased	RAB11B, member RAS oncogene family	protein coding gene	17:33742483-33761870	1.02
613mt-Nd6	####	####	####	####	1.03	2.08	2.02	0.04	1.06	1.02	increased	mitochondrially encoded NADH dehydrogenase 6	protein coding gene	MT:11741-14070	1.02
614Nrn1	####	9.66	####	11.01	0.05	0.11	2.02	3.20	3.19	1.02	increased	neurtin 1	protein coding gene	13:36725624-36734477	1.02
615Gm10481	0.40	0.48	0.36	0.88	1.20	2.43	2.02	0.27	1.28	1.01	increased	predicted gene 10481	pseudogene	5:13042432-13043434	1.01
616Aldh1a3	0.18	1.01	0.17	0.48	5.75	2.85	0.50	2.52	1.51	1.01	Decreased	aldehyde dehydrogenase family 1, subfamily A3	protein coding gene	7:66390891-66427517	1.01
617Rhob	38.04	23.13	34.88	10.53	0.61	0.30	0.50	0.72	1.73	1.01	Decreased	ras homolog family member B	protein coding gene	12:8497762-8499985	1.01
618Olfm2	1.43	3.35	0.27	0.31	2.34	1.16	0.50	1.23	1.22	1.01	Decreased	olfactomedin 2	protein coding gene	9:20667985-20728219	1.01
619Nymrin	0.77	0.05	0.37	0.05	0.07	0.14	2.01	1.87	2.86	1.01	increased	NYN domain and retroviral integrase containing	protein coding gene	14:55854035-55874735	1.01
620Cd1d1	2.99	1.41	2.32	2.19	0.47	0.95	2.01	1.09	0.08	1.01	increased	CD1d1 antigen	protein coding gene	3:86995833-86999441	1.01
621Wnt9a	0.18	0.11	0.99	0.29	0.58	0.29	0.50	0.78	1.78	1.01	Decreased	wingless-type MMTV integration site family, member 9A	protein coding gene	11:59306927-59333552	1.01
622Ppan	79.00	19.45	####	12.43	0.25	0.12	0.50	2.02	3.03	1.01	Decreased	peter pan homolog	protein coding gene	9:20888174-20892178	1.01
623Npr3	1.99	1.66	2.35	0.98	0.83	0.42	0.50	0.26	1.27	1.01	Decreased	natriuretic peptide receptor 3	protein coding gene	15:11839895-11905674	1.01
624Meaf6	21.30	7.51	17.31	12.24	0.35	0.71	2.01	1.50	0.50	1.00	increased	MYST/Esa1-associated factor 6	protein coding gene	4:125085133-125113238	1.00
625Rpl17-ps8	0.35	0.60	0.40	1.38	1.73	3.46	2.00	0.79	1.79	1.00	increased	ribosomal protein L17, pseudogene 8	pseudogene	X:96306177-96574485	1.00
626Tst	4.37	1.93	12.50	11.05	0.44	0.88	2.00	1.18	0.18	1.00	increased	thiosulfate sulfurtransferase, mitochondrial	protein coding gene	15:78399555-78405859	1.00