



**Newcastle  
University**

**Combining CRISPR-Cas13 and proximity  
biotinylation approaches for discovery  
of novel androgen receptor variant splicing  
factors in castrate resistant prostate cancer**

**Thesis submitted in partial fulfilment of the requirement of the degree of Doctor of Philosophy**

**Nicholas Brittain**

**190545226**

Translational and Clinical Research Institute

Faculty of Medical Sciences

Newcastle University

June 2023



## Abstract

The relevance of androgens to prostate cancer (PCa) has long been appreciated. The cellular effects of androgens are exerted by their interaction with the ligand-binding domain (LBD) of the androgen receptor (AR). Targeting AR signalling through suppression of circulating testosterone with androgen deprivation therapy (ADT) or direct receptor inhibition using antiandrogens is the clinical mainstay for treatment of metastatic PCa. Whilst these can elicit substantial clinical responses, a significant number of patients will relapse despite castrate levels of serum testosterone.

It is now recognised that this disease state, castrate resistant prostate cancer (CRPC), continually activates AR through a number of mechanisms including alternative mRNA splicing of androgen receptor-variants (AR-Vs). AR-Vs splice cryptic exons (CEs) in place of the LBD, obviating the need for androgen binding and rendering them immune to targeting with antiandrogens. AR-Vs, in particular AR-V7 which splices cryptic exon 3 (CE3), are linked with therapeutic resistance and poor clinical outcomes.

There is currently a lack of knowledge regarding factors that mediate splicing of AR-V7. Here, two contemporary biotechnologies have been employed to identify novel splicing factors. This work has combined RNA-targeting CRISPR protein Cas13 with proximity biotinylation enzyme APEX2 to biotinylate, enrich, and identify local protein interactors of CE3 mRNA in CRPC cell line CWR22Rv1. This has created a rich source of potential AR-V7 splicing regulators, of which TRA2B has shown particular promise. Subsequent validation has confirmed TRA2 proteins as regulators controlling alternative splicing decisions of the *AR* gene, and their clinical relevance is vindicated by analysis of patient datasets.

Therefore, development of these methodologies represents a powerful proof of concept that could in theory be applied to any mRNA region of interest.

## Acknowledgements

I would firstly like to thank my primary supervisor, Dr Luke Gaughan, for the opportunity to come to Newcastle University and for his advice and support throughout what has been a highly innovative and engaging PhD project. This project has been a real challenge, in the best possible way, and was an ambitious, novel and truly creative idea. I would also like to thank him for providing me with the freedom to build my skillset in multiple other areas outside of the lab, which will prove invaluable in the next stages of my development as a scientist. Lastly, he has been highly supportive of my career decisions, which I have appreciated enormously.

I would also like to thank the entire Newcastle University Solid Tumour Target Discovery group for their support and friendship throughout. Particularly Ryan Nelson, with whom I have shared and received many ideas and pieces of advice and is the only other man I know who can quote David Brent word for word; Beth Adamson and Laura Walker, for their support and making 2.017 bay 1 a great place to work over these last few years; and Laura Wilson, who is truly vital to the lab and is always on hand to compare notes as a fellow new parent.

I would also like to thank my parents, Liz and Chris, who have always supported me so much in my life and career, which has given me a platform to pursue what I am truly passionate about.

Additionally, this project would not have been possible without the funding of Cancer Research UK, to whom I am indebted.

Most of all, I would like to thank my incredible wife Amy, who has backed me right from when I first had the idea of applying for a PhD. Her love and support throughout have made all my hard work over the last few years possible, which I will forever be thankful for. Finally, I would like to thank our beautiful son Rowan for all the joy and laughter he has brought to our worlds, time spent together always provides a welcome relief from my work.

## Table of contents

Chapter 1 – Introduction .....	1
1.1 The prostate, carcinogenesis and disease burden .....	1
1.1.1 Anatomy and development of the prostate.....	1
1.1.2 Prostate cancer: clinical statistics.....	3
1.1.3 Prostate carcinogenesis: risk factors and disease progression .....	4
1.1.4 Molecular biology and genetics of prostate carcinogenesis.....	6
1.2 The androgen receptor .....	8
1.2.1 Androgen receptor: molecular structure and canonical signalling .....	8
1.2.2 Non-canonical AR signalling .....	13
1.2.3 AR roles in prostate biology and benign disease.....	14
1.3 The androgen receptor and prostate cancer .....	17
1.3.1 A molecular link between androgens and prostate cancer .....	17
1.3.2 Androgen deprivation therapy and antiandrogens: silver bullets? .....	18
1.4 Castrate resistant prostate cancer .....	24
1.4.1 The clinical challenge of castrate resistant prostate cancer .....	24
1.4.2 AR-mediated mechanisms of castrate resistance and therapeutic strategies .....	25
1.5 Alternative splicing .....	32
1.5.1 Alternative splicing: processes and regulation.....	32
1.5.2 Alternative splicing and cancer .....	38
1.6 Androgen receptor variants: alternatively spliced drivers of CRPC .....	40
1.6.1 Currently identified androgen receptor variants .....	40
1.6.2 AR-V structure and function: a clinical challenge.....	42
1.6.3 Current understanding of AR-V alternative splicing.....	46
1.6.4 Pharmacological inhibition of splicing.....	49
1.7 CRISPR-Cas13 and proximity biotinylation.....	51

1.7.1 CRISPR-Cas13: RNA-targeting Cas nucleases.....	51
1.7.2 Proximity biotinylation unveils protein interactomes.....	52
1.7.3 CRISPR and proximity biotinylation combine: proof of concept.....	54
<b>Chapter 2 - Aims and objectives.....</b>	<b>56</b>
<b>Chapter 3 - General materials and methods .....</b>	<b>58</b>
3.1 Cell lines and cell culture reagents.....	58
3.2 Protein harvest and western blotting .....	58
3.3 RNA extraction and RT-qPCR.....	60
3.4 Plasmid transfection.....	61
3.5 CasRx gRNA design .....	61
3.6 Statistical analyses and plotting of results .....	62
<b>Chapter 4 - Initial technical optimisation and pilot experiments .....</b>	<b>63</b>
4.1 Introduction and rationale .....	63
4.2 Specific materials and methods .....	66
4.2.1 Plasmids and molecular cloning .....	66
4.2.2 Protein structure prediction and modelling.....	66
4.2.3 CasRx gRNA transfection in HEK293FT .....	67
4.2.4 dCasRx-APEX2 protein proximity biotinylation using transient plasmid expression .....	67
4.2.5 Biotinylated protein pulldown with streptavidin (whole-cell lysates).....	68
4.3 Development of CasRx/dCasRx-APEX2 expression constructs .....	70
4.4 Protein structure prediction modelling of dCasRx-APEX2 fusion protein .....	74
4.5 Preliminary CasRx gRNA design algorithm testing.....	79
4.6 Validation of dCasRx-APEX2 utility for protein proximity biotinylation .....	82
4.7 Discussion .....	86

## Chapter 5 Development of transgenic CWR22Rv1 and CE3-targeted proteomics workflows **88**

5.1 Introduction and rationale .....	88
5.2 Specific materials and methods .....	91
5.2.1 Lentiviral production, cell line transduction and cell sorting.....	91
5.2.2 CasRx gRNA transfection in CWR22Rv1(CasRx) and CWR22Rv1(dCasRx-APEX2) ...	92
5.2.3 dCasRx-APEX2 protein proximity biotinylation using CWR22Rv1(dCasRx-APEX2) .	92
5.2.4 Biotinylated protein pulldown with streptavidin (nuclear-cytoplasmic fractionation).....	92
5.2.5 Immunofluorescence.....	93
5.2.6 Multiple structure alignment analysis of AlphaFold 2 structural predictions.....	94
5.2.7 Protein immunoprecipitation .....	94
5.2.8 RNA immunoprecipitation.....	95
5.2.9 dCasRx-APEX2 RNA biotinylation and streptavidin-pulldown assay.....	95
5.2.10 dCasRx-APEX2 proteomics sample preparation .....	98
5.2.11 Label-free mass spectrometry sample preparation and submission .....	98
5.2.12 Proteomics data analysis .....	99
5.3 CasRx(TLCV2)/dCasRx-APEX2(TLCV2) enable creation of transgenic CWR22Rv1 derivatives for use in CasRx and APEX2 workflows .....	101
5.4 Synthetic gRNAs enable targeting of AR CE3 mRNA in CWR22Rv1(CasRx) .....	103
5.5 CWR22Rv1(dCasRx-APEX2) enable live-cell proximity biotinylation of proteins .....	111
5.6 dCasRx-APEX2 is targetable to CE3 mRNA .....	116
5.7 Proteomics workflows successfully enrich splicing proteins and expected CE3 mRNA interactors .....	128
5.8 Discussion .....	144
<b>Chapter 6 Selection and validation of novel AR-variant splicing factors.....</b>	<b>149</b>
6.1 Introduction and rationale .....	149

6.2 Specific materials and methods .....	152
6.2.1 Analysis of publicly available RNA-Seq clinical cohorts.....	152
6.2.2 Analysis of publicly available microarray datasets.....	152
6.2.3 Determination of gene essentiality .....	152
6.2.4 Patient survival analysis.....	153
6.2.5 Amino acid multiple alignment between TRA2A and TRA2B.....	153
6.2.6 siRNA transfection .....	153
6.2.7 CWR22Rv1 TRA2 knockdown RNA-Seq analysis.....	154
6.2.8 Sulforhodamine B (SRB) assay .....	155
6.2.9 Phosphorodiamidate morpholino and decoy RNA oligomer transfection.....	156
6.3 Publicly available prostate cancer data facilitates splicing factor selection.....	157
6.4 Experimental validation confirms TRA2B as an AR-V7 splicing factor.....	171
6.5 Transcriptomic analyses further characterise TRA2 functions in CWR22Rv1 .....	180
6.6 TRA2 knockdown elicits contrasting proliferative responses in CWR22Rv1 and VCaP .....	198
6.7 TRA2 may regulate AR-V splicing at a CE3 splicing enhancer .....	202
6.8 Discussion.....	207
<b>Chapter 7 Conclusions and future work.....</b>	<b>213</b>
<b>Chapter 8 References .....</b>	<b>219</b>

## Glossary of key abbreviations

<b>A4</b>	Androstenedione
<b>AA</b>	Abiraterone acetate
<b>ADT</b>	Androgen deprivation therapy
<b>AF1</b>	Activation function 1
<b>AF2</b>	Activation function 2
<b>APEX2</b>	Engineered ascorbate peroxidase 2
<b>AR</b>	Androgen receptor
<b>ARE</b>	Androgen response element
<b>ARSi</b>	Androgen receptor signalling inhibitor
<b>ARTA</b>	Androgen receptor targeted agents
<b>AR-FL</b>	Androgen receptor full-length
<b>AR-V7</b>	Androgen receptor variant 7
<b>AR-Vs</b>	Androgen receptor variants
<b>BAn</b>	Biotin-aniline
<b>BP</b>	Biotin-phenol
<b>BPH</b>	Benign prostate hyperplasia
<b>CasRx</b>	RfxCas13d
<b>CDS</b>	Coding sequence
<b>CE</b>	Cryptic exon
<b>CE3</b>	Cryptic exon 3
<b>ChIP</b>	Chromatin immunoprecipitation
<b>CPM</b>	Counts per million
<b>CRISPR</b>	Clustered regularly interspersed short palindromic repeats
<b>CRPC</b>	Castrate resistant prostate cancer
<b>CSPC</b>	Castrate sensitive prostate cancer
<b>CYP17A1</b>	Cytochrome P450 17 $\alpha$ -hydroxylase/17,20-lyase
<b>DBD</b>	DNA-binding domain
<b>dCasRx</b>	Dead RfxCas13d
<b>DEG</b>	Differentially expressed gene
<b>DEU</b>	Differential exon usage

<b>DGEA</b>	Differential gene expression analysis
<b>DHEA</b>	Dehydroepiandrosterone
<b>DHEAS</b>	DHEA sulphate
<b>DHT</b>	Dihydrotestosterone
<b>DNA</b>	Deoxyribonucleic acid
<b>ECM</b>	Extracellular matrix
<b>ER</b>	Estrogen receptor
<b>ESE</b>	Exonic splicing enhancer
<b>ESS</b>	Exonic splicing silencer
<b>FACS</b>	Fluorescence-activated cell sorting
<b>FBS</b>	Foetal bovine serum
<b>FDR</b>	False-discovery rate
<b>FPKM</b>	Fragments per kilobase million
<b>GnRH</b>	Gonadotrophin releasing hormone
<b>GO</b>	Gene ontology
<b>GR</b>	Glucocorticoid receptor
<b>gRNA</b>	Guide RNA
<b>GSEA</b>	Gene set enrichment analysis
<b>H<sub>2</sub>O<sub>2</sub></b>	Hydrogen peroxide
<b>HAT</b>	Hypothalamus-anterior pituitary-testes
<b>HEPN</b>	Higher eukaryotes and prokaryotes nucleotide-binding
<b>hnRNP</b>	Heterogeneous nuclear ribonucleoproteins
<b>HSPC</b>	Hormone sensitive prostate cancer ( <i>aka</i> CSPC)
<b>iBAQ</b>	Intensity based absolute quantification
<b>ISE</b>	Intronic splicing enhancer
<b>ISS</b>	Intronic splicing silencer
<b>KEGG</b>	Kyoto encyclopaedia of genes and genomes
<b>LBD</b>	Ligand binding domain
<b>LC-MS/MS</b>	Liquid chromatography tandem mass spectrometry
<b>LH</b>	Luteinising hormone
<b>mCRPC</b>	Metastatic castrate resistant prostate cancer

<b>mPCa</b>	Metastatic prostate cancer
<b>MR</b>	Mineralocorticoid receptor
<b>mRNA</b>	Messenger RNA
<b>MV</b>	Missing value
<b>NE</b>	Neuroendocrine
<b>NICE</b>	National Institute for Health and Care Excellence
<b>NLS</b>	Nuclear localisation signal
<b>NSAA</b>	Non-steroidal antiandrogen
<b>NT</b>	Non-targeting
<b>NTD</b>	N-terminal transactivation domain
<b>ORF</b>	Open reading frame
<b>PCa</b>	Prostate cancer
<b>PCF</b>	Prostate Cancer Foundation
<b>PIN</b>	Prostatic intraepithelial neoplasia
<b>PMO</b>	Phosphorodiamidate morpholino oligomer
<b>POI</b>	Protein of interest
<b>PR</b>	Progesterone receptor
<b>PSA</b>	Prostate specific antigen
<b>RBP</b>	RNA-binding protein
<b>RIP</b>	RNA immunoprecipitation
<b>RNA</b>	Ribonucleic acid
<b>RNAi</b>	RNA interference
<b>ROAST</b>	Rotation gene set testing
<b>RRM</b>	RNA recognition motif
<b>RS domain</b>	Arginine-serine rich domain
<b>SAA</b>	Steroidal antiandrogen
<b>SDM</b>	Steroid-depleted media
<b>siRNA</b>	Small interfering RNA
<b>snRNP</b>	Small nuclear ribonucleoprotein
<b>SR protein</b>	Serine-arginine rich protein
<b>SRPM</b>	Spliced reads per million

<b>SS</b>	Splice site
<b>ssRNA</b>	Single-stranded RNA
<b>SU2C</b>	Stand Up to Cancer
<b>TCGA</b>	The Cancer Genome Atlas
<b>TRA2</b>	Transformer 2
<b>TRA2A</b>	Transformer 2 alpha homolog
<b>TRA2B</b>	Transformer 2 beta homolog
<b>UGE</b>	Urogenital epithelia
<b>UGM</b>	Urogenital sinus mesenchyme
<b>UGS</b>	Urogenital sinus

## List of figures

Figure 1.1 – Anatomical location of the prostate gland and glandular zones.....	1
Figure 1.2 - Structure and epithelial cell subtypes of prostatic ducts .....	2
Figure 1.3 - Visual summary of key stages in embryonic prostate epithelia development .....	3
Figure 1.4 - Summary of prostate cancer clinical statistics .....	4
Figure 1.5 – Stages of progression from normal prostate epithelia to metastatic prostate cancer.....	5
Figure 1.6 – Androgen receptor genetic structure and canonical signalling.....	10
Figure 1.7 - Androgen receptor functional domain structures.....	13
Figure 1.8 – Hormonal axis control of prostate AR signalling .....	15
Figure 1.9 - AR signalling and prostate 'Hallmarks of Cancer' .....	18
Figure 1.10 – Mechanisms of AR signalling inhibition by ADT and antiandrogens .....	19
Figure 1.11 - Antiandrogen structure and NSAA modulation of AR LBD conformation.....	22
Figure 1.12 - Typical PCa progression to castrate resistance .....	24
Figure 1.13 - Mechanisms enabling sustained AR signalling in CRPC - including AR-Vs, the focus of this work.....	26
Figure 1.14 - Alternative splicing patterns.....	32
Figure 1.15 - Stages of exon splicing by the spliceosome.....	34
Figure 1.16 (previous page) – Spliceosomal snRNP function involves extensive protein/RNA interactions .....	36
Figure 1.17 - SR proteins and hnRNPs bind RNA-encoded regulatory elements to influence alternative splicing.....	37
Figure 1.18 - Exonic structure of prominent AR-variants .....	41
Figure 1.19 - AR-V7 elicits resistance to abiraterone and NSAs in preclinical models of CRPC .....	43
Figure 1.20 - AR-V7 expression correlates with development of CRPC, hormonal therapy exposure and poor clinical outcomes .....	44
Figure 1.21 - Cas13 is a novel CRISPR nuclease usable for studying RNA biology.....	52
Figure 1.22 - Proximity biotinylation methods enable proteomic characterisation of interactomes .....	54
Figure 1.23 - Proximity biotinylation can combine with Cas9/Cas13 to elucidate sequence-specific DNA/RNA protein interactomes .....	55
Figure 2.1 - Graphical summary of PhD project aims and objectives, with relevant chapter numbers noted.....	57
Figure 4.1 - Expression of CasRx/dCasRx plasmids is detectable by western blot and GFP microscopy.....	71
Figure 4.2 - Plasmid criteria and transgenic CWR22Rv1 generation strategy .....	72
Figure 4.3 (previous page) - CasRx and dCasRx-APEX2 were successfully subcloned into TLCV2 .....	73
Figure 4.4 - Protein disorder prediction of dCasRx reveals inherent N and C-terminal flexibility.....	76
Figure 4.5 - Protein structure modelling of dCasRx-APEX2 indicates a separation of functional fusion protein domains.....	78

Figure 4.6 - A published algorithm enables CasRx gRNA design for transcript targeting .....	80
Figure 4.7 - Application of a CasRx gRNA design algorithm in HEK293FT elicits TP53 knockdown by transfection with synthetic gRNA and pXR001 .....	81
Figure 4.8 - dCasRx-APEX2 plasmid pXR002-APEX2 enables live-cell protein biotinylation and streptavidin enrichment from lysates in HEK293FT .....	83
Figure 4.9 - Endogenous biotinylated proteins negatively impact APEX2 effectiveness in CWR22Rv1.....	84
Figure 5.1 - Lentiviral transduction with CasRx(TLCV2) and dCasRx-APEX2(TLCV2) enables creation of doxycycline-inducible CWR22Rv1 derivatives .....	102
Figure 5.2 - Use of synthetic gRNA in CWR22Rv1(CasRx) elicits AR mRNA knockdown .....	105
Figure 5.3 - Transfection of 25 nM CasRx gRNA using RNAiMAX mediates efficient cellular uptake in CWR22Rv1(CasRx).....	107
Figure 5.4 - Knockdown using synthetic gRNAs in CWR22Rv1(CasRx)is maintained over 72 hours .....	110
Figure 5.5 - Nuclear/cytoplasmic fractionation enables effective enrichment of biotinylated proteins in CWR22Rv1(dCasRx-APEX2).....	112
Figure 5.6 (previous page) - Immunofluorescence visualises nuclear biotinylation in CWR22Rv1(dCasRx-APEX2).....	114
Figure 5.7 - Multiple structure alignment analyses demonstrate high structural similarity between dCasRx and dCasRx-APEX2 .....	117
Figure 5.8 - Schematic of RNA immunoprecipitation assay to verify dCasRx-APEX2 binding with CE3 mRNA .....	118
Figure 5.9 - RNA immunoprecipitation of dCasRx-APEX2 targeted at <i>TP53</i> mRNA fails to consistently enrich <i>TP53</i> transcripts in HEK293FT.....	120
Figure 5.10 - RNA immunoprecipitation of dCasRx-APEX2 targeted at CE3 mRNA fails to consistently enrich AR-V7 transcripts in CWR22Rv1(dCasRx-APEX2) .....	122
Figure 5.11 - Biotin-aniline mediated APEX2 proximity biotinylation provides an effective means to validate dCasRx-APEX2 binding at CE3 mRNA .....	124
Figure 5.12 - RNA biotinylation with biotin-aniline enables a CE3 pulldown assay in CWR22Rv1(dCasRx-APEX2).....	126
Figure 5.13 - CWR22Rv1(dCasRx-APEX2) CE3-targeted proteomics experimental setup.....	128
Figure 5.14 - Nuclear lysates from proteomics experimental arms, detailed in 5.14, are utilised for LC-MS/MS .....	130
Figure 5.15 - Simplified summary of computational analysis pipeline for mass spectrometry data .....	132
Figure 5.16 - Proportion of MVs and valid assigned iBAQ values from initial n = 4 MaxQuant processing .....	133
Figure 5.17 - CE3-targeted dCasRx-APEX2 proteomics enriches previously published AR-V7 splicing regulators and selects for nuclear/RNA-binding proteins.....	135
Figure 5.18 - Assaying tryptic peptide concentrations enables a significantly greater proportion of MaxQuant protein identifications.....	137
Figure 5.19 - Injection of equal peptide concentrations for LC-MS/MS enables greater protein intensity homogeny between samples .....	139

Figure 5.20 - Experimental optimisation of proteomics workflows significantly improved dataset power .....	140
Figure 5.21 - STRING analysis of proteins enriched by AR g2 demonstrates significant enrichment for the expected biological processes.....	141
Figure 5.22 - Previously identified regulators of AR-V7 generation are enriched by our dCasRx-APEX2 workflow .....	143
Figure 6.1 - Correlation of enriched proteins' gene expression with AR-V7 splicing in an mCRPC patient cohort.....	158
Figure 6.2 (previous page) - Target gene expression data may provide an auxiliary readout of AR-V7 generation.....	160
Figure 6.3 - Correlation of enriched protein genes with clinically identified AR-V7 signatures provides additional evidence for protein selection .....	162
Figure 6.4 - A panel of genes encoding dCasRx-APEX2-enriched splicing factors are associated with AR-V7 expression in TCGA-PRAD samples .....	164
Figure 6.5 - Summary of CWR22Rv1 CE3 interactome triangulation with publicly available datasets.....	165
Figure 6.6 - TRA2B and THRAP3 emerge as lead splicing candidates for functional analysis .....	167
Figure 6.7 - Supplementary analyses further vindicate selection of TRA2B and THRAP3 as candidates.....	169
Figure 6.8 - SART1 was selected as an additional factor upregulated in mCRPC and associated with poor PCa survival .....	170
Figure 6.9 - Initial assessment of AR-FL and AR-V7 mRNA levels upon target depletion in CWR22Rv1.....	171
Figure 6.10 - Dual TRA2 depletion in CWR22Rv1 alters AR mRNA splicing decisions to significantly deplete AR-V7 levels.....	173
Figure 6.11 - Effects of dual TRA2A/B depletion in CWR22Rv1 are also observed at the protein level .....	174
Figure 6.12 - AR-V7 splicing is controlled by TRA2 in VCaP .....	176
Figure 6.13 - Enzalutamide mediates an elevation in AR-V7 in VCaP, which is quashed by TRA2 depletion.....	178
Figure 6.14 - TRA2 protein TRA2B specifically correlates with AR-V7 in clinical patient cohorts .....	182
Figure 6.15 - Both TRA2 paralogs are associated with poorer clinical outcomes in prostate cancer.....	183
Figure 6.16 - Adapter trimming with <i>Cutadapt</i> successfully eliminates Illumina adapter contamination.....	185
Figure 6.17 - A computational pipeline enables differential gene expression analysis of RNA-Seq data .....	186
Figure 6.18 - Principal component analysis and hierarchical clustering demonstrate sample expression similarities and differences between experimental arms.....	188
Figure 6.19 - TRA2 paralogs are efficiently depleted by siRNAs and combined depletion results in significantly greater differential expression .....	190

Figure 6.20 - Overlap analysis of TRA2 knockdown significantly DEGs alludes to both redundant and unique expression changes.....	191
Figure 6.21 - Combined TRA2A/B depletion drives a switch from AR-V7 to AR-FL signalling in CWR22Rv1.....	192
Figure 6.22 - Additional gene set enrichments further demonstrate an <i>AR</i> splicing switch.....	193
Figure 6.23 - Differential exon usage analysis of RNA-Seq reads confirms a TRA2 control of an <i>AR</i> splicing switch.....	195
Figure 6.24 - AR CE3 and exon 4 are amongst the most differentially utilised exons in the annotated transcriptome upon TRA2A/B depletion in CWR22Rv1.....	196
Figure 6.25 - Alternative cryptic exon inclusion is also regulated by TRA2.....	197
Figure 6.26 - (previous page) CWR22Rv1 and VCaP cells exhibit contrasting proliferative responses to TRA2 depletion .....	200
Figure 6.27 - Oligonucleotide approaches represent a novel means to modulate CRPC splicing activity.....	203
Figure 6.28 - Transfection with CE3 block PMO effectively depletes AR-V7, but not AR-FL, levels to confirm functional relevance of the target mRNA region.....	206
Figure 6.29 - Mechanistic summary of TRA2 proteins' putative role in AR-V7 splicing .....	212

## List of tables

Table 1.1 – Literature summary of prominent genetic/molecular alterations in PCa development.....	7
Table 1.2 - Literature summary of currently identified AR-V splicing/processing regulators	49
Table 3.1 - Culture media conditions for cell lines .....	58
Table 3.2 - Buffers used for protein harvest, PAGE and western blot transfer.....	59
Table 3.3 - Primary antibodies used for western blot .....	60
Table 3.4 - Primers used for qPCR analysis.....	61
Table 4.1 - Summary of software and corresponding websites used for protein structure prediction.....	67
Table 4.2 - CasRx gRNA sequences used for <i>TP53</i> mRNA targeting in HEK293FT .....	67
Table 4.3 - Amino acid frequency of occurrence in naturally occurring multidomain protein linkers.....	75
Table 5.1 - CasRx gRNA sequences used for <i>AR</i> mRNA targeting in CWR22Rv1(CasRx) and CWR22Rv1(dCasRx-APEX2).....	92
Table 6.1 - siRNA sequences used in knockdown experiments.....	154



# Chapter 1 – Introduction

## 1.1 The prostate, carcinogenesis and disease burden

### 1.1.1 Anatomy and development of the prostate

The prostate is a walnut-sized gland found in males, surrounding the urethra at the base of the penis and bladder (Figure 1.1A). Through the prostatic ducts it secretes the lipids, enzymes and metal ions that make up prostatic fluid, crucial to the protection and function of spermatozoa (Kumar and Majumder, 1995). Three distinct regions make up the glandular ‘zones’ of the prostate: the peripheral, transition and central zones. The majority of glandular epithelium is found in the peripheral zone, comprising ~70% of the prostate (Ittmann, 2018). The transition zone consists of two portions located either side of the prostatic urethra, making up only 5-10% of the gland in young males, however it commonly enlarges with age as the site of benign prostate hyperplasia (BPH). The central zone is a cone shaped tissue surrounding the ejaculatory ducts, the point of which lies at the ducts’ confluence with the urethra (Figure 1.1B) (Ittmann, 2018; Hammerich *et al.*, 2008). In addition to glandular epithelium, the prostate contains a fibromuscular stroma dominated by smooth muscle cells, fibroblasts and elastic connective tissue (Hägglöf and Bergh, 2012; Shafik *et al.*, 2006).

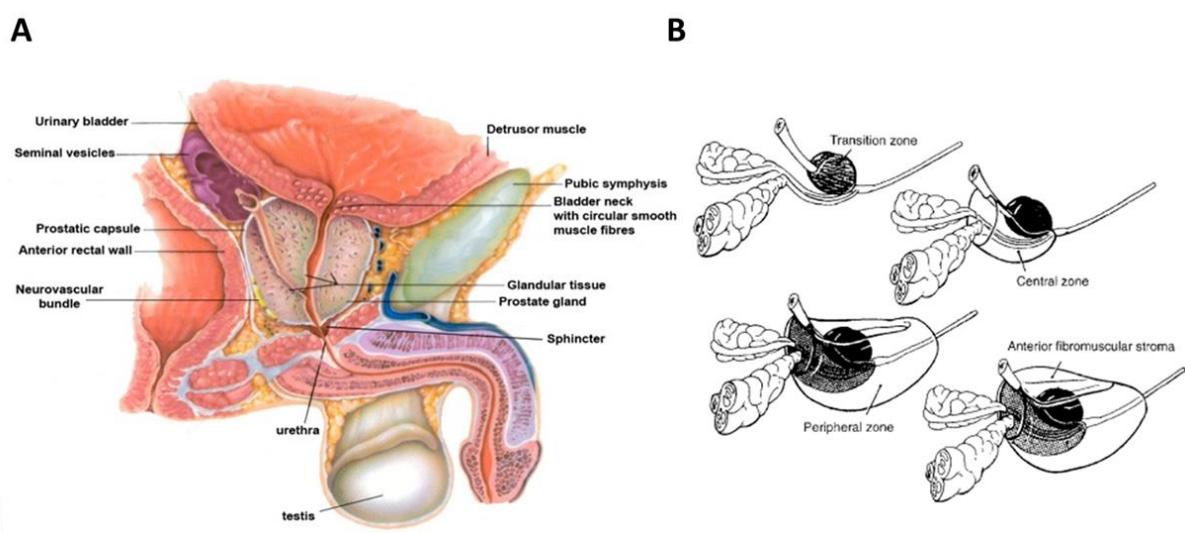
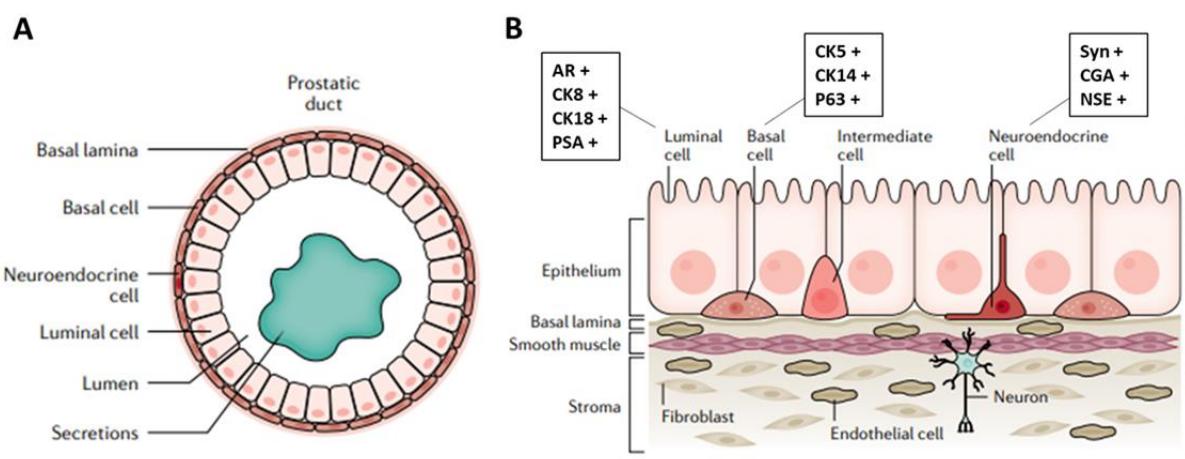


Figure 1.1 – Anatomical location of the prostate gland and glandular zones

**A.** Illustration of prostate gland location between the base of the penis and bladder, surrounding the ejaculatory ducts and urethra (Figure taken from Roberts *et al.*, 2000). **B.** Location of the peripheral, transition and central zones, in relation to the seminal vesicles and urethra (Figure taken from Hammerich *et al.*, 2008)

Cytologically, prostate glandular tissue contains three epithelial cell types: luminal, basal and neuroendocrine (NE) (Liu and True, 2002). Prostatic lumens are lined by luminal cells, which perform the organ's major secretory functions. These are surrounded by a layer of basal cells, responsible for extracellular matrix (ECM) deposition, luminal cell survival and ductal structural integrity (Figure 1.2A) (Rebello *et al.*, 2021; Kurit *et al.*, 2004). NE cells are a rare subpopulation, comprising only ~1% of prostate epithelia. Their function is poorly characterised, they are known to secrete peptide hormones and are proposed to communicate with other epithelial cell types in a paracrine manner (Butler and Huang, 2021). Each epithelial cell type can be identified biochemically based on protein expression (Figure 1.2B). Furthermore an 'intermediate' epithelial phenotype has been identified, with expression characteristics of both luminal and basal cells (Hudson *et al.*, 2001; van Leenders *et al.*, 2003).



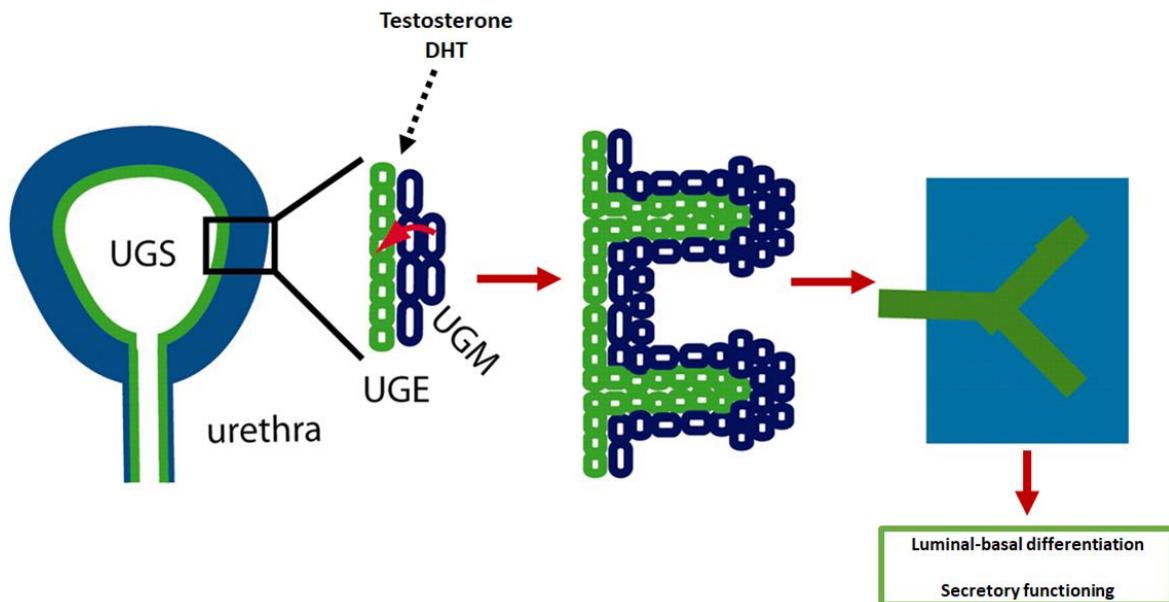
**Figure 1.2 - Structure and epithelial cell subtypes of prostatic ducts**

**A.** Luminal epithelial cells line the lumen of prostatic ducts and carry out the major secretory functions. These lie above a layer of basal cells and rare neuroendocrine subpopulations (Figure taken from Rebello *et al.*, 2021)

**B.** Prostate epithelial cells express unique combinations of protein markers, and are apical to a layer of extracellular matrix and stroma (Figure adapted from Rebello *et al.*, 2021)

Human prostate development starts around embryonic week 10. At this point epithelial buds emerge from the embryonic urogenital sinus (UGS), driven by interactions between urogenital sinus epithelia (UGE) and urogenital sinus mesenchyme (UGM). Developmental responses to this interaction rely on androgens, produced by foetal testes from around week 8 (Cunha *et al.*, 2010; Cunha *et al.*, 2018). Epithelial outgrowths accelerate between weeks 11-14, whereupon they develop into tubulo-acinar glands (Kasper, 2008; Cunha *et al.*, 2018; Kellokumpu-Lehtinen *et al.*, 1980). Epithelia differentiate into basal and luminal phenotypes

as the prostate grows throughout male fetal development. Cunha *et al.* (2018) divide prostate development into six stages: i) pre-bud UGS, ii) epithelial bud development from UGE, iii) bud elongation and branching, iv) epithelial cord canalization, v) luminal-basal epithelial differentiation, and vi) secretory cytodifferentiation (Figure 1.3).

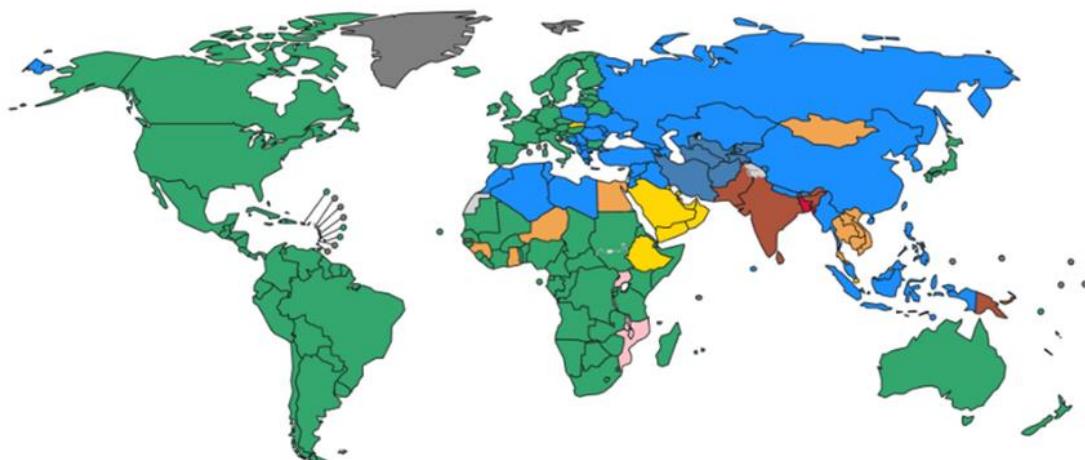
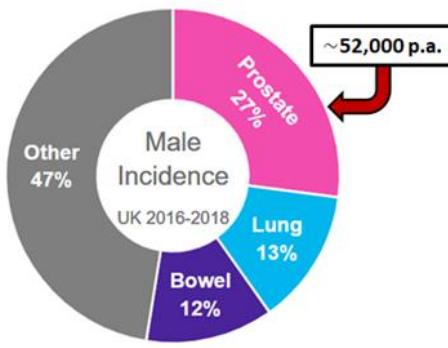
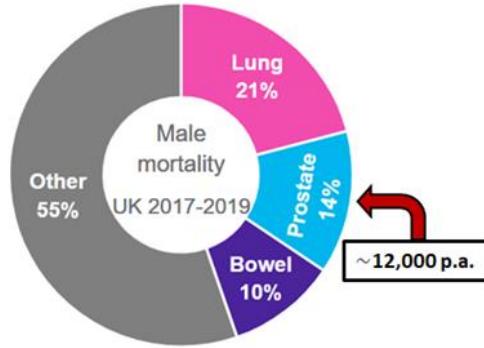


**Figure 1.3 - Visual summary of key stages in embryonic prostate epithelia development**

Interactions between UGE (green) and UGM (blue) at the embryonic UGS induce epithelial bud development and UGE elongation into the UGM, a process that relies on androgens testosterone and dihydrotestosterone (DHT). UGE buds then continue to elongate, branch and canalize, forming channels. Over time UGE differentiates into the epithelial cell subtypes that form functional prostate glandular epithelium as outlined in Figure 1.2 (Figure adapted from Meeks and Schaeffer, 2011)

### 1.1.2 Prostate cancer: clinical statistics

Prostate cancer (PCa) is the second most diagnosed cancer in men globally, whilst in many countries including the UK and USA PCa incidence exceeds that of any other non-cutaneous cancer (Figure 1.4A) (Sung *et al.*, 2021). Furthermore, incidence rates per 100,000 men are projected to increase up to at least the year 2035 (Smittenaar *et al.*, 2016). The UK sees approximately 52,000 cases and 12,000 deaths from PCa annually comprising, respectively, 27% and 14% of all male cancer diagnoses and deaths (Figure 1.4B, C) (Cancer Research UK, 2019).

**A****B****C**

**Figure 1.4 - Summary of prostate cancer clinical statistics**

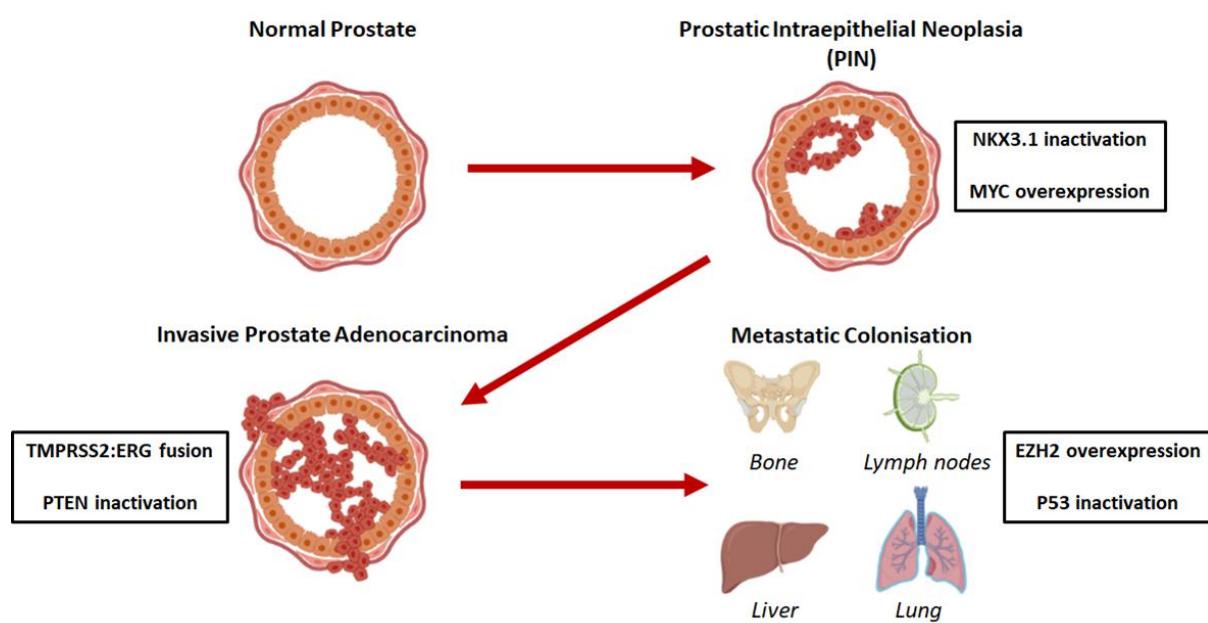
**A.** Global summary of countries coloured by most diagnosed non-cutaneous cancer type in males. Countries in dark green are those in which prostate cancer is most common (Figure taken from Sung *et al.*, 2021) **B.** UK summary of highest incidence rate cancers in males, with number of prostate cancer cases per annum highlighted (Figure adapted from Cancer Research UK, 2019) **C.** Summary of most prevalent contributors to male cancer mortality in the UK, with number of prostate cancer deaths per annum highlighted (Figure adapted from Cancer Research UK, 2019)

### 1.1.3 Prostate carcinogenesis: risk factors and disease progression

Some 75% of PCa cases develop in the peripheral zone of the gland (Adler *et al.*, 2012). Furthermore, the vast majority (>90%) of PCa arises from glandular epithelium as adenocarcinoma (Leslie *et al.*, 2022; Alizadeh and Alizadeh, 2014; Grignon 2004). Epidemiological analyses identify inherent risk factors associated with PCa development. These include: i) age, as for all cancers risk of PCa development rises sharply with age; ii) race, black ethnicity has been noted as a significant risk factor; and iii) inherited mutations, including in genes involved in mismatch repair (*MLH1*, *MSH6*), homologous recombination (*BRCA1/2*, *PALB2*, *ATM*), and transcription factors such as the *HOXB13* G48E mutation (Bostwick *et al.*, 2004; Jones and Chinegwundoh, 2014; Lynch *et al.*, 2016; Vietri *et al.*, 2021; Ewing *et al.*, 2012). Lifestyle exposures linked to PCa have also been researched including diet,

alcohol consumption, smoking and exposure to heavy metals *eg* cadmium. However, evidence related to these exogenous factors is inconclusive (Bostwick *et al.*, 2014; Leitzmann and Rohrmann, 2012).

Multiple stages exist between healthy prostate and invasive metastatic disease (mPCa) (Figure 1.5). It is recognised that intraluminal hyperplasia, termed prostatic intraepithelial neoplasia (PIN), exists as a precursor to PCa (McNeal and Bostwick, 1986; Bostwick and Brawer, 1987). PIN is characterized by luminal epithelial proliferation and in high-grade cases significant basal cell disruption may occur. PIN also shares cytological and histological changes with PCa; however, it is crucially distinguished by a lack of basement membrane penetration (Brawer, 2005; Bostwick and Qian, 2004; Bostwick, 1995). Transition to PCa is signified by significant disruption of and/or invasion through the basement membrane into the surrounding stroma, where it may progress locally (Bostwick and Brawer, 1987). Advanced mPCa involves metastatic colonization of other anatomical sites, most commonly osteoblastic lesions of the spine, pelvis and ribs (Logothetis and Lin, 2005; Bagi, 2003). Other frequently observed sites of metastasis are the lymph nodes, liver and lungs (Gandaglia *et al.*, 2014; Pezaro *et al.*, 2014; Bubendorf *et al.*, 2000).



**Figure 1.5 – Stages of progression from normal prostate epithelia to metastatic prostate cancer**

Prostate epithelia can evolve from localised intraepithelial proliferation (PIN) to invasive adenocarcinoma and metastatic cancer. Numerous molecular and genomic events involved in this progression have been characterised, some of the most prominently studied of which are highlighted above (black boxes). A more

extensive summary of the molecular regulators of prostate carcinogenesis can be found in Table 1.1 (Figure created using biorender.com)

The most widely used assessment of PCa tumours by biopsy, the Gleason grade, scores tumour tissue from 1 to 5, with a grade of 1 indicating well-differentiated, uniform glands and a grade of 5 comprising sheets of anaplastic cells with necrotic regions. Gleason scoring is a valuable prognostic marker, with higher grades predictive of more aggressive disease (Hoogland *et al.*, 2014; Humphrey, 2004; Egevad *et al.*, 2002; Gleason and Mellinger, 1974; Gleason, 1966). PCa has long been identified as a multifocal disease, with observations seen of a prostate 'field effect' containing multiple independent neoplasia (Arora *et al.*, 2004; Cheng *et al.*, 1998; Miller and Cygan, 1994; Andreou and Cheng, 2010). This multifocality makes tumour grading and prognostication challenging, as one prostate gland may contain multiple cancers with differing grades (Andreou and Cheng, 2010; Arora *et al.*, 2004).

#### **1.1.4 Molecular biology and genetics of prostate carcinogenesis**

Numerous somatic molecular and genetic events involved in PCa development have been identified (Figure 1.5, black boxes. Table 1.1). For example homeobox gene *NKX3.1* maps to locus 8p21 which is frequently deleted in PIN (Emmert-Buck *et al.*, 1995). Mouse knockout models demonstrate *NKX3.1* haploinsufficiency, as loss of one allele is sufficient to induce PIN (Bhatia-Gaur *et al.*, 1999; Abdulkadir *et al.*, 2002; Magee *et al.*, 2003). *MYC* oncogene overexpression has been observed in both high-grade PIN and advanced PCa, suggesting it is another early event in prostate carcinogenesis (Gurel *et al.*, 2004; Iwata *et al.*, 2010; Ellwood-Yen *et al.*, 2003; Qiu *et al.*, 2022). Another prominent event is a genetic translocation on chromosome 21, creating a fusion between the promoter of androgen-responsive gene *TMPRSS2* and the coding sequence of proto-oncogene and ETS family member *ERG* (or, less commonly, ETS member *ETV1*). The resultant *TMPRSS2:ERG* fusion, detected in approximately 50% of PCa, is then constitutively activated by androgen signalling in PCa where it drives proliferation (Tomlins *et al.*, 2005; Tomlins *et al.*, 2008; Park *et al.*, 2014; Wang *et al.*, 2017).

Many other abnormalities beyond the scope of this review have been implicated in PCa progression. These include loss or inactivation of tumours suppressors *PTEN* and *TP53*, activation of the MEK/ERK pathway and overexpression of transcriptional regulator protein EZH2 (Jamaspishvili *et al.*, 2018; Teroerde *et al.*, 2021; Nickols *et al.*, 2019; Varambally *et al.*,

2002; Wu *et al.*, 2019). A summary of the most prominently characterised alterations with associated references can be found in Table 1.1. Androgen receptor (AR) signalling is implicated in all stages from prostate development through to advanced mPCa and is the primary focus of this review.

Gene/Alteration	Approximate frequency in prostate cancer		References
	Primary	Metastatic	
<b>NKX3.1/loss</b>	20-35%	75%	Emmert-Buck <i>et al.</i> , 1995; Bhatia-Gaur <i>et al.</i> , 1999; Bowen <i>et al.</i> , 2000; Abdulkadir <i>et al.</i> , 2002; Magee <i>et al.</i> , 2003
<b>PTEN/loss</b>	20%	50%	Jamaspishvili <i>et al.</i> , 2018; Wang <i>et al.</i> , 2003; Cairns <i>et al.</i> , 1997
<b>TP53/loss</b>	40-50%	40-50%	Abida <i>et al.</i> , 2019; Zhou <i>et al.</i> , 2006; Osman <i>et al.</i> , 1999; Eastham <i>et al.</i> , 1995; Bookstein <i>et al.</i> , 1993; Isaacs <i>et al.</i> , 1991
<b>RB1/loss</b>	15%	20%	Abida <i>et al.</i> , 2019; Thangavel <i>et al.</i> , 2017; Maddison <i>et al.</i> , 2004; Zhou <i>et al.</i> , 2006; Burkhardt <i>et al.</i> , 2019
<b>SPOP/loss</b>	10-15%	10-15%	Kim <i>et al.</i> , 2013; Boysen <i>et al.</i> , 2015; Barbieri <i>et al.</i> , 2012; Bernasocchi and Theurillat, 2022
<b>CHD1/loss</b>	10%	20%	Oh-Hohenhorst <i>et al.</i> , 2021; Augello <i>et al.</i> , 2019; Zhang <i>et al.</i> , 2020; Huang <i>et al.</i> , 2012
<b>GSTP1/promoter hypermethylation</b>	80%	Unspecified	Lin <i>et al.</i> , 2001; Nakayama <i>et al.</i> , 2004; Martignano <i>et al.</i> , 2016; Cairns <i>et al.</i> , 2001
<b>TMPRSS2:ERG/fusion</b>	40-50%	40-50%	Tomlins <i>et al.</i> , 2005; Tomlins <i>et al.</i> , 2008; Park <i>et al.</i> , 2014; Wang <i>et al.</i> , 2017
<b>MYC/overexpression</b>	10%	35%	Gurel <i>et al.</i> , 2004; Iwata <i>et al.</i> , 2010; Ellwood-Yen <i>et al.</i> , 2003; Qiu <i>et al.</i> , 2022
<b>SPINK1/overexpression</b>	10%	Unspecified	Ateeq <i>et al.</i> , 2011; Flavin <i>et al.</i> , 2014; Lu <i>et al.</i> , 2020; Tomlins <i>et al.</i> , 2008
<b>EZH2/overexpression</b>	Frequency unspecified. An established positive correlation exists between EZH2 levels and prostatic progression from benign tissue, to primary tumours and metastases		Varambally <i>et al.</i> , 2002; Wu <i>et al.</i> , 2019; Bryant <i>et al.</i> , 2007; Labb�� <i>et al.</i> , 2017
<b>AR/mutation, amplification, alternative splicing</b>	Variable, discussed in sections 1.4 onwards		Tan <i>et al.</i> , 2015 (review article – see subsequent sections for more extensive review of AR in prostate cancer)

**Table 1.1 (legend overleaf) – Literature summary of prominent genetic/molecular alterations in PCa development**

Literature synopsis of the most well characterised genetic and molecular changes involved in PCa progression and their approximate frequencies, where applicable, in primary and metastatic PCa. ‘loss’ = loss of function through either mutation or deletion, ‘overexpression’ = increased expression through either amplification or other mechanisms. Roles of the androgen receptor (AR) are discussed extensively in subsequent sections

## 1.2 The androgen receptor

### 1.2.1 Androgen receptor: molecular structure and canonical signalling

The androgen receptor (AR) is a nuclear steroid hormone receptor, a family of proteins including the estrogen (ER), progesterone (PR) and glucocorticoid (GR) receptors (Guiochon-Mantel *et al.*, 1996). AR is expressed in a range of cell types, with the highest levels being seen in the reproductive (prostate, testes), liver, muscle and adipose tissues where it is responsible for exerting the biological effects of androgens including androstenedione, testosterone and dihydrotestosterone (DHT) (Hunter *et al.*, 2018; Mooradian *et al.*, 1987). The importance of AR to numerous tissues has been demonstrated by knockout mouse models, in which either global or tissue-specific *AR* knockout mice exhibit perturbations in the brain, bone, cardiovascular, immune and haematopoietic systems, in addition to tissues outlined previously (Davey and Grossmann *et al.*, 2016; Rana *et al.*, 2014). The *AR* locus is at chromosome Xq11-12, spanning approximately 90kb of DNA and comprising 8 exons that encode N-terminal transactivation domain (NTD), DNA-binding (DBD), hinge (H) and ligand-binding (LBD) domains (Figure 1.6A) (Gelmann *et al.*, 2002).

Canonical, or genomic, AR signalling is mediated by intracellular binding of androgen ligands, the highest affinity of which is DHT synthesised from testosterone by 5 $\alpha$ -reductase (Kinter and Anekar, 2022; Randall, 1994, Grino *et al.*, 1990). In the prostate, testosterone enters cells and undergoes conversion to DHT before binding the LBD of the receptor. Unbound by ligand, AR is cytoplasmically sequestered in a high affinity ligand-binding state by chaperone proteins including HSP70 and HSP90, which dissociate upon DHT binding (Fang *et al.*, 1996; Pratt and Toft, 2003; Veldscholte *et al.*, 1992; Pratt and Toft, 1997). DHT binding induces a conformational change in receptor structure, whereby intramolecular N/C terminal interactions take place followed by receptor homodimerisation through intermolecular DBD and N/C terminal interactions (van Royen *et al.*, 2012; Schaufele *et al.*, 2005). Ligand stimulation also results in AR phosphorylation, largely in its NTD. Phosphorylation at S81 by CDK1, CDK5 and CDK9 has been particularly well studied, where cell line models demonstrate robust increases in pS81-AR upon DHT treatment (Chen *et al.*, 2012; Chen *et al.*, 2006; Hsu *et al.*, 2011; Gordon *et al.*, 2009; Wen *et al.*, 2020). DHT-bound, phosphorylated AR translocates to the nucleus, directed by interactions between its nuclear localisation signal (NLS), nuclear

import protein importin- $\alpha$  and actin-binding protein filamin (Cutress *et al.*, 2008; Ni *et al.*, 2013; Ozanne *et al.*, 2000). AR homodimers predominantly function as transcriptional activators, binding to DNA androgen response elements (AREs) in the promoters and enhancers of target genes alongside chromatin remodelling proteins (CBP/P300, LSD1) and coactivators (p160 family) to recruit RNA Pol II, facilitating transcription (Figure 1.6B) (Denayer *et al.*, 2012; Louie *et al.*, 2003; Frønsdal *et al.*, 1998; Metzger *et al.*, 2005; Jin *et al.*, 2013).

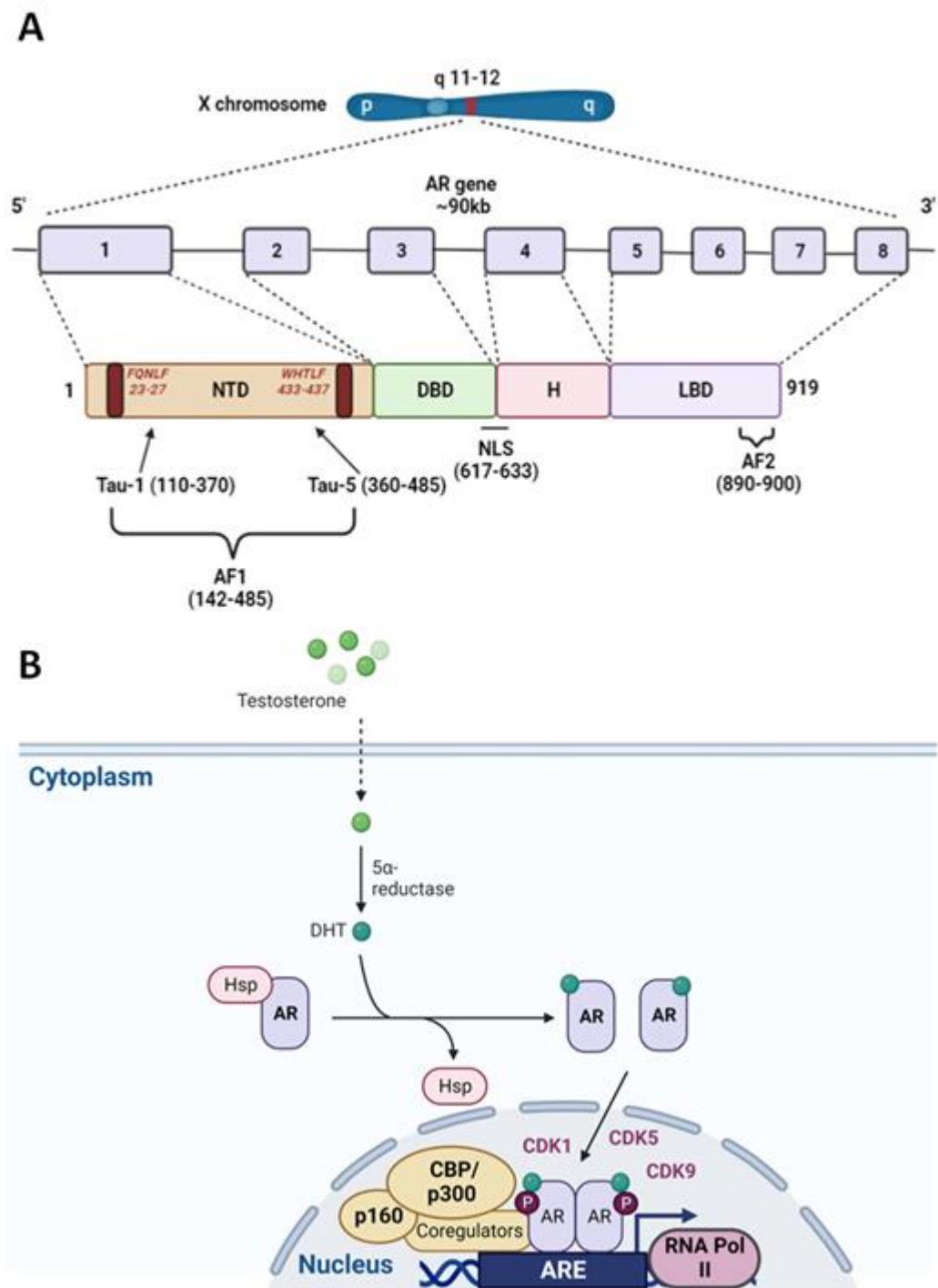


Figure 1.6 – Androgen receptor genetic structure and canonical signalling

**A.** The ~90kb AR gene is located on chromosome Xq11-12 and encodes 8 exons. Structural and functional details of the N-terminal (NTD), DNA-binding (DBD), hinge (H) and ligand-binding (LBD) domains, as well as the transcription activation units (Tau-1, Tau-5), activation function domains (AF1, AF2) and nuclear-localisation signal (NLS) are explored in text (exon sizes not to scale) **B.** Testosterone enters prostate cells and is metabolised to dihydrotestosterone (DHT) by 5 $\alpha$ -reductase, which bind the AR LBD. DHT-bound AR is phosphorylated, homodimerises and translocates to the nucleus to bind androgen response elements (AREs), where along with coregulators it recruits transcriptional machinery to drive expression of target genes (Figures created using biorender.com)

### ***N-terminal transactivation domain (NTD)***

The NTD, encoded by exon 1, is responsible for AR-mediated transactivation, in part, by interacting with p160 coactivators and other transcriptional machinery at AREs (Nakka *et al.*, 2013; Bevan *et al.*, 1999; McEwan and Gustafsson, 1997). The activation function 1 (AF1) region, located between amino acids 142-485, is crucial to NTD transactivation, within which lie two transcriptional activation units Tau-1 and Tau-5 indispensable for AR transcriptional activity (Bevan *et al.*, 1999; Jenster *et al.*, 1995; Callewaert *et al.*, 2006; McEwan, 2004). The NTD also contains FQNL and WHTLF motifs, responsible for N/C terminal interactions upon ligand binding (Figure 1.6A) (He *et al.*, 2000; Doesburg *et al.*, 1997). NTD mediated transactivation is further influenced by polyglutamine (CAG) and polyglycine (GGC) repeats, the length of which vary considerably among the population and, in the case of CAG, inversely correlate with AR activity (Tut *et al.*, 1997; Chamberlain *et al.*, 1994; Sasaki *et al.*, 2003; Irvine *et al.*, 1995). Despite recognition of these motifs and domains, a precise crystal structure of the NTD is undefined due to its inherently disordered nature and conformational plasticity (Lavery and McEwan, 2008; Reid *et al.*, 2002).

### ***DNA-binding domain (DBD)***

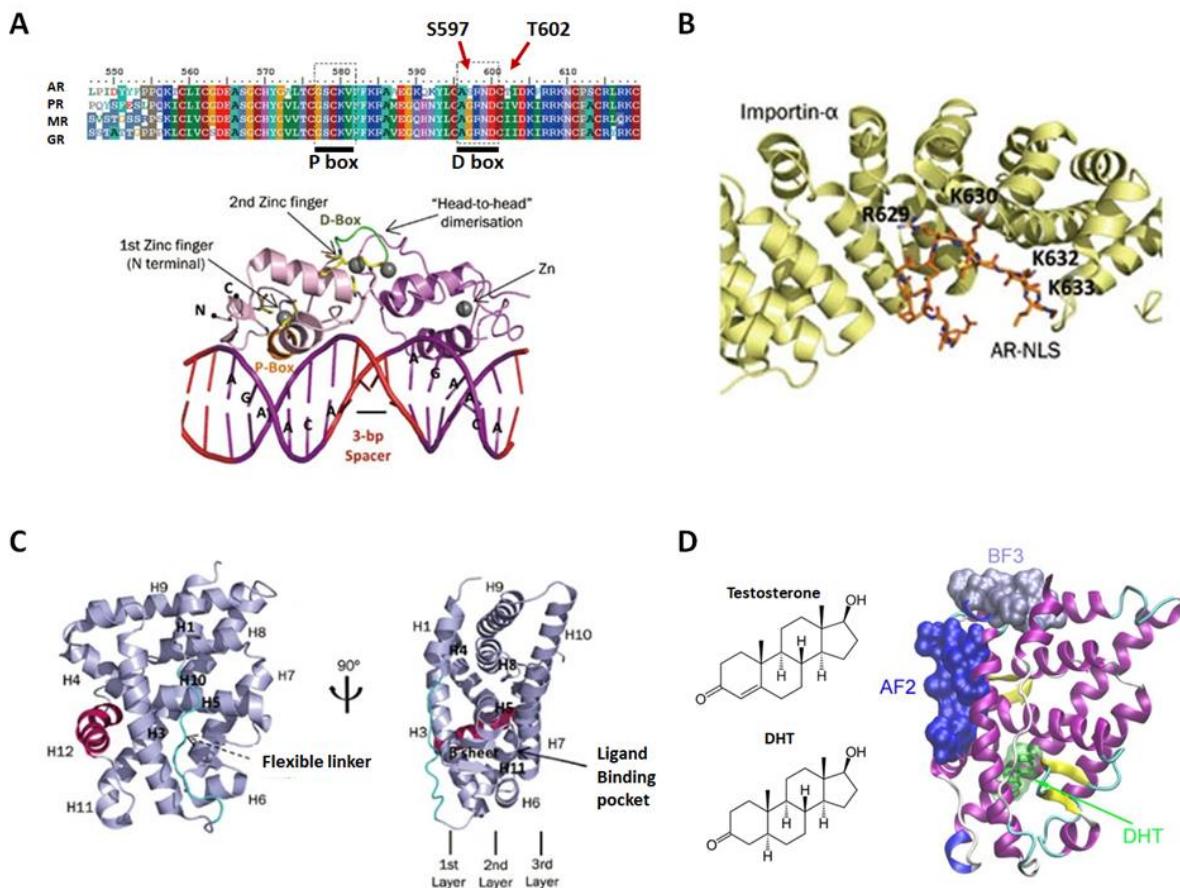
The AR DBD encoded by exons 2 and 3 is highly conserved, sharing considerable homology with the GR, PR and mineralocorticoid receptor (MR) (Figure 1.7A). The DBD comprises two zinc fingers, where the  $\alpha$ -helix of the N-terminal zinc finger directly interacts, via its 'P-box' motif, with nucleotides in either the generic hormone response hexamer sequence 2x 5'-AGAACCA-3' (separated by a 3-bp spacer) or specific AREs *eg* 5'-GGTTCT-3'. The second zinc finger stabilises interactions through contact with the sugar-phosphate backbone (Beato *et al.*, 1995; Zhou *et al.*, 1997; Claessens *et al.*, 1996; Schoenmakers *et al.*, 2000, Tan *et al.*, 2015). The precise mechanism of AR DBD selectivity for AREs is unclear. It has been proposed that stronger dimerisation relative to other steroid hormone receptors may play a role. This is mediated by a serine and threonine in the 'D-box' of the second zinc finger, S597 and T602, which are unique to AR and enable greater hydrogen bonding between monomers. This increased interaction strength relative to other steroid hormone receptors is purported to stabilise AR dimerization for retention at AREs (Shaffer *et al.*, 2004; van Royen *et al.*, 2012; Claessens *et al.*, 2008; Tan *et al.*, 2015) (Figure 1.7A).

### ***Hinge region (H)***

Exon 4 encodes the hinge region, largely responsible for nuclear import via an NLS at its junction with the DBD: 617-RKCYEAGMTLGARKLKK-634 (Figure 1.6A). Androgen binding has been proposed to mediate structural changes of AR which expose the NLS for binding and nuclear translocation, in which the AR NLS fits within the inner concave surface of importin- $\alpha$  (Figure 1.7B) (Cutress *et al.*, 2008; Ni *et al.*, 2013). Structural studies have also suggested an important role for the AR hinge region in ARE selectivity as its sequence is poorly conserved and unique to the AR (Cutress *et al.*, 2008; Clinckemalie *et al.*, 2012). Furthermore, it has been shown as involved in regulating coactivator recruitment and transactivation, and is subject to extensive post-translational modification (Haelens *et al.*, 2007; Clinckemalie *et al.*, 2012).

### ***Ligand-binding domain (LBD)***

The LBD of AR is encoded by exons 5-8 and is well characterised crystallographically. It consists of eleven  $\alpha$ -helices, lacking the H2 of other nuclear receptor LBDs in place of a flexible linker region (Figure 1.7C) (Matias *et al.*, 2000, Tan *et al.*, 2015). The LBD contains a ligand binding pocket (LBP) (Figure 1.7C), formed by  $\alpha$ -helices H3, H5, H10, H11 and H12, the latter of which forms the core of ARs ligand-dependent activation function 2 (AF2) domain (Figure 1.6A) and acts as a 'lid' to close the LBP upon agonist binding (Matias *et al.*, 2000; Bevan *et al.*, 1999). DHT is the most potent natural AR agonist, which differs from testosterone by a single double bond on ring A (Figure 1.7D) (Kinter and Anekar, 2022). Agonist binding renders conformational changes in the LBD, exposing AF2 in a state amenable to interaction with LXXLL motif-containing coactivator proteins, as well as the NTD FQNLF motif, driving N/C terminal interactions (Figure 1.7D) (van Royen *et al.*, 2012; Schaufele *et al.*, 2005; He *et al.*, 2000; He *et al.*, 2004; Dubbink *et al.*, 2006; Bevan *et al.*, 1999; Askew *et al.*, 2012; Heery *et al.*, 1997; McInerney *et al.*, 1998).



**Figure 1.7 - Androgen receptor functional domain structures**

**A.** The DBD of AR shares considerable homology with the PR, MR and GR. AR's D-box contains unique residues S597 and T602 (highlighted by arrows) that mediate strong homodimerisation. The P-box directly interacts with nucleotides at hormone response elements in DNA (Figure adapted from Tan *et al.*, 2015) **B.** ARs NLS, encoded in the hinge region, fits into the concave inner surface of importin- $\alpha$  upon ligand binding (Figure taken from Tan *et al.*, 2015) **C.** Eleven  $\alpha$ -helices comprise the LBD of AR, which has a flexible linker in place of the helix H2 of other nuclear receptors.  $\alpha$ -helices H3, H5, H10, H11 and H12 form the ligand-binding pocket. H12 (highlighted in maroon) acts as a 'lid' to close the pocket upon ligand binding and forms the core of ARs AF2 domain (Figure adapted from Tan *et al.*, 2015) **D.** AR agonists testosterone and DHT differ by a single double bond. DHT binding (green) to the LBD mediates exposure of AF2 (blue) to LXXLL-motif containing coactivators, as well as the NTD FQNL motif, driving intra- and inter- AR monomer interactions (Figure adapted from Xu *et al.*, 2013)

### 1.2.2 Non-canonical AR signalling

The AR/DHT axis also has roles separate to canonical DNA-binding transcriptional regulation. Occurring independently of nuclear translocation, AR can rapidly activate kinases including Src, MAPK/ERK and PI3K/Akt (Migliaccio *et al.*, 2007; Peterziel *et al.*, 1999; Kang *et al.*, 2004; Baron *et al.*, 2004; Kousteni *et al.*, 2001). AR can also undergo plasma membrane localisation to interact with caveolins upon DHT binding (Deng *et al.*, 2017; Pedram *et al.*, 2007). Translocation to the plasma membrane can influence non-canonical activation of secondary messengers and can even potentiate canonical AR signalling at AREs by enhancing nuclear

translocation (Cinar *et al.*, 2007; Lu *et al.*, 2001; Li *et al.*, 2018). Furthermore, AR is capable of directly binding to and regulating the activity of other transcription factors such as AP-1, ETS and NF $\kappa$ B independently of DNA-binding (Kallio *et al.*, 1995; Schneikert *et al.*, 1996; Palvimo *et al.*, 1996).

Non-canonical roles of the AR have been reviewed extensively elsewhere, in-depth discussion of which is beyond the scope of this thesis (Liao *et al.*, 2013; Zamagni *et al.*, 2019; Leung and Sadar, 2017; Davey and Grossmann, 2016). Subsequent sections will primarily focus on canonical, nuclear AR function.

### **1.2.3 AR roles in prostate biology and benign disease**

AR signalling is crucial to prostate development and function. UGE/UGM expression of AR and 5 $\alpha$ -reductase is seen during early gestation, and the developmental fate of the UGS is dictated by embryonic week 8 as the foetal testes begin to produce testosterone (Majumder and Kumar, 1997; Levine *et al.*, 1996; Sajjad *et al.*, 2004). Siiteri and Wilson, 1974; Word *et al.*, 1989). Moreover, the developmental importance of androgens/AR signalling is highlighted by models where XY embryos lacking functional AR fail to develop a prostate, whilst XX UGS can undergo prostatic differentiation upon DHT treatment (Toivanen and Shen, 2017; Aaron *et al.*, 2016; Francis and Swain, 2018).

The prostate gland weighs just a few grams at birth, before significant expansion to approximately 20 grams by the end of puberty (Kumar and Majumder, 1995). This growth is attributable to augmented hypothalamus-anterior pituitary-testes (HAT) axis activity, triggered by pubertal increases in gonadotrophin-releasing hormone (GnRH) production (Abreu and Kaiser, 2016; Ebling, 2005). Concomitant increases in circulating testosterone occur, converted to DHT in the prostate to drive AR signalling (Figure 1.8). AR signalling in the pubescent prostate is demonstrated by rising prostate specific antigen (PSA, encoded by the *KLK3* gene) secretions in correlation with serum testosterone levels (Juul, Müller and Skakkebaek, 1997; Vieira *et al.*, 1994). PSA is a well-established marker of AR transactivation, as the *KLK3* gene contains functional AREs in its enhancer/promoter (Riegman *et al.*, 1991; Huang *et al.*, 1999; Cleutjens *et al.*, 1997).

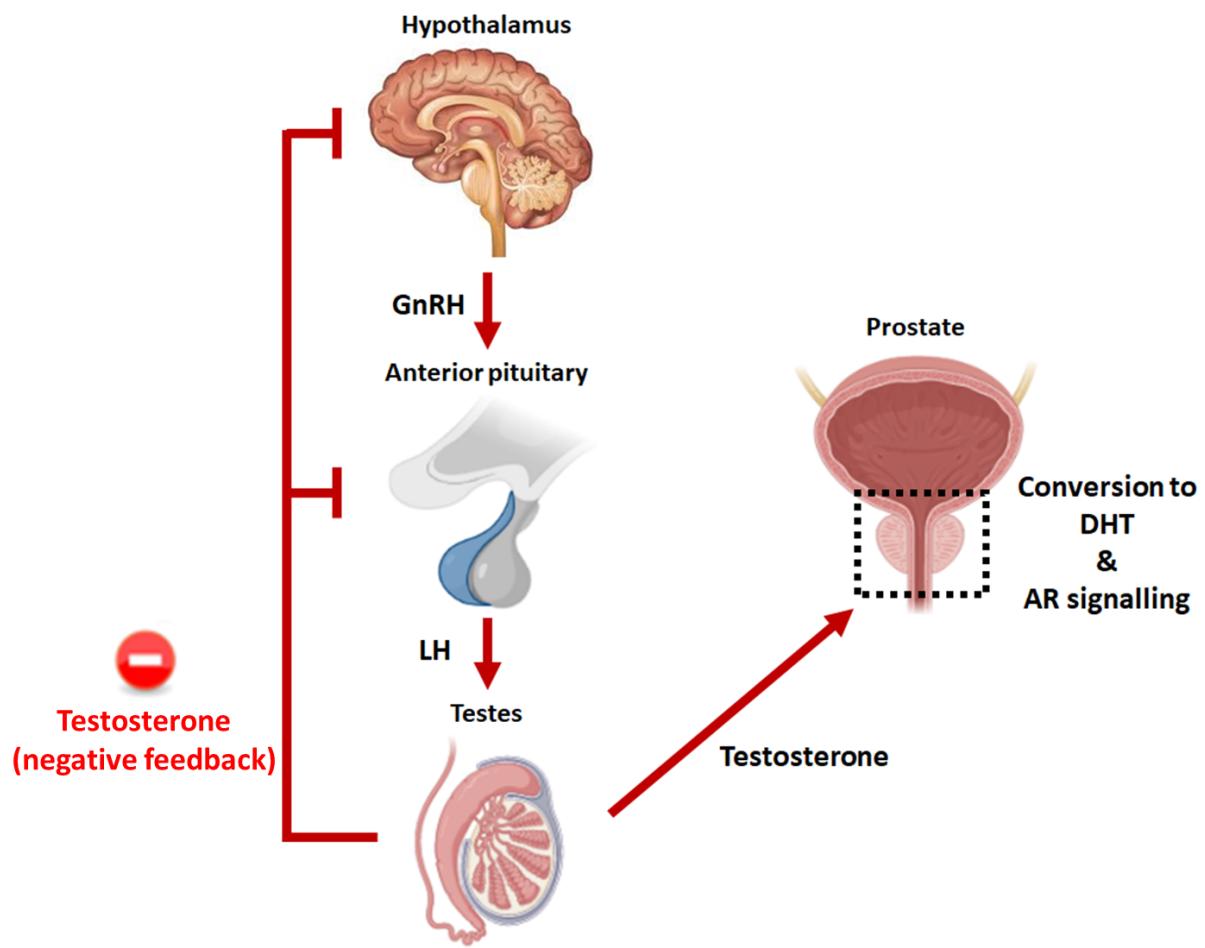


Figure 1.8 – Hormonal axis control of prostate AR signalling

Production of GnRH by the hypothalamus rises in puberty. This binds GnRH receptors of the anterior pituitary, stimulating release of luteinising hormone (LH), which in turn stimulates release of testosterone from Leydig cells of the testes. Prostate cells expressing 5 $\alpha$ -reductase convert this to the potent AR agonist DHT, driving AR signalling. A negative feedback loop (red) exists, whereby increases in testosterone suppress GnRH and/or LH production (Figure created using biorender.com)

Prostate knockout studies reveal a role for AR in different tissue compartments. Models of AR deletion in prostate fibroblasts underscore the importance of stromal signalling to prostate epithelial proliferation, development and secretory differentiation; mediated by paracrine action of growth factors *eg* FGF-10 and IGF-1 (Yu *et al.*, 2012; Ruan *et al.*, 1999; Donjacour, Thomson and Cunha, 2003; Singh *et al.*, 2014; Lai *et al.*, 2012; Cunha *et al.*, 1992). Similar perturbations are seen upon AR knockout in stromal smooth muscle (Welsh *et al.*, 2011; Yu *et al.*, 2011). However, restricting AR knockout to epithelial cells does not critically hinder overall growth and ductal development; rather the primary effect is failure to develop functionally mature luminal epithelia (Simanainen *et al.*, 2007, Wu *et al.*, 2007). This collectively demonstrates that AR function in mature prostate development and homeostasis

is complex and cell-type specific. Rodent castration models further highlight the importance of androgens in homeostasis of the adult prostate, where orchectomy results in apoptosis and tissue atrophy, an effect that is reversible by testosterone administration (Isaacs, 1984; Kyprianou and Isaacs, 1988; Sugimura *et al.*, 1986).

AR also plays a role in non-cancerous prostate pathologies, the most common of which is benign prostate hyperplasia (BPH), estimated to be present in roughly 10% of men in their 30s and upwards of 80% of those in their 70s and 80s (Roehrborn, 2005). BPH is a non-malignant proliferation of stromal and epithelial cells in the prostate transition zone (McNeal, 1990; Ng and Baradhi, 2022). The role of DHT in BPH is evidenced by management with 5 $\alpha$ -reductase inhibitors finasteride and dutasteride, both of which reduce prostate volume (Gormley *et al.*, 1992; Thomson, 2005).

Given that serum testosterone decreases with age, the apparent inverse correlation between circulating androgen and rates of BPH seems somewhat paradoxical (Xia *et al.*, 2021; Stanworth and Jones, 2008). However, studies show that intraprostatic DHT concentrations are remarkably stable in the face of changing serum testosterone, and that the two do not always correlate (Thirumalai *et al.*, 2016; Swerdloff *et al.*, 2017; Page *et al.*, 2006). Research demonstrates increased nuclear AR in the periurethral zone, the most common site of BPH (Monti *et al.*, 1998). Additionally, more intense AR staining and elevated AR target gene expression is observed in BPH tissues relative to normal prostate (Nicholson *et al.*, 2013; O'Malley *et al.*, 2009). Further evidence for AR involvement in BPH comes from studies using direct AR antagonists, which effectively reduce prostate volume (Stone and Clejan, 1991; Eri and Tveter, 1993; Narayan *et al.*, 1996). How AR shifts from maintaining normal prostate homeostasis to being a driver of BPH is poorly understood; although chronic prostate inflammation, metabolic syndrome, and/or tissue hypoxia may alter the microenvironmental context in which AR operates (Gandaglia *et al.*, 2013; Gacci *et al.*, 2015; Wu *et al.*, 2016; Kim *et al.*, 2013). Research to date thus highlights ARs role in healthy prostate development and differentiation, as well as disease.

## 1.3 The androgen receptor and prostate cancer

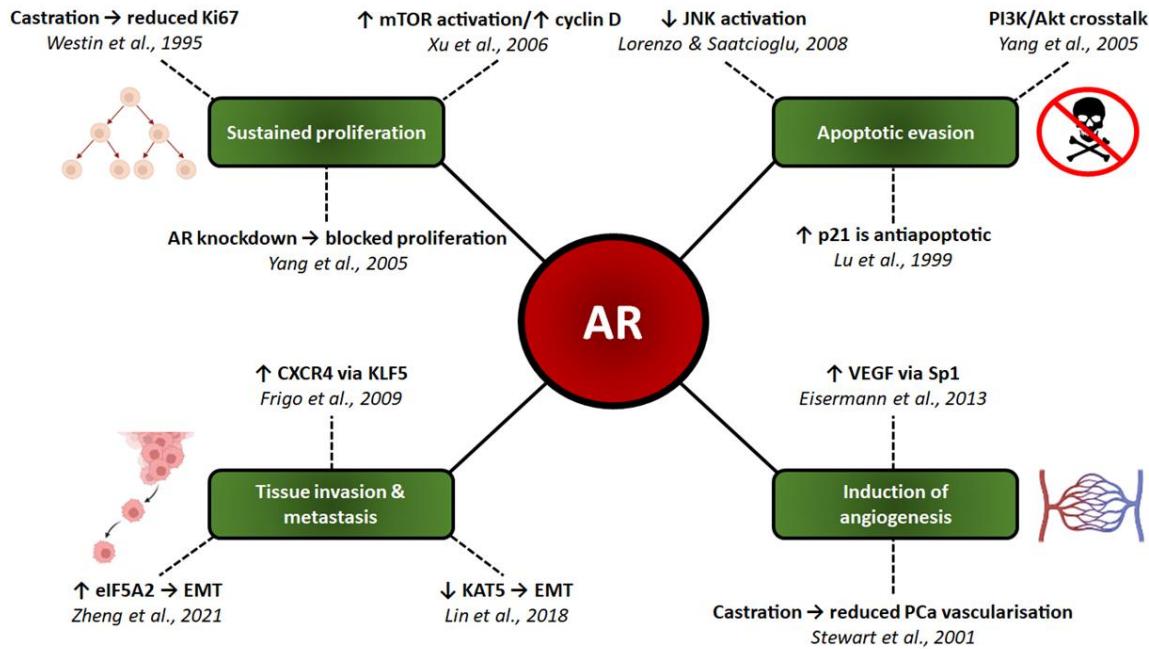
### 1.3.1 A molecular link between androgens and prostate cancer

AR is highly expressed in luminal epithelial cells of the adult prostate, where it suppresses proliferation and maintains terminal secretory differentiation (Wu *et al.*, 2007; Simanainen *et al.*, 2009; Whitacre *et al.*, 2002). The question thus remains how PCa progression is hormone-dependent when most cases have a luminal epithelial/AR+ phenotype.

Progression to cancer relies on additional molecular events outlined previously (Table 1.1) which cooperate with AR signalling. MYC oncogene overexpression is a frequently observed aberration, seen in a majority of PCa cases (Koh *et al.*, 2010). Intriguingly, *in vitro* experiments show that AR suppresses *c-MYC* transcription in normal prostate epithelium whilst in PCa the opposite occurs (Antony *et al.*, 2014; Vander Griend, Litvinov and Isaacs, 2014). *NKX3.1*, an AR target gene, is a suppressor of prostate epithelial proliferation and functions in a negative feedback loop to downregulate AR. This feedback is lost by deletion/inactivation of either *NKX3.1* or its positive regulator *PTEN*, both common in PCa, perturbing the balance of proliferation, survival and death in PCa cells (Bhatia-Gaur *et al.*, 1999; Cairns *et al.*, 1997; Thomas, Preece and Bentel, 2010; Lei *et al.*, 2006; Bowen *et al.*, 2019). Additionally, *NKX3.1* loss correlates with appearance and expression of oncogenic, AR-driven chromosomal translocations (Lin *et al.*, 2009; Bowen, Zheng and Gelmann, 2015; Thangapazham *et al.*, 2014). *TPRSS2*, encoding a serine protease found in seminal fluid, contains AR binding sites in its promoter and enhancer, thus enabling AR/DHT mediated overexpression of the *ERG* oncogene in the context of the frequently seen *TPRSS2:ERG* fusion (Clinckemalie *et al.*, 2013; Wang *et al.*, 2007; Sahu *et al.*, 2011; Adamo and Ladomery, 2016). Deletion of the locus encoding chromatin remodelling protein CHD1 occurs in ~15% of PCa cases and is implicated in modifying the AR transactivation landscape to a pro-tumourigenic transcriptome (Augello *et al.*, 2019). Furthermore, inactivating mutations in E3 ubiquitin ligase SPOP, which ordinarily targets AR for degradation, are seen in ~10% of primary PCa tumours (Geng *et al.*, 2014; An *et al.*, 2014).

Irrespective of specific molecular alterations mediating an AR-driven phenotype, PCa largely maintains addiction to its activity. Indeed, AR has been linked with many of the 'Hallmarks of

Cancer' in PCa, emphasizing its role in disease progression (Figure 1.9) (Hanahan and Weinberg, 2000/2011; Hanahan, 2022).



**Figure 1.9 - AR signalling and prostate 'Hallmarks of Cancer'**

Numerous lines of evidence link AR signalling to the 'Hallmarks of Cancer' in PCa (summary not exhaustive). Acquisition of molecular and genetic aberrations, outlined in Section 1.3.1, enables AR to switch from a mediator of healthy prostate homeostasis and function to a driver of PCa progression (Figure created using biorender.com)

### 1.3.2 Androgen deprivation therapy and antiandrogens: silver bullets?

Hormonal involvement in PCa has long been known. The pioneering work of Huggins and Hodges in the 1940s demonstrated that either orchietomy or injection of androgens/estrogens could alter metastatic disease burden (Huggins and Hodges, 1941; Huggins, Stevens and Hodges, 1941). The decades since have seen this approach evolve, and manipulation of androgens and/or direct inhibition of AR signalling is now the standard of care for mPCa (Crawford, 2004). This approach, termed hormone therapy, is offered to all men in the UK with stage IV metastatic disease (NICE, 2021). Two main approaches to hormone therapy exist: androgen deprivation therapy (ADT), or alteration of the hormonal signalling cascades and metabolic pathways that synthesise androgens, effectively starving AR of its ligand; and antiandrogens (also termed AR signalling inhibitors (ARSi) or AR targeted agents (ARTA)), direct antagonists of the AR protein which prevent nuclear translocation and DNA-binding activity, preventing AR transactivation (Figure 1.10).

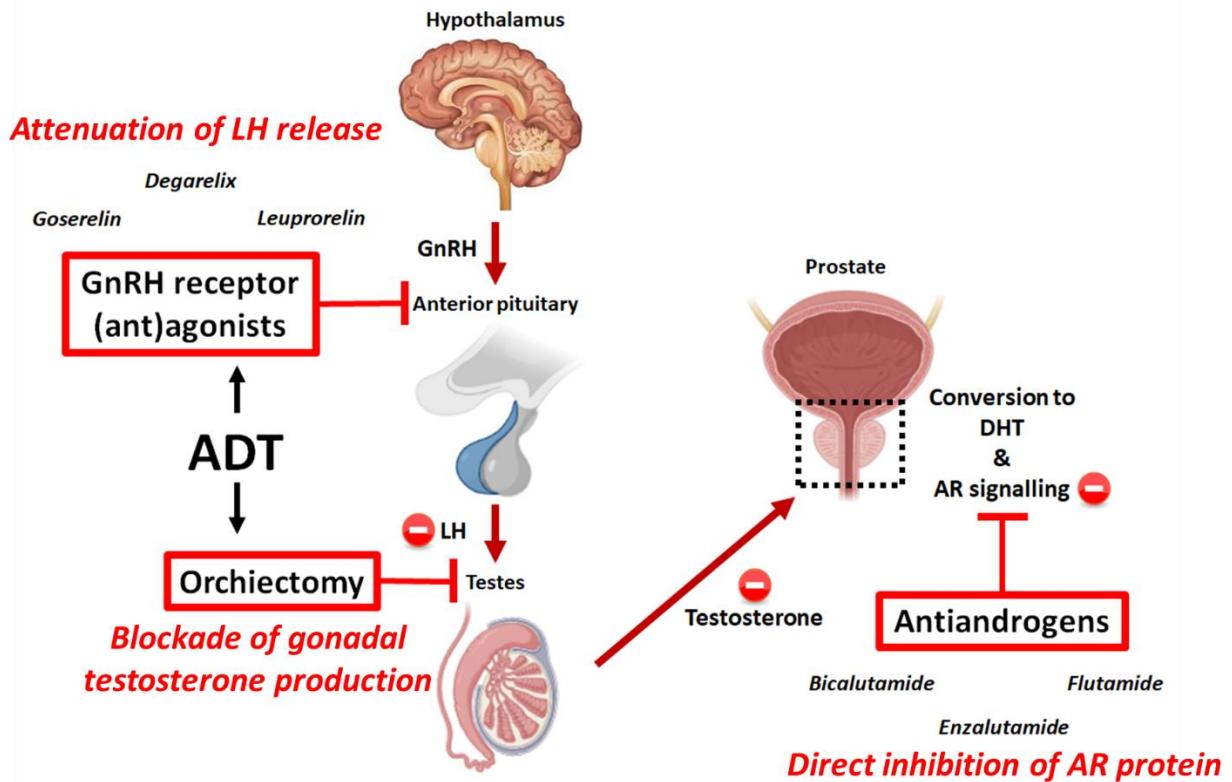


Figure 1.10 – Mechanisms of AR signalling inhibition by ADT and antiandrogens

ADT aims to starve PCa cells of testosterone and by extension, DHT. This was historically achieved by bilateral orchiectomy, replaced by chemical castration via GnRH receptor regulation. GnRH receptor agonists (goserelin, leuprorelin) overstimulate the GnRH receptor in the anterior pituitary. This may result in short-term exacerbation of symptoms by a ‘flare-effect’, however in time GnRH receptor is downregulated, blocking LH stimulation of testosterone release. GnRH receptor antagonists (degarelix) directly inhibit receptor activity instead. Antiandrogens (bicalutamide, flutamide, enzalutamide) are AR antagonists, which directly block AR nuclear translocation and transactivation capabilities in PCa cells. Primary mode of action for each treatment approach is summarised in red italic text (Figure created using biorender.com)

### **Androgen deprivation therapy (ADT)**

Normal circulating testosterone concentration ranges between ~260-920 ng/dL in healthy adult males (Travison *et al.*, 2017). ADT aims to reduce this to castrate levels, defined as < 50 ng/dL, with < 20 ng/dL maximizing therapeutic benefit (Gomella, 2009; Wang, Dai and Ye, 2017). This was historically achieved via bilateral subcapsular orchiectomy, which quickly and effectively ablates serum testosterone (Oefelein *et al.*, 2000; Novara *et al.*, 2009). However, due to the irreversible nature of surgery and psychological morbidities of testicular removal, chemical castration using systemically administered compounds is now preferred (Skoog *et al.*, 2011; Anderson *et al.*, 2008). This is achieved using GnRH receptor agonists and/or antagonists. GnRH receptor agonists (eg goserelin, leuprorelin) bind with high affinity in the anterior pituitary, thereby stimulating LH/testosterone production and, in the short term,

exacerbating disease. This initial increase in LH release, the ‘flare effect’, may worsen symptoms in metastatic disease (Waxman *et al.*, 1985; van Poppel and Nilsson, 2008). In time, GnRH receptor is downregulated, quashing LH release and reducing testosterone to castrate levels in ~3 weeks (Thompson, 2001; Heber *et al.*, 1982; Cooke and Sullivan, 1985). To avoid flare effects and enable immediate suppression, GnRH receptor antagonists (*eg* degarelix) may be used as an alternative. These newer compounds directly inhibit the receptor, outcompeting GnRH to prevent activity (Jiang *et al.*, 2001; Samant *et al.*, 2005). Thus, receptor antagonists rapidly achieve chemical castration within 2-3 days and do not elicit flare effects (Klotz *et al.*, 2008; Shore, 2013). Both compound classes attenuate testosterone production, and decisions regarding choice of agonist/antagonist largely depend on clinical presentation at time of therapy. Cost-effectiveness analysis by NICE determined that due to the higher price of degarelix, its use in mPCa over GnRH agonists is only warranted in patients suffering from spinal metastases/spinal cord compression, for whom an initial testosterone flare is of greater concern (Uttley *et al.*, 2017).

### ***Antiandrogens***

Antiandrogens act directly within PCa cells, preventing canonical AR signalling. These consist of two classes, steroidal (SAA) and non-steroidal (NSAA) antiandrogens, differing in structure whereby the former have a steroidal four-ringed composition (Figure 1.11A). Cyproterone acetate, a derivative of progesterone, is the only SAA currently used in mPCa on the NHS (NICE, 2022). Mechanistically, it competes with testosterone/DHT to bind the AR LBD, preventing nuclear translocation and ARE binding (Brinkmann *et al.*, 1983; Bohl *et al.*, 2007). It also exerts negative feedback on LH release (Schröder, 1993). SAAs are non-selective and capable of binding to GR, MR and PR. Consequently, they have greater likelihood of side effects including impotence, osteoporosis and potentially fatal hepato- and cardiotoxicities (Wirth, Hakenberg and Froehner, 2007; Fourcade and McLeod, 2004). Furthermore, cyproterone acetate can act as a partial AR agonist to stimulate AR transactivation (Poyet and Labrie, 1985; Labrie, 1993). It is mainly indicated for use with GnRH receptor agonists to neutralise flare effects (Sugiono *et al.*, 2005), and NSAAs are now largely preferred.

Numerous NSAAs have been developed including flutamide, bicalutamide and enzalutamide, all created on the same principle of selective, potent AR antagonism to compete with native DHT-AR interactions (Figure 1.11A, B). Flutamide was the first available NSAA, and was shown

in trials to improve survival for treatment-naïve mPCa patients in combination with leuprorelin vs the GnRH agonist alone (Crawford *et al.*, 1989; Goldspiel and Kohler, 1990). CYP1A2 metabolises flutamide to the more potent 2-hydroxyflutamide, which has a higher binding affinity for AR (Shet *et al.*, 1997; Wakeling *et al.*, 1981; Peets, Faye and Neri, 1974). Initially, 2-hydroxyflutamide was classed as a ‘pure antiandrogen’ which binds AR and promotes nuclear translocation, yet prevents gene transactivation (Kemppainen, Lane and Sar, 1992; Labrie, 1993). However *in vitro* characterisation revealed that 2-hydroxyflutamide can behave as an AR agonist at higher concentrations, contradicting its status as a silent antagonist (Wong *et al.*, 1995; Nguyen, Yao and Pike, 2007). Higher potency antagonists were subsequently developed with superior efficacy, pharmacokinetics and tolerability, one of which, bicalutamide, has largely replaced flutamide (Sarosdy, 1999; Schellhammer *et al.*, 1995). Bicalutamide exhibits up to 4-fold higher AR binding affinity than 2-hydroxyflutamide (Kolvenbag, Furr and Blackledge, 1998). Mechanistic studies suggest it promotes AR nuclear translocation and that the bicalutamide-ligated receptor binds DNA but fails to effectively undergo N-C terminal interactions or recruit coactivators (Masiello *et al.*, 2002; Farla *et al.*, 2005).

Development of novel NSAs continued, leading to synthesis of second-generation antiandrogens that include enzalutamide, apalutamide and darolutamide (Figure 1.11B). Enzalutamide (or MDV3100), developed in 2006, was the first such compound, proving superior to older NSAs in all regards. Firstly, it has far higher AR binding affinity, 5-8 fold higher than bicalutamide and just 2-3 fold lower than ARs native ligand DHT. Secondly, enzalutamide blocks AR nuclear translocation, DNA-binding and coactivator recruitment, thereby inhibiting AR signalling at numerous stages and preventing any partial agonist activity. Finally, trials have unequivocally demonstrated the benefits of enzalutamide over bicalutamide using a range of clinical outcome measures and disease settings (Tran, Ouk *et al.*, 2009; Shore *et al.*, 2016; Penson *et al.*, 2022; Penson *et al.*, 2016; Siemens *et al.*, 2018; Vaishampayan *et al.*, 2021). Enzalutamide has thus recently been approved by NICE as a first-line therapy in combination with ADT for mPCa (NICE, 2021). Apalutamide and darolutamide share a similar mechanism of action to enzalutamide, however they offer greater antiandrogenic potency, pharmacokinetic advantages or the ability to inhibit AR mutants (the latter is explored in greater depth in Section 1.4.2) (Higano, 2019).

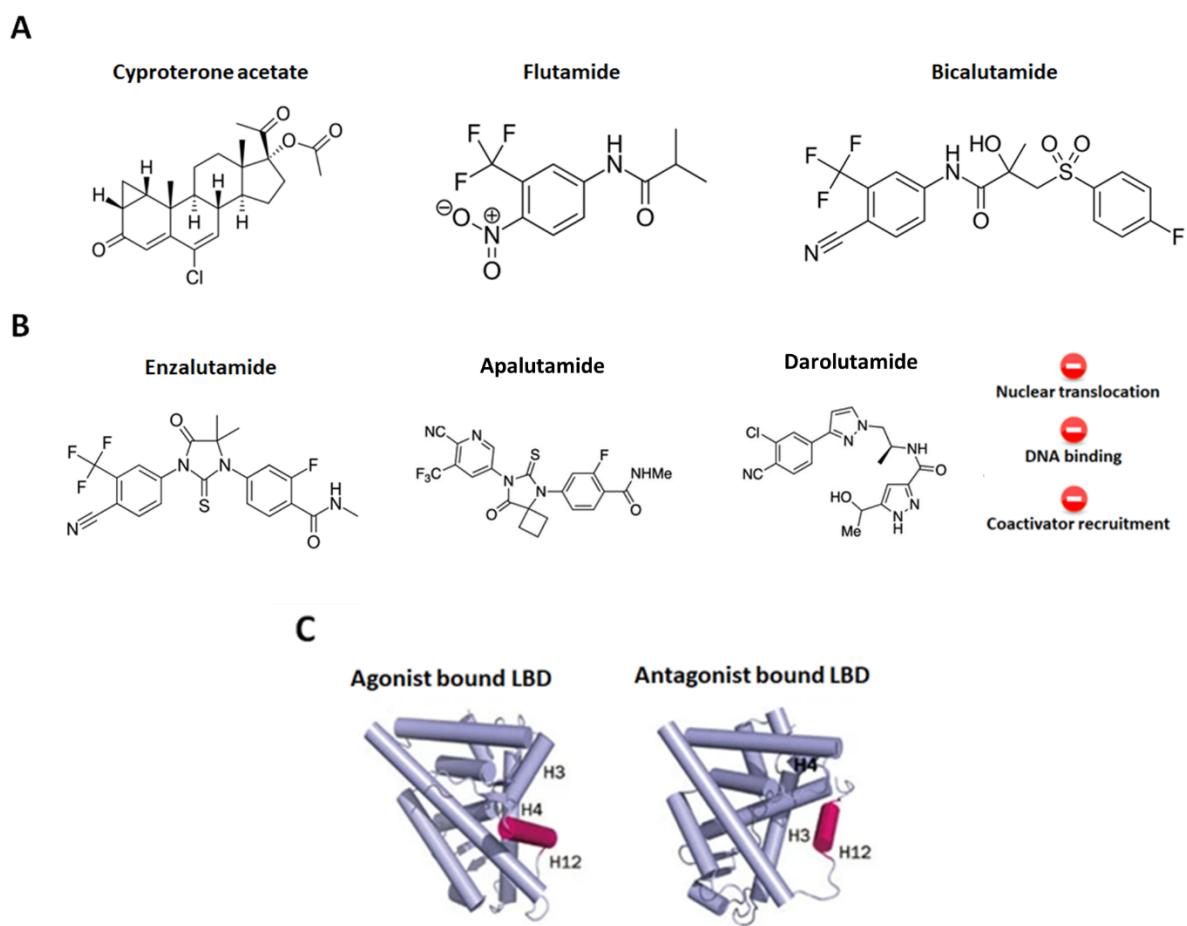


Figure 1.11 - Antiandrogen structure and NSAA modulation of AR LBD conformation

**A.** Chemical structure of steroidal antiandrogen cyproterone acetate, exhibiting a characteristic four-ring steroid configuration, and first generation non-steroidal antiandrogens (NSAAs) flutamide and bicalutamide **B.** Chemical structure of second generation NSAAs enzalutamide, apalutamide and darolutamide which inhibit AR signalling at multiple stages as listed **C.** Structural schematic of ARs LBD when agonist or antagonist bound. DHT binding repositions H12 (maroon) to close the ligand binding pocket and adopt a conformation with H3 and H4 comprising the AF2 domain. Conversely, NSAAs are believed to mediate an alternative repositioning of H12, disrupting AF2 structure and preventing coactivator recruitment and N/C terminal interactions (Figures adapted from Tan *et al.*, 2015).

The crystal structure of AR bound to different NSAs is undetermined, potentially due to NSAAs-bound AR failing to dissociate from chaperone proteins during purification (Bohl *et al.*, 2005; Tan *et al.*, 2015). Work using mutant AR, molecular dynamics simulations and/or inferences from other antagonist-bound nuclear hormone receptors suggest that NSAAs binding at ARs LBD destabilises normal LBP closure and AF2 surface formation by a repositioning of H12, inhibiting both recruitment of coactivators and receptor homodimerisation (Figure 1.11C) (Bohl *et al.*, 2005; Osguthorpe and Hagler, 2011; Sakkiah *et al.*, 2018; Sakkiah *et al.*, 2016; Gim *et al.*, 2021; Brzozowski *et al.*, 1997; Nadal *et al.*, 2017).

Availability of ADT and NSAAAs has led to the concept of maximal androgen blockade (MAB), in which ADT and NSAAAs are used in combination (Labrie *et al.*, 1982; Chodak, 2005). The early evidence supporting such a regime was mixed. MAB trials performed with first generation NSAAAs flutamide or nilutamide provided only a modest survival benefit with greater toxicity, thus these compounds are now mainly combined with ADT in the short-term context of flare effect prevention (Prostate Cancer Trialists Collaborative Group, 2001; Schmitt *et al.*, 1999; Samson *et al.*, 2002; Thompson, 2001). MAB performed with bicalutamide improves survival benefit and reduces toxicity vs other first generation NSAAAs, and using second generation compound enzalutamide further enhances patient benefit and time to mPCa progression (Akaza *et al.*, 2009; Klotz and Schellhammer, 2005; Ramsey *et al.*, 2005; Davis *et al.*, 2019; Armstrong *et al.*, 2019; Vaishampayan *et al.*, 2021). Irrespective of modality, patient benefit from AR signalling inhibition is often finite. A significant number of patients will relapse and progress to a state in which mPCa continues to progress despite continued ADT and/or NSAA usage. This disease state, termed castrate resistant prostate cancer, forms the focus of subsequent sections.

## 1.4 Castrate resistant prostate cancer

### 1.4.1 The clinical challenge of castrate resistant prostate cancer

Castrate resistant prostate cancer (CRPC, or mCRPC when metastatic) arises in a substantial number of patients treated with ADT and/or NSAAAs. Estimates vary considerably and suggest ~10-55% of castrate-sensitive mPCa cases (mCSPC) will develop CRPC in an average timeframe of 1-5 years post-therapy. (Hirst, Cabrera and Kirby, 2012; Kirby, Hirst and Crawford, 2011; Sweeney *et al.*, 2015; Svensson *et al.*, 2021; Wenzel *et al.*, 2021; Miyake *et al.*, 2019; Okamoto *et al.*, 2020; Lin *et al.*, 2019; Tamada *et al.*, 2018). CRPC is defined as, in a patient with castrate levels of serum testosterone, either: i) a greater than 25% rise in PSA levels over two or more consecutive tests taken over a week apart (biochemical progression), or ii) radiographical evidence of disease advancement (radiological progression) (Morote *et al.*, 2022). Historically, chemotherapy with docetaxel has been used to prolong survival, however disease inevitably progresses with a median survival of 1-3 years post-mCRPC diagnosis (Figure 1.12) (Tannock *et al.*, 2004; Petrylak *et al.*, 2004; James *et al.*, 2016; NICE, 2006; Moreira *et al.*, 2017; Ryan *et al.*, 2015; Armstrong *et al.*, 2020).

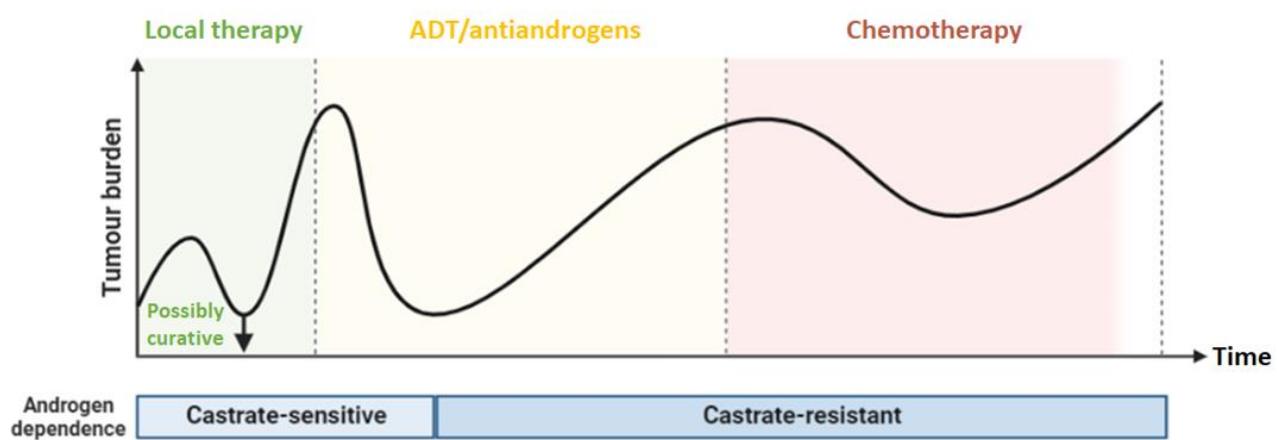


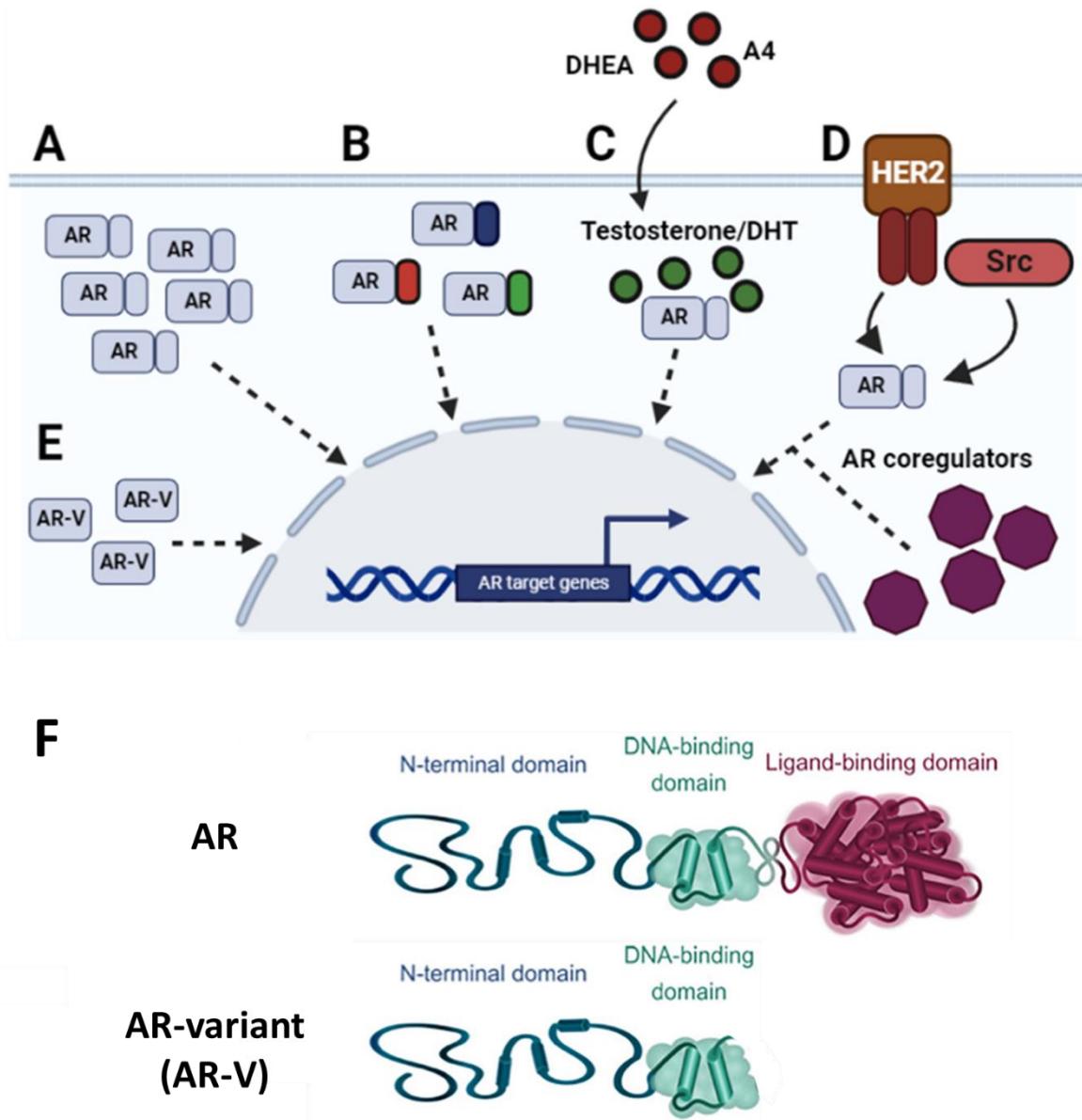
Figure 1.12 - Typical PCa progression to castrate resistance

Localised PCa may be treated curatively by surgery and/or radiotherapy, this may be supplemented with ADT. Men with recurrent or metastatic disease are universally offered ADT, which is initially highly effective alone or in combination with antiandrogens. However, significant numbers of patients transition from CSPC to CRPC which grows despite ablation of circulating testosterone. Historical use of docetaxel in this setting has provided some survival benefit, though CRPC inevitably progresses (Figure created using biorender.com).

### 1.4.2 AR-mediated mechanisms of castrate resistance and therapeutic strategies

Numerous lines of evidence show that despite castrate conditions, AR signalling can still be activated to drive CRPC. Indications of this first came from studies demonstrating post-ADT AR amplification, and that AR mutations are associated with progression to androgen independence (Visakorpi *et al.*, 1995; Gregory *et al.*, 2001; Tilley *et al.*, 1996; Taplin *et al.*, 1995). Furthermore, work performed using PCa cell lines and xenografts revealed similar overexpression and mutations to occur in model systems (Chen *et al.*, 2004; Veldscholte *et al.*, 1990; Zhao *et al.*, 1999). The relevance of AR to CRPC has since been confirmed and a variety of mechanisms are characterised, whilst the continued use of PSA as a biochemical marker of CRPC highlights ARs frequent role in the disease (Morote *et al.*, 2022).

Other processes exist by which CRPC evolves total independence from AR, such as neuroendocrine differentiation (Parimi *et al.*, 2014). These are beyond the remit of this work and have been reviewed extensively elsewhere (Crona and Whang, 2017; Hoang *et al.*, 2016; Sahin *et al.*, 2018). Here, mechanisms of CRPC resulting in continued AR signalling are the primary focus (Figure 1.13).



**Figure 1.13 - Mechanisms enabling sustained AR signalling in CRPC - including AR-Vs, the focus of this work**

**A**: Amplification or overexpression of AR can sensitise PCa cells to residual levels of androgen, enabling continued AR transactivation in the context of ADT.

**B**: Point mutations in the ligand binding domain can elicit structural changes to AR that prevent NSAA inhibition, and can even cause NSAAAs to function as AR agonists.

**C**: CRPC can upregulate expression of steroidogenic enzymes enabling synthesis of testosterone, and subsequently DHT, from adrenal androgens *eg* DHEA and A4.

**D**: AR can be activated by phosphorylation through tyrosine kinases including HER2 and Src, circumventing the need for androgen binding. Alterations to coregulator expression may also potentiate AR transactivation by these alternative means.

**E**: Truncated forms of the androgen receptor lacking a ligand binding domain, termed androgen receptor variants (AR-Vs), are generated by alternative splicing. These can mediate AR target gene transcription without ligand stimulation.

**F**: Schematic depicting protein structure comparison between AR, which contains a full ligand-binding domain, and the general structure of an AR-variant that lacks this domain.

### ***AR amplification/overexpression***

Gene amplification or overexpression was one of the earliest characterised AR aberrations linked with CRPC. Visakorpi *et al.* observed that 30% of tumours recurring after ADT had high-level *AR* amplifications, none of which were seen in the same patients prior to therapy (Visakorpi *et al.*, 1995). Amplification is also shown to correlate with increased *AR* mRNA, as demonstrated by research illustrating castrate-resistant tumours to exhibit higher *AR* expression than androgen-dependent PCa (Linja *et al.*, 2001; Latil *et al.*, 2001; Koivisto *et al.*, 1997). This trend was also noted at the protein level using immunohistochemistry of matched castrate sensitive/resistant tissue of the same patient (Edwards *et al.*, 2003). Furthermore, such a mechanism has been validated by *in vitro* and *in vivo* experimental evidence using androgen deprived models, which were adapted to grow in such conditions and overexpress *AR* through multiple mechanisms including alterations to the epigenetic landscape via LSD1 (Waltering *et al.*, 2009; Kokontis *et al.*, 1994; Chen *et al.*, 2004; Sirotnak *et al.*, 2004; Cai *et al.*, 2011).

There is debate as to the relative contributions of *AR* gene amplification vs overexpression by other means (Karantanos, Corn and Thompson, 2013; Jernberg, Bergh and Wikström, 2017). Regardless of underlying mechanism, it is hypothesised that increased receptor levels sensitise CRPC to low levels of androgen (Visakorpi *et al.*, 1995; Waltering *et al.*, 2009; Chen *et al.*, 2004). The clinical relevance of this has been validated by work demonstrating *AR* amplification in tumour biopsies that progress on enzalutamide (McKay *et al.*, 2021). Additionally, ‘liquid biopsies’ of circulating tumour DNA (ctDNA) have found great utility as a non-invasive assessment of *AR* amplification, further linking this genomic aberration with therapy resistance (Wyatt *et al.*, 2016; Conteduca *et al.*, 2017; Del Re *et al.*, 2021). Theoretically, development of increasingly potent NSAAs could circumvent this mechanism by preventing action of all available cellular *AR*, given ADT appears insufficient to prevent amplification/overexpression mediated *AR* signalling (Figure 1.13A).

### ***AR point mutations***

Point mutations in the LBD are also causally linked with sustained *AR* signalling in CRPC, where they are estimated to occur in ~20% of CRPCs mutually exclusively with *AR* amplification (Beltran *et al.*, 2013). The first published *AR* mutation was a threonine to alanine substitution,

T877A, identified in the LNCaP cell line derived from a PCa lymph node metastasis and demonstrated to render the NSAA flutamide an AR agonist (Veldscholte *et al.*, 1990; Veldscholte *et al.*, 1992; Liu *et al.*, 2017). Solving the crystal structure of AR<sub>T877A</sub> revealed it maintains the ability to bind androgens with high structural similarity to wild-type AR, however the introduction of an alanine elicits subtle structural alterations to the LBD purported to enable greater binding promiscuity (Sack *et al.*, 2001). The binding affinities of position 877 substitutions vary by amino acid, as unlike T877A the introduction of aspartic acid, lysine or tyrosine eliminates androgen binding capability (Ris-Stalpers *et al.*, 1993; Steketee *et al.*, 2002). Subsequent work in clinical samples highlighted an identical mutation in NSAA-resistant PCa biopsies (Suzuki *et al.*, 1993; Gaddipati *et al.*, 1994; Taplin *et al.*, 1999).

Numerous clinically relevant AR mutations have since been identified. W741C/L, has, whilst rare, been observed in relapsed bicalutamide treated patients (Haapala *et al.*, 2001). AR<sub>W741C/L</sub> causes bicalutamide to act as an AR agonist, yet is still antagonised by the less potent NSAA flutamide (Hara *et al.*, 2003; Bohl *et al.*, 2005; Liu *et al.*, 2016). KUCaP, an androgen-dependent PCa xenograft model carrying AR<sub>W741C/L</sub>, even accelerated in growth when treated with bicalutamide (Yoshida *et al.*, 2005). Moreover, work performed in the LNCaP cell line demonstrated a hyper-mutable *AR* gene state induced by bicalutamide treatment, highlighting the selective pressures applied to NSAA treated cells that may provide an ideal environment for mutant selection (Hara *et al.*, 2005).

Further mutations have since been catalogued that impact efficacy of contemporary antiandrogens. F876L confers both *in vitro* and *in vivo* resistance to enzalutamide and apalutamide and is detectable in the plasma DNA of patients treated with the latter compound (Joseph *et al.*, 2013; Korpal *et al.*, 2013). Additionally, AR<sub>F876L</sub> can co-occur with another mutation T878A, which combine to convert enzalutamide into a strong AR agonist (Prekovic *et al.*, 2016). This particular mode of therapy evasion may be overcome by newer antiandrogens capable of targeting common mutations. Darolutamide, approved for use by the NHS in 2020, is efficacious against many currently known *AR* mutations and proves efficacious against CRPC with minimal toxicity (Figure 1.13B). (Lallous *et al.*, 2021; Fizazi *et al.*, 2019)

### ***Non-androgenic transactivation***

Mechanisms occur in which AR transactivation is independent of canonical receptor/DHT binding. This is referred to in the literature by various names including 'outlaw AR', 'signalling cross-talk' or 'ligand independent transactivation' (Chandrasekar *et al.*, 2015; Schalken and Fitzpatrick, 2016; Imamura and Sadar, 2016). Additionally, expression changes in AR coregulatory proteins have been observed (Culig, 2016). For this review, these processes are collectively described as 'non-androgenic transactivation', referring to AR transactivation occurring outside of classical DHT/receptor binding at physiological levels of androgen.

Tyrosine kinase signalling is implicated in AR transactivation (Figure 1.13C). *In vitro* cell line models demonstrate that receptor tyrosine kinase HER2 can drive AR target gene transcription (Yeh *et al.*, 1999; Mellinghoff *et al.*, 2004; Gregory *et al.*, 2005). This occurs by AR phosphorylation at S515 via MAPK (Ponguta *et al.*, 2008; Yeh *et al.*, 1999). Work performed in PCa xenografts illustrates augmented HER2 expression in ADT-resistant tumours, whilst retroviral HER2 overexpression enabled androgen-dependent populations to grow in castrate conditions (Craft *et al.*, 1999). Furthermore, HER2 expression levels negatively correlate with clinical outcomes in CRPC (Maillet *et al.*, 2021; Edwards *et al.*, 2006). CRPC models also demonstrate enhancement of AR transactivation by intracellular kinase Src which phosphorylates AR at Y534, a modification elevated in ADT-relapsed patient samples (Asim *et al.*, 2008; Guo *et al.*, 2006). Additionally, an siRNA screen showed knockdown of Src inhibitory kinase CSK to phenocopy treatment of androgen depleted cells with synthetic androgen R1881 (Yang *et al.*, 2015). Unfortunately, inhibition of these kinase pathways has yet to garner clinical benefit in CRPC (Ziada *et al.*, 2004; Lara *et al.*, 2009; Araujo *et al.*, 2013). Deeper preclinical evidence and a more thorough mechanistic understanding may help identify which patients are most likely to benefit.

AR coactivator proteins may also potentiate aberrant receptor activity. Histone acetyltransferase p300, involved in remodelling of the chromatin landscape at AREs and elevated in CRPC, enhances DHT-independent transactivation by IL-6 (Welti, Sharp and Brooks *et al.*, 2021; Debes *et al.*, 2002; Debes *et al.*, 2005). Moreover, increased expression of p160 family members TIF2, SRC-1 and SRC-3 is associated with CRPC and abnormal AR transcriptional activity (Agoulnik *et al.*, 2006; Ueda *et al.*, 2002; Zou *et al.*, 2006; Tien *et al.*, 2013; Coutinho *et al.*, 2016). Conversely, corepressors *eg* NCOR1 and SMRT mediate a closed,

transcriptionally repressive histone microenvironment, are downregulated in CRPC and linked with bicalutamide resistance (Lopez *et al.*, 2016; Godoy *et al.*, 2012). Therefore, alterations to AR coregulators could modify chromatin status and transcriptional machinery recruitment at AREs, enabling continued AR transactivation in suboptimal conditions (Figure 1.13C). p300 inhibitors for CRPC have consequently entered clinical trials (NCT05488548).

### ***Alternative steroid metabolism***

ADT targets the HAT axis, ultimately blocking gonadal testosterone production. CRPC can thwart this by utilising alternative sources of androgen to fuel AR. The adrenal glands are a prominent source producing dehydroepiandrosterone (DHEA), DHEA sulphate (DHEAS) and androstenedione (A4), all of which can be enzymatically converted to testosterone and DHT (Nassar and Leslie, 2022; Labrie *et al.*, 2001). This led to historical use of adrenalectomy in treating relapsed PCa (Harrison, Thorn and Jenkins, 1953; Bowers, 1962; Mahoney and Harrison, 1972; Bhanalaph, Varkarakis and Murphy, 1974). Suppression of adrenal androgens is now performed pharmacologically. The first attempts to do this utilised ketoconazole, an imidazole used to treat fungal infections (Trachtenberg, Halpern and Pont, 1983; Trachtenberg and Pont, 1984). Ketoconazole is an inhibitor of both cholesterol side-chain cleavage enzyme (P450scc), which converts cholesterol to pregnenolone, and cytochrome P450 17 $\alpha$ -hydroxylase/17,20-lyase (CYP17A1), responsible for various enzymatic conversions of testosterone precursors including 17 $\alpha$ -hydroxypregnenolone to DHEA and 17 $\alpha$ -hydroxyprogesterone to A4 (Santen *et al.* 1983, Vasaitis, Bruno and Njar, 2011). Ketoconazole has proven efficacious in CRPC by numerous trials (Small, Baron and Bok, 1997; Small *et al.*, 1997; Small *et al.*, 2004). However, therapeutic doses are usually high resulting in potentially fatal hepatotoxicity, and due to wide ranging perturbations to steroid metabolism and possible adrenal insufficiency it is necessary to supplement its use with corticosteroids, commonly dexamethasone (Paul and Breul, 2000; Sinawe and Casadesus, 2022; Vasaitis, Bruno and Njar, 2011; De Coster *et al.*, 1987).

Subsequent interest in nullifying extragonadal androgen production resulted in development of abiraterone acetate (AA) (Barrie *et al.*, 1994; Potter *et al.*, 1995). AA, specifically its active metabolite abiraterone, potently and selectively inhibits CYP17A1, effectively suppressing formation of DHEA and A4 with fewer effects on wider corticosteroid metabolism (Schweizer and Antonorakis, 2012; Bryce and Ryan, 2012). Consequently, studies comparing AA with

ketoconazole demonstrated superior PSA reduction and survival benefit with significantly fewer toxicities using the former compound (Ryan *et al.*, 2010; Peer *et al.*, 2014; Kim *et al.*, 2014). When coadministered with low dose prednisone to combat the side effect of cortisol reduction, AA has proven effective in numerous trials (Fizazi *et al.*, 2012; de Bono *et al.*, 2011; Ryan *et al.*, 2015; Auchus *et al.*, 2014). As a result, abiraterone was approved on the NHS as first-line therapy for mCRPC without docetaxel in 2016 (NICE, 2016).

The rationale behind these approaches is that enzymes responsible for conversion of adrenal androgens to testosterone, including CYP17A1, AKR1C3, SRD5A1 and UGT2B15, are upregulated in CRPC metastases (Stanbrough *et al.*, 2006; Montgomery *et al.*, 2008). However, alterations to enzymes involved in *de novo* androgen synthesis from cholesterol increase in response to abiraterone, suggesting that CRPC is capable of further enhancing steroidogenic pathways as a means to quash AA efficacy (Locke *et al.*, 2008; Cai *et al.*, 2011; Mostaghel *et al.*, 2019) (Figure 1.13D). Nevertheless even if androgen biosynthesis is completely suppressed, alternative mechanisms of resistance enable continued AR signalling in the absence of ligand. These include the primary focus of this work, androgen receptor variants.

### ***Androgen receptor variants***

For more than two decades the existence of short, truncated forms of AR has been observed (Wilson and McPhaul, 1996; Gregory, He and Wilson, 2001). Subsequent work has identified many of these as alternatively spliced forms of the androgen receptor lacking a ligand binding domain, termed androgen receptor variants. Androgen receptor variants are capable of driving AR target gene transcription in the absence of androgen and are now recognised as another mechanism driving CRPC (Wadosky and Koochekpour, 2007) (Figure 1.13E, F). Before these androgen receptor variants are explored in-depth, an appreciation for the biology and processes governing alternative splicing of mRNA is warranted and forms the focus of the next section.

## 1.5 Alternative splicing

### 1.5.1 Alternative splicing: processes and regulation

The vast majority of eukaryotic genes consist of multiple exons forming the coding sequence (CDS), which undergoes mRNA translation into amino acid and protein products (Jorquera *et al.*, 2016). Exons are interspersed with longer introns, transcribed together as pre-mRNA before processing occurs to splice out introns, join exons, add a 5' methylguanosine cap and 3' poly(A) tail to create mature mRNA (Wang and Farhana, 2022). In 1977, it was demonstrated that a single adenoviral pre-mRNA could be spliced into multiple different mature mRNA products, a process that was then revealed in cellular mRNA transcribed from the *IgM* gene (Berget *et al.*, 1977; Early *et al.*, 1980). This has subsequently been described as alternative splicing, which can occur in numerous ways and is estimated to occur in 95% of human multiexon genes, vastly increasing proteome diversity to ~70,000 proteins from ~20,000 genes (Pan *et al.*, 2008; Aebersold *et al.*, 2018) (Figure 1.14).

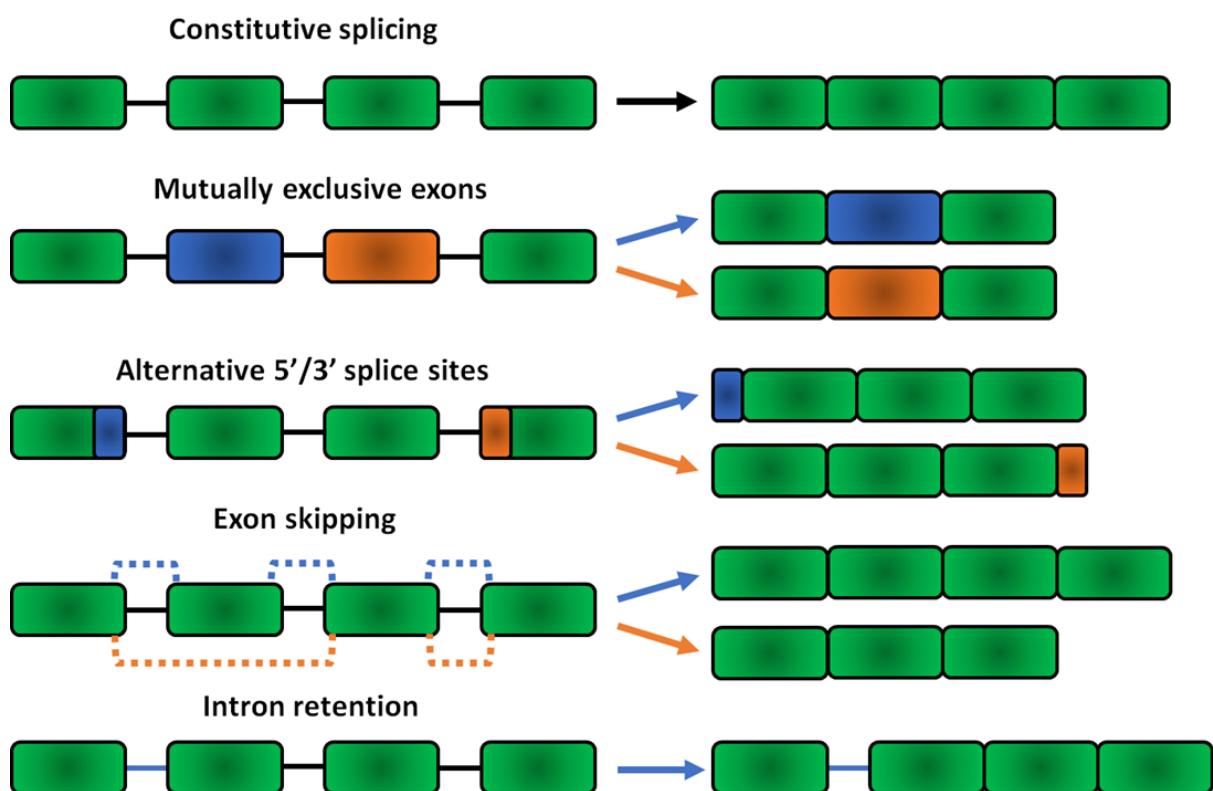


Figure 1.14 - Alternative splicing patterns

Numerous types of alternative splicing have been identified, the most common of which are illustrated above. Exons (green boxes), separated by longer introns (black lines, length not to scale), can be joined in numerous configurations to form multiple mature mRNA products. Constitutive splicing refers to a linear configuration in

which full exons are joined sequentially from their 5' to 3' genomic order. Alternative splicing enables exons to be included in a mutually exclusive manner, alternative 5' and 3' sites within exons may be utilised, exons can be skipped altogether, or introns may be retained.

Splicing is performed by the spliceosome, a dynamic assembly of small nuclear ribonucleoprotein complexes (snRNPs). pre-mRNA splice sites are demarcated by a conserved GU at the 5' splice-site (5' SS), an AG at the 3' splice-site (3' SS) which is preceded by a polypyrimidine tract (poly(Y)) of variable length, and a branch point (BP) adenosine located on average 25 bases upstream of the 3' SS (Will and Lührmann, 2011; Taggart *et al.*, 2012; Sickmier *et al.*, 2006).

The core snRNPs, numbered U1 to U6, sequentially associate with splice junctions to mediate a series of processes: first, U1 recognises and binds the 5' SS whilst splicing factors SF1 and U2AF65 aid recruitment of U2 at the BP, forming the prespliceosome/A complex (Figure 1.15A); next, the U4/U6.U5 tri-snRNP is recruited along with the NineTeen Complex (NTC) to create the pre-catalytic spliceosome/B complex before U1 and U4 dissociation and structural rearrangements result in the catalytically active B\* complex (Figure 1.15B); the B\* complex then catalyses a transesterification reaction in which the 2' OH of the BP adenosine performs a nucleophilic attack on the 5' SS, generating an intron/3' exon lariat known as the C complex (Figure 1.15C); finally, a second transesterification takes place involving nucleophilic attack of the phosphodiester bond at the 3' SS by the 5' exon, before the remaining intron lariat is degraded and snRNPs are recycled for further splicing reactions (Figure 1.15D) (Wilkinson, Charenton and Nagai, 2020; Will and Lührmann, 2011). Other proteins prominently involved in the energy-dependent RNA/protein rearrangements during splicing are RNA helicases with ATP/GTPase activity including DEAD-box and DEAH-box family members (Figure 1.15, blue boxes) (Cordin and Beggs, 2013).

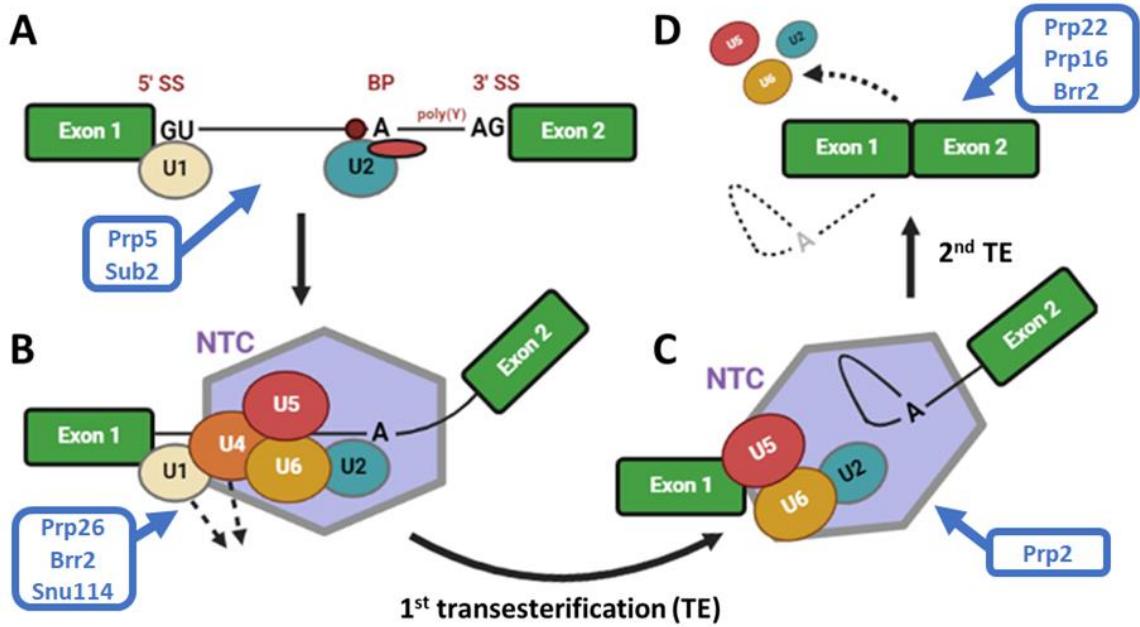


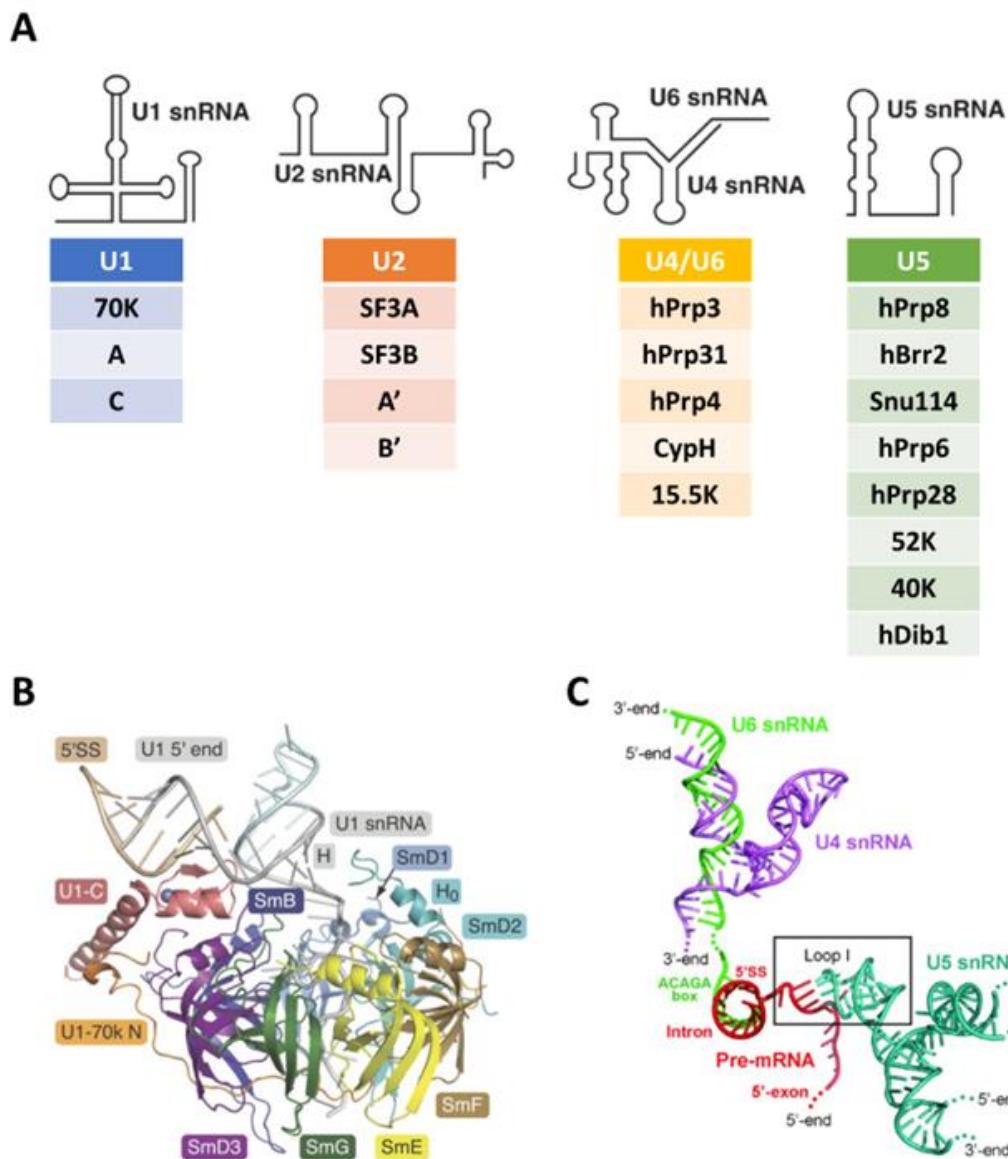
Figure 1.15 - Stages of exon splicing by the spliceosome

A. U1 binds GU at the 5' SS, whilst splicing factors SF1 and U2AF65 (maroon circle and oval) aid recruitment of U2 at the BP adenosine to form the A complex B. The U4/U6.U5 tri-snRNP is recruited along with the NTC to create the pre-catalytic spliceosome/B complex. U1 and U4 subsequently dissociate to form the B\* complex C. Catalytically activated, the B\* complex mediates the first splicing transesterification reaction creating an intron/3' exon lariat or C complex D. A second transesterification then takes place between the 5' exon and 3' SS, joining exons. The remaining intron lariat is degraded and spliceosomal snRNPs are recycled for further splicing events. DEAD- and DEAH-box ATP/GTPases are involved throughout for energy dependent structural rearrangements (blue boxes and arrows) (Figure created using biorender.com)

Spliceosomal snRNPs U1, U2 and U4/U6.U5 are multiprotein complexes with both a protein and small nuclear RNA (snRNA) component of < 200 nucleotides (Jurica and Moore, 2003). Common to all snRNPs is a core of 7 Sm proteins, E, F, G, D1, D2, D3 and B/B', which assemble into a heteroheptameric ring around an Sm consensus sequence on snRNA, RAU<sub>(4-6)</sub>GR (Urlaub *et al.*, 2001; Branlant *et al.*, 1982).

snRNP specificity is conferred by the structure of each snRNA, which recruit a range of proteins to generate a unique complex (Figure 1.16A). Splicing relies on base pairing between snRNA and mRNA sequences. The 5' end of U1 snRNA binds with the 5' SS in an interaction stabilised by U1C to define the 5' exon/intron boundary (Figure 1.16B) (Kondo *et al.*, 2015; Malca, Shomron and Ast, 2003). Conversely, U2 snRNP does not define the 3' SS alone. SF1 and U2AF65 bind the BP and poly(Y) tract before being displaced by DEAD-box helicases Prp5 and Sub2 which recruit U2, only then does snRNA-BP base pairing occur, stabilised by U2's SF3A/B complexes (Berglund, Abovich and Rosbah, 1998; Liang and Cheng, 2015; Fleckner *et*

al., 1997; Urabe *et al.*, 2021). Furthermore, base pairing dictates tri-snRNP assembly and function. U4 and U6 snRNAs have high complementarity, notwithstanding an unpaired region of U6 snRNP termed the ACAGA box that forms a helix with the 5' SS, displacing U1, whilst U5 interacts with nucleotides in the 5' exon (Figure 1.16C) (Wan *et al.*, 2016; Charenton, Wilkinson and Nagai, 2019; Kandels-Lewis and Séraphin, 1993; Artemyeva-Isman and Porter, 2021; Newman, 1997). Formation of the active B\* complex then requires Brr2 mediated unravelling of U4:U6 snRNA bonds which are supplanted by U2:U6 base pairing in the U2/U5/U6 catalytic spliceosome (Maeder, Kutach and Guthrie, 2009; Sun and Manley, 1995; Madhani and Guthrie, 1992).



**Figure 1.16 (previous page) – Spliceosomal snRNP function involves extensive protein/RNA interactions**

**A.** Spliceosomal snRNPs contain an snRNA component surrounded by a core of 7 Sm proteins (not listed), plus a complement of proteins unique to that snRNP (Figure adapted from Will and Lührmann, 2011) **B.** Crystal structure of human U1 snRNP in complex with the 5' SS of a target splice site, 3.3 Å resolution. Electrostatic interactions between U1C and the RNA backbone stabilise snRNP binding (Figure taken from Kondo *et al.*, 2015) **C.** 3.8 Å resolution cryo-EM structure showing U4/U6.U5 snRNAs binding both the 5' SS and 5' exon, as well as extensive U4:U6 snRNA complementarity (proteins not shown) (Figure taken from Wan *et al.*, 2016)

If the snRNP-directed processes outlined above occurred at all potential splicing boundaries then no alternative splicing would take place, transcripts would be constitutively spliced in a predictable manner. Therefore, deeper layers of regulation must be at play to influence the final exon composition of a transcribed pre-mRNA.

The ‘strength’ of a splice site, or its likelihood of inclusion in the final mRNA product, is dictated by numerous factors. In yeast a stringent GUAUGU is largely found at the 5' SS whilst the BP adenosine is conserved within UACUAAC (BP adenosine underlined in bold) (Qin *et al.*, 2016; Davis *et al.*, 2000; Spingola *et al.*, 1999). Humans have more permissive splice sites in which only the 5' GU, 3' AG, BP adenosine and poly(Y) tract are strictly invariant, while neighbouring sequences have greater variety (Padgett, 2012). Human splicing patterns have recently been analysed at single-cell resolution, illustrating the great diversity of splicing events within tissue compartments and cell types (Olivieri *et al.*, 2021). Therefore, the need for complex regulatory processes is apparent. Splicing regulators can be split into two broad categories: *trans*-acting splicing factors, proteins that bind RNA to influence spliceosome assembly and function; and *cis*-regulatory elements, sequences encoded in RNA which the former bind to (Figure 1.17A). The most well studied *trans*-acting proteins are the serine-arginine rich (SR) proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs) (Dvinge, 2018).

SR proteins contain a characteristic C-terminal protein-protein interaction domain enriched for arginine and serine residues, the RS domain. They undergo extensive functional regulation via phosphorylation and are largely seen as promoters of splicing that encourage inclusion of exons in the final mRNA product (Jeong, 2017). SR proteins exert their influence by binding to exonic or intronic splicing enhancer sequences (ESEs/ISEs); selectivity for which is influenced by their RNA recognition motif (RRM) domain (Figure 1.17B) (Mayeda *et al.*, 1999; Cáceres *et al.*, 1997). Once bound, SR proteins promote splicing by numerous mechanisms. For example,

SF2/SRSF1 binds U1 snRNP components *eg* U1-70K (which itself has an RS domain), strengthening interactions with the 5' SS (Jamison *et al.*, 1995; Cho *et al.*, 2011). Furthermore, SR proteins have been shown to bridge interactions between U1 and U2 snRNPs, and to assist in U4/U6.U5 tri-snRNP recruitment to the spliceosome (Shao *et al.*, 2012; Boukis *et al.*, 2004; Roscigno and Garcia-Blanco, 1995; Fetzer *et al.*, 1997). SR proteins are even capable of compensating for U1 loss of function to select the 5' SS (Tarn and Steitz, 1994).

Conversely, hnRNPs are viewed as splicing repressors which bind to exonic or intronic splicing silencers (ESSs/ISSs). hnRNPs contain a greater diversity of domains than SR proteins, they may have an RRM, a K homology (KH) or arginine-glycine-glycine (RGG) domain to mediate RNA binding, whilst protein-protein interactions are dictated by their glycine, proline or acidic-rich domains (Figure 1.17C) (Geuens, Bouhy and Timmerman, 2016). Upon binding they are capable of antagonising SR protein function, for example hnRNPA1 competes with SF2/SRSF1 to bind RNA and the relative amount of these opposing proteins influences exon inclusion (Eperon *et al.*, 2000; Zerbe *et al.*, 2004; Rooke *et al.*, 2003). hnRNPL can bind within exons and stimulate enhanced interactions between U1 and the 5' SS. Rather than this promoting splicing, interactions are hyperstabilised and unable to be dissociated by U6, preventing catalytic spliceosome formation (Chiou, Shankarling and Lynch, 2013; McClory, Lynch and Ling, 2018; Rothrock, House and Lynch, 2005).

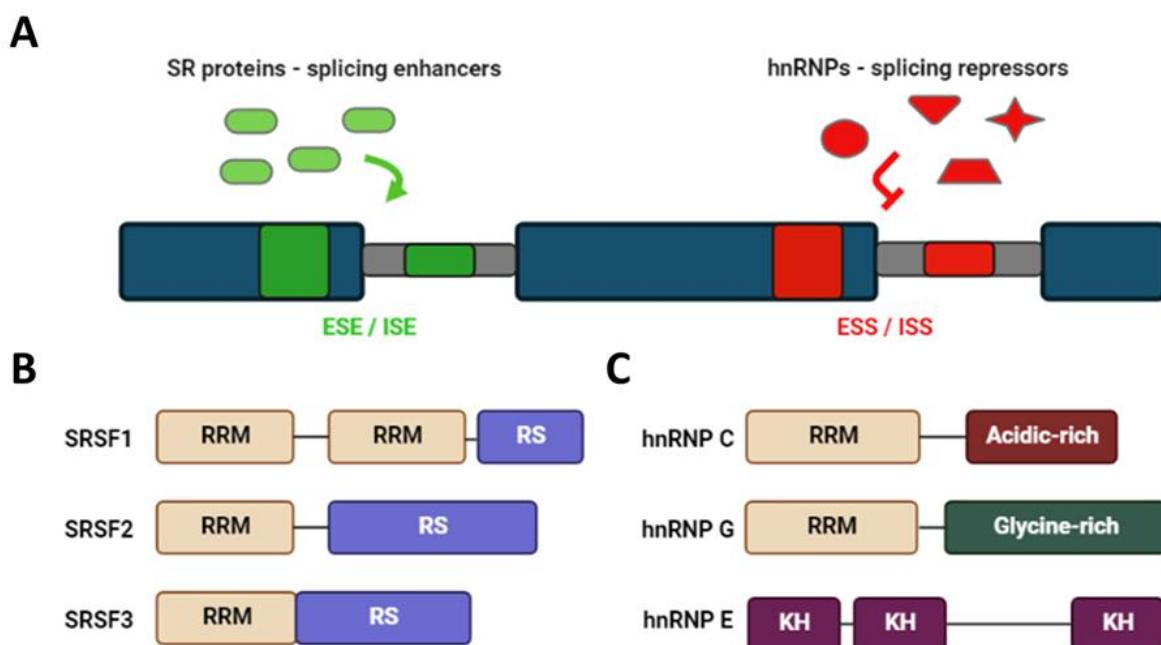


Figure 1.17 - SR proteins and hnRNPs bind RNA-encoded regulatory elements to influence alternative splicing

**A.** SR proteins and hnRNPs are the two most well characterised classes of *trans*-acting splicing factors. They bind ESE/ISEs or ESS/ISSs to either enhance or inhibit spliceosome activity, thus influencing alternative splicing decisions **B.** Domain structure of three example SR proteins. SR proteins all contain a C-terminal arginine/serine rich RS domain responsible for protein-protein interactions, and at least one RRM **C.** Domain structure of three example hnRNPs. hnRNPs are a more variable group of proteins, which may bind to RNA through either RRM, KH or RGG domains. Their protein-protein interaction domains may be glycine, proline or acidic-rich (Figure created using biorender.com)

This SR protein/hnRNP dichotomy is however something of an oversimplification, splicing control is far more complex and context dependent. SRSF10 promotes exon inclusion when bound within certain exons, whilst it results in exclusion of others (Zhou *et al.*, 2014). SR proteins can also act as splicing repressors when bound within introns (Kanopka, Mühlemann and Akusjärvi, 1996). Interestingly, work has shown that inserting an ESE sequence into an intronic region converts SR proteins into splicing inhibitors (Ibrahim *et al.*, 2005). Moreover the effects of SR protein loss are unpredictable, at certain transcript loci their depletion results in failure to recruit splicing factors, whilst at other exons knockdown enhances binding efficiency of other SR proteins (Pandit *et al.*, 2013). hnRNPs also exhibit varying influence on splicing, hnRNPL can bind to both ISEs and ESSs to enhance or repress splicing (Hui *et al.*, 2003; Rothrock, House and Lynch, 2005). Additionally, hnRNPA1/A2 can act pleiotropically within the same gene, binding intronically near the 5' SS inhibits *SMN1/2* exon 7 inclusion, yet promotes inclusion when bound near the 3' SS (Qiu *et al.*, 2022).

Detailed knowledge of alternative splicing regulation is still growing, and research is only beginning to uncover the complexity of these processes. The fate of exons during splicing is thus determined by an interplay of protein-protein and protein-RNA interactions which may exert opposing effects, the net result of which ultimately dictate final mRNA sequences (Crawford and Patton, 2006; Zahler *et al.*, 2004; Charlet-B *et al.*, 2002; Zhu, Mayeda and Krainer, 2001).

### 1.5.2 Alternative splicing and cancer

Alternative splicing dysregulation is widespread in cancer. Splicing factor mutations are frequently observed, particularly in blood cancers where *SF3B1* and *SRSF2* mutations occur in over 40% of patients with certain haematological malignancies (Anczuków and Krainer, 2016). Splicing factor alterations in solid tumours are more commonly gene amplifications or overexpression. *SRSF6* has been observed as amplified in 37% of colon cancers and overexpressed at the RNA level relative to normal tissue controls in ~50% of lung and breast tumours (Cohen-Eliav *et al.*, 2013). Expression of numerous splicing factors including SRSF1

and SRSF3 is enhanced in ovarian cancers, whilst hnRNPA2/B1 levels correlate with poor prognosis in glioblastoma (Iborra *et al.*, 2013; Golan-Gerstl *et al.*, 2011). Splicing regulators can also act as tumour suppressors, loss of which promotes oncogenesis. RBM5 protein levels are decreased in over 70% of primary lung cancers, and experimental lung cancer models show that *RBM10* knockdown augments proliferation (Oh *et al.*, 2002; Hernández *et al.*, 2016).

Deregulation of splicing factors is also seen in PCa. Sam68 is frequently overexpressed, and its levels correlate with alternative splicing of *CCND1* into a truncated, more oncogenic isoform (Paronetto *et al.*, 2010). Furthermore PCa *in vitro* models show that androgen deprivation upregulates *SRRM4* to alter splicing of neuroendocrine suppressor *REST*, resulting in AR-independent neuroendocrine differentiation (Li *et al.*, 2017). SRPK1 is a kinase that phosphorylates SRSF1 to regulate its activity. When the former is depleted, it results in alterations to pro-angiogenic *VEGF* splicing, creating an antiangiogenic isoform VEGF165b which suppresses growth and microvessel density of PCa xenografts (Mavrou *et al.* 2015; Rennel *et al.*, 2008). Entire review articles can, and have, been written in great depth exploring the role of alternative splicing in cancer/PCa (Zhang *et al.*, 2021; Olander and Lee, 2019). Instead, the relevance of alternative splicing in this review is how it pertains to generation of androgen receptor variants in CRPC, which is the emphasis of following sections.

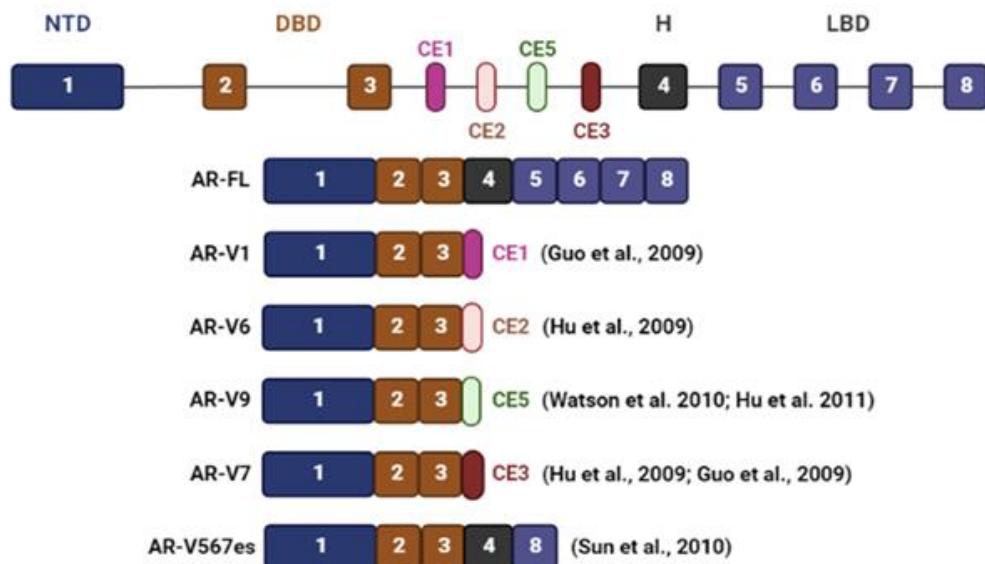
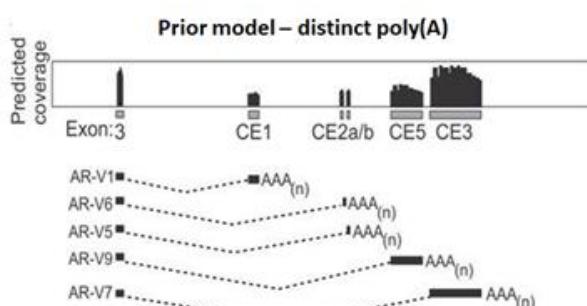
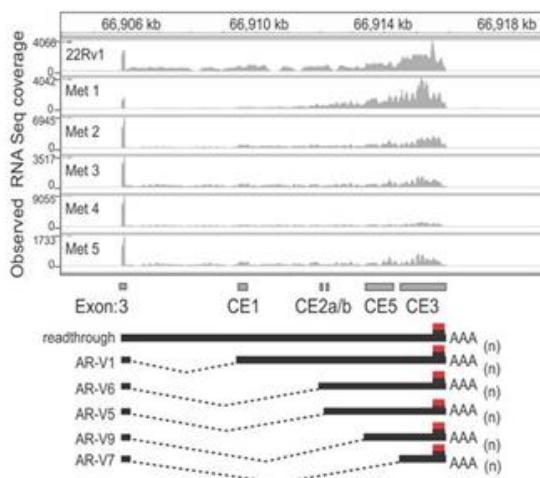
## 1.6 Androgen receptor variants: alternatively spliced drivers of CRPC

### 1.6.1 Currently identified androgen receptor variants

Short, truncated forms of AR protein have been known in PCa for more than twenty years. Work initially focused on a C-terminally truncated AR generated by proteolytic cleavage, which was historically referred to as the AR-A isoform (Wilson and McPhaul, 1996; Gregory, He and Wilson, 2001). Further studies demonstrated that this cleavage event was mediated by calpains, and it was hypothesised that genomic rearrangements, such as the *AR* exon 3 duplication seen in CRPC cell line CWR22Rv1, may alter AR protein structure in such a way that makes it amenable to attack by proteases (Libertini *et al.*, 2007; Tepper *et al.*, 2002).

Subsequent analysis of *AR* transcripts in CWR22Rv1 by 3' rapid amplification of cDNA ends (3' RACE), a method for identification of novel 3' transcript sequences using a primer anchored within a known 5' end (in this case *AR* exon 1), revealed that short *AR* isoforms can be generated at the mRNA level and aren't necessarily post-translational cleavage events (Dehm *et al.*, 2008). Numerous groups went on to report existence of truncated *AR* mRNA isoforms, termed androgen receptor variants (AR-Vs). The majority of these share the common feature of a full NTD and DBD CDS, whilst the usual LBD is replaced by 3' alternative splicing of short 'cryptic exons' (CE) arising from intronic sequences to create a premature stop codon. It is notable that not all AR-Vs are generated by 3' cryptic exon splicing, such as AR-V567es (567 'exon skip'), which includes the hinge region and exon 8 but omits exons 5-7 (Figure 1.18A).

Initially, it was proposed that CE splicing and polyadenylation occurred separately for each AR-V. This would give rise to AR-Vs containing a sequence composition of NTD-DBD-CE-poly(A) with distinct 3' termini (Figure 1.18B). AR-V7, the most prominently encountered and studied variant, encodes cryptic exon 3 (CE3) at its 3' end. Transcriptomic analysis by RNA-Seq in CWR22Rv1 and CRPC metastases demonstrated *AR* CE read coverage to occur through CE3 regardless of upstream CE. That is to say, all detected AR-Vs appeared to incorporate a common CE3 and poly(A) sequence regardless of where the initial 3' splice site selection had occurred (Figure 1.18C) (Van Etten *et al.*, 2017). This apparent commonality between AR-Vs may have implications for their regulation (Section 1.6.3 ).

**A****B****C****Proposed model – common poly(A)****Figure 1.18 - Exonic structure of prominent AR-variants**

**A.** Mature AR mRNA usually comprises eight exons encoding the domains that form the standard, or full-length, receptor (AR-FL). Aberrant alternative splicing processes give rise to a range of AR-Vs in CRPC, which generally share their NTD and DBD sequence with AR-FL but differ in their ligand binding domain. Many identified AR-Vs contain an intron-derived CE on their 3' end to generate premature in-frame stop codons, whilst AR-V567es skips exons 5-7 (Figure created using biorender.com) **B.** Early work on AR-Vs proposed that each variant terminated in a unique 3' CE and poly(A) sequence (Figure adapted from Van Etten *et al.*, 2017) **C.** Transcriptomic analysis of CWR22Rv1 cells and CRPC metastatic biopsies demonstrated a common 3' CE3 and poly(A) sequence, regardless of which upstream CE had been selected as a 3' splice site (Figure adapted from Van Etten *et al.*, 2017)

The existence of AR-Vs is not merely an mRNA phenomenon. A range of *in vitro* CRPC models express AR-V protein, most notably the CWR22Rv1, VCaP and LNCaP95 cell lines (Sharp *et al.*, 2019). The presence of AR-V protein has also been exhibited in CRPC patient material where AR-V7, for which variant-specific antibodies have been developed, is the most consistently observed and correlates with disease progression (Guo *et al.*, 2009; Hu *et al.*, 2009; Sharp *et al.*, 2019; Scher *et al.*, 2016; Sowalsky *et al.*, 2022).

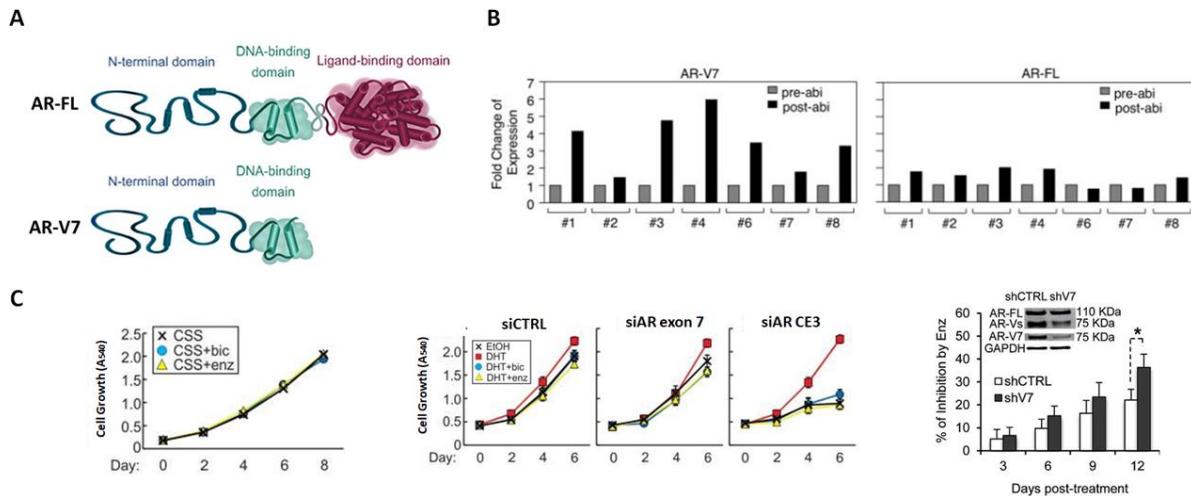
### **1.6.2 AR-V structure and function: a clinical challenge**

AR-Vs association with CRPC progression lies in their protein structure and function. The normal, or ‘full-length’, AR (AR-FL) contains four defined structural domains which collectively enact AR signalling, as outlined previously (Section 1.2.1). Aberrant alternative splicing results in a truncated AR-V protein structure that lacks the LBD due to translation of in-frame stop codons (Figure 1.19A). Rather than this disabling AR function, the outcome is a constitutively active form of AR that does not require androgen binding for transactivation of canonical AR target genes including *KLK2*, *PSA*, *FKBP5* and *TMPRSS2* (Dehm *et al.*, 2008; Guo *et al.*, 2009; Hu *et al.*, 2009). A CRISPR-engineered CWR22Rv1 knockout line developed by our lab further demonstrates the ability of AR-Vs to continually drive AR signalling in the complete absence of AR-FL protein (Kounatidou *et al.*, 2019). Analysis of the AR-V7 cistrome using ChIP-seq in a range of both cell line and primary CRPC models has revealed significant DNA-binding overlap between AR-FL and AR-V7, however a considerable number of loci are uniquely bound by AR-V7 (Sugiura *et al.*, 2020; Basil *et al.*, 2022; Chen *et al.*, 2018). These cistromic studies are consistent with RNA-Seq results showing that AR-V7 shares a canonical AR-signalling transcriptome with AR-FL, yet also mediates variant-specific expression signatures (Basil *et al.*, 2022; Nagandla *et al.*, 2020; Hu *et al.*, 2012).

AR-V constitutive activity relies on the fact that the N-terminal AF1 region of AR is the primary determinant of transactivation ability and coactivator recruitment, whilst AF2 deletion fails to impact function (Bevan *et al.*, 1999; Wärnmark *et al.*, 2003). This contrasts with ER where AF2/AF1 synergy is essential for receptor activity (Kumar *et al.*, 2011). Although AR-V7 lacks the NLS of AR-FL, it undergoes nuclear translocation via active transport by nucleoporins in a mechanism distinct from importin- $\alpha$  dependent AR-FL transport (Kim, Au and Jamalruddin *et al.*, 2022; Chan, Li and Dehm, 2012). Interestingly, AR-V7 and AR-V567es can either

homodimerise or heterodimerise with AR-FL, and exhibit androgen-independent transactivation whether as a homo- or heterodimer (Xu *et al.*, 2015; Cao and Qi *et al.*, 2014).

This constitutive transactivation would in theory confer resistance to hormone therapy, which is demonstrated in preclinical xenograft models treated with CYP17A1 inhibitor AA that markedly elevate AR-V7 splicing upon progression (Figure 1.19B) (Liu *et al.*, 2016; Yu *et al.*, 2014; Hu *et al.*, 2012; Mostaghel *et al.*, 2011). Furthermore, AR-Vs lack the LBD binding site of all currently available NSAAAs. This is reflected in studies using AR-V7 expressing cell line and mouse models of CRPC that are resistant to NSAA treatment (Figure 1.19C) (Li *et al.*, 2013; Cao, Qi *et al.*, 2014; Zhao *et al.*, 2020).

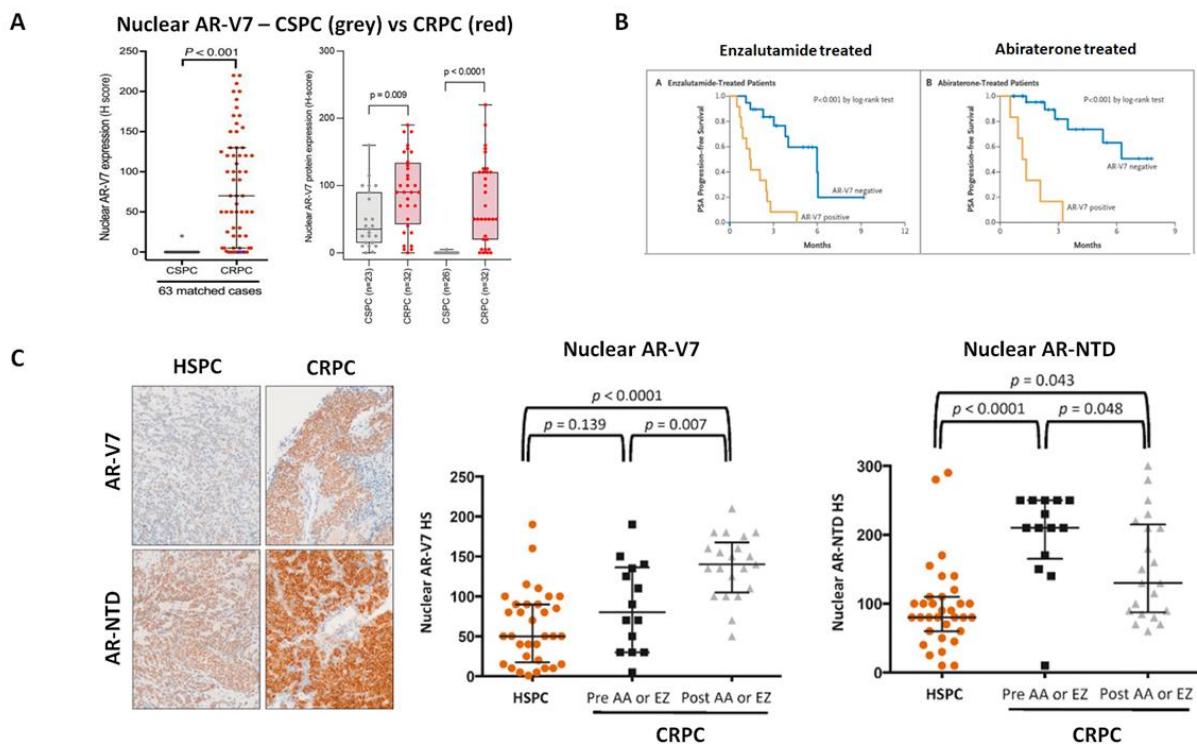


**Figure 1.19 - AR-V7 elicits resistance to abiraterone and NSAAAs in preclinical models of CRPC**

**A.** AR-Vs, of which AR-V7 is the most commonly encountered, lack the full LBD of full-length AR (AR-FL). This results in a C-terminally truncated protein structure that retains the NTD (including the crucial AF1 domain) and DBD, retaining transactivation ability whilst omitting the need for ligand stimulation and lacking the binding site of all available NSAAAs (Figure adapted from Kuznik *et al.*, 2021) **B.** VCaP PCa xenografts undergo increased AR-V7 expression after treatment with abiraterone. AR-V7 increases to a greater degree than AR-FL, indicating a splicing effect rather than a general increase in AR gene expression (Figure adapted from Yu *et al.*, 2014) **C.** CRPC cell line CWR22Rv1 grown in androgen-depleted conditions exhibits resistance to both bicalutamide and enzalutamide (left panel), while siRNA targeting of AR-V7 CE3 sensitises cells to these NSAAAs (middle panel). shRNA targeting of AR-V7 in a CWR22Rv1 xenograft potentiates the ability of enzalutamide to inhibit tumour growth (right panel) (Figures adapted from Li *et al.*, 2013 and Cao, Qi *et al.*, 2014)

The relevance of AR-Vs to therapeutic resistance is further exemplified in clinical samples. IHC analysis of CSPC vs CRPC indicates enhanced nuclear AR-V7 staining in the latter samples (Figure 1.20A) (Sharp *et al.*, 2019; Sowalsky *et al.*, 2022). Likewise, qPCR analysis of AR-V7 transcripts in CRPC circulating tumour cells reveals striking survival differences between AR-V7 positive and negative patients treated with enzalutamide or AA (Figure 1.20B) (Antonarakis *et al.*, 2014). Interestingly, tissue staining intensity of both AR-V7 and AR-NTD

(representing total AR) increases in CRPC relative to CSPC (or HSPC *ie* hormone-sensitive), though dividing samples into pre- or post-AA/enzalutamide treatment shows that AR-V7 levels rise after these therapeutic interventions. However, AR-NTD intensity decreases post-AA/enzalutamide, indicating a splicing event rather than simply an increase in total AR (Figure 1.20C) (Welti *et al.*, 2016).



**Figure 1.20 - AR-V7 expression correlates with development of CRPC, hormonal therapy exposure and poor clinical outcomes**

**A.** CRPC (red) exhibits higher nuclear AR-V7 expression than CSPC (grey), measured by IHC. This association was found in both matched and unmatched tumour samples, and with two independent antibodies (Figures adapted from Sharp *et al.*, 2019 and Sowalsky *et al.*, 2022) **B.** Circulating tumour cells from CRPC patients were profiled by qPCR for AR-V7 expression status. Patients with detectable AR-V7 (orange) display significantly worse PSA progression-free survival ( $p < 0.001$ ) than patients with no measurable AR-V7 (blue) (Figure adapted from Antonarakis *et al.*, 2014) **C.** IHC identifies higher AR-V7 and total AR (AR-NTD) expression in CRPC than CSPC (or HSPC *ie* hormone-sensitive). AR-V7 staining is enhanced by treatment with abiraterone or enzalutamide, however total AR (AR-NTD) expression falls, suggesting a splicing effect rather than simply an increase in AR gene expression (Figures adapted from Welti *et al.*, 2016)

The clinical challenge presented by AR-Vs has created interest in inhibiting either the AR NTD or DBD. The former is challenging due to the inherently disordered nature and lack of stable NTD structures, hampering traditional structure-based drug discovery efforts (McEwan, 2012; Lavery and McEwan, 2008). Nevertheless, a range of NTD-targeting compounds have been developed of which one example, EPI-506, inhibits transactivating protein-protein interactions at AF1 to impede growth of AR-V7 driven CRPC xenografts (Yang *et al.*, 2016).

EPI-506 subsequently entered clinical trials, however this was terminated due to lack of PSA response, poor pharmacokinetics and excessive pill burden (Le Moigne *et al.*, 2019; Maurice-Dror *et al.*, 2022; NCT02606123). Additional anti-NTD agents such as EPI-7386 have been synthesised which also showed encouraging preclinical results (Hong *et al.*, 2020; Le Moigne *et al.*, 2021). Consequently, EPI-7386 is now recruiting for clinical trials (NCT04421222, NCT05075577). Given extensive DBD homology with other nuclear hormone receptors (Figure 1.7A), selective inhibition of the AR by targeting this domain is likely to be challenging. Although compounds targeting the DBD have been created which show preclinical efficacy, this strategy is yet to reach clinical trials for CRPC (Radaeva *et al.*, 2021; Elgehama *et al.*, 2021; Yang *et al.*, 2020; Dalal *et al.*, 2014).

Inhibition of AR NTD cochaperone proteins may present a therapeutic alternative to inhibiting AR-Vs directly. For example BAG1L, a member of the BAG1 cochaperone family, is a known molecular interactor of AR capable of binding within the NTD to recruit transcriptional coactivators at AR AF1 (Lee *et al.*, 2019). Targeting of BAG1L has thus recently been touted as a possible therapeutic strategy in CRPC by having the potential to inhibit all known AR-Vs, however its clinical utility is debated due to challenges in BAG1L druggability (Neeb *et al.*, 2022). Heat shock proteins HSP40 and HSP70 have also been shown to play key chaperoning roles in AR-V7 transactivation by binding to the AR NTD (Moses *et al.*, 2018; Dong *et al.*, 2019), as has KIF15 (Gao *et al.*, 2021). All of these indirect approaches have shown preclinical promise, and it remains to be seen whether they becomes a viable route to inhibit AR-V activity.

Therefore, efforts to inhibit the remaining functional domains of AR-Vs are ongoing and yet to yield results. An alternative approach is to block their production altogether, which requires mechanistic knowledge of their generation by alternative splicing.

### 1.6.3 Current understanding of AR-V alternative splicing

Knowledge of splicing regulatory mechanisms governing AR-V generation in CRPC is burgeoning. Studies to date have utilised *in silico* NGS analyses of PCa patient datasets, drawing correlations between expression of AR-Vs and candidate regulators for subsequent validation by molecular biology techniques.

#### ***Splicing factors/RNA-binding proteins***

Kawamura *et al.* selected a panel of 309 genes encoding splicing factors or RNA-binding proteins (RBPs) based on their gene ontology (GO) annotation. These were then filtered based on upregulation in both mCRPC vs localised, treatment naïve PCa and in patients expressing high vs low AR-V7 RNA (Grasso *et al.*, 2012; TCGA, 2015). Depletion of overlapping genes in CWR22Rv1 using CRISPR-Cas9 revealed U2 snRNP component SF3B2, which is targetable by SF3B inhibitor pladienolide B, to be implicated in AR-V7 generation (Kawamura *et al.*, 2019). Other work by Nadiminty *et al.* demonstrated that hnRNPA1 is overexpressed in prostate tumours compared to benign controls. Also, the authors found that hnRNPA1 levels positively correlated with AR-V7 expression and were enhanced in PCa cell lines conditioned for resistance to enzalutamide (Nadiminty *et al.*, 2015). Furthermore, the plant flavonoid quercetin suppresses hnRNPA1 and AR-V7 splicing, which is exhibited to resensitise enzalutamide-resistant cells using *in vitro* studies (Tummala *et al.*, 2017). Numerous studies have shown other splicing factors and RBPs to be involved in AR-V7 generation including U2AF65, SRSF1, SF3B1, DDX39, SFPQ, Sam68 and RBM39, where interactions between proteins and AR mRNA were validated by use of RNA immunoprecipitation techniques (Liu *et al.*, 2014; Wang *et al.*, 2017; Nakata *et al.*, 2017; Takayama *et al.*, 2017; Stockley *et al.*, 2015; Melnyk *et al.*, 2020).

#### ***Alternative poly(A) selection***

AR-V CEs are spliced in place of LBD exons at the 3' end of RNA. Therefore, studies have elucidated mechanisms by which 3' end processing factors responsible for pre-mRNA cleavage and polyadenylation, essential to maturation and translation (Danckwardt, Hentze and Kulozik, 2008), are involved in AR-V generation. Work performed in CWR22Rv1 by Van Etten *et al.* observed a poly(A) signal in AR intron 3, found at the 3' end of CE3 and common to all observed AR-Vs (Figure 1.18C). Blocking this sequence with morpholino oligomers led

to reduced AR-V splicing and increased AR-FL expression, whilst gene silencing experiments revealed the cleavage and polyadenylation specificity factor (CPSF) complex was central to this as knockdown of CPSF1 and CPSF3 also suppressed AR-V production (Van Etten *et al.*, 2017). Research has also demonstrated that expression of SIAH1, an E3 ubiquitin ligase that targets CPSF1 for degradation, negatively correlates with clinical PCa progression, further corroborating previous observations. Moreover the authors showed SIAH1-mediated quashing of AR-V7 generation could be resisted by CPSF1 overexpression (Xia *et al.*, 2022). Recent work by Sun *et al.* exhibited that cyclin K-CDK12 dysfunction through mutation/deletion promotes alternative poly(A) site selection in *AR* intron 3, they also showed that pharmacological inhibition of CDK12 increases AR-V7 expression in numerous PCa cell lines (Sun *et al.*, 2022). CDK12's role in poly(A) regulation has been documented previously in DNA repair genes, and this link with *AR* control may explain the shorter times to PSA progression seen in *CDK12* mutant CRPC patients on abiraterone or enzalutamide (Krajewska *et al.*, 2019; Dubbury, Boutz and Sharp, 2018; Reimers, Yip *et al.*, 2020).

The relative importance of alternative splicing vs alternative polyadenylation in AR-V generation is debated, particularly given the previously discussed data demonstrating that all CEs share a common CE3 and poly(A) tail (Van Etten *et al.*, 2017). However, alternative splicing must still remain an initiating event in AR-V production, otherwise all AR-Vs would simply exist as the AR intron 3 sequence up to the site of alternative polyadenylation, and the range of AR-Vs currently observed would not exist. Furthermore CPSF1, a prominent factor reported to be involved in AR-V alternative polyadenylation, recognises a canonical AAUAAA hexamer, found in CE3 (Kumar *et al.*, 2019; Van Etten *et al.*, 2017). Therefore it seems that rather than recognising an 'alternative' polyadenylation site, such factors may simply bind and process a canonical site that is made available via alternative splicing of CEs. Interestingly, recent work has shown the existence of 'spliced polyadenylated introns', purported to be removed from transcripts by the spliceosome to prevent widespread use of intronic polyadenylation sites in eukaryotes (Vlasenok, Margasyuk and Pervouchine, 2022). Collectively, it is likely that although polyadenylation is indeed necessary for AR-V maturation, alternative splicing remains the key event in their generation and has thus been the focus of studies to date.

### ***Chromatin-bound regulators***

A body of work has also highlighted how chromatin bound proteins alter recruitment of splicing machinery in PCa. KDM3A, a histone lysine demethylase, has been shown via ChIP assay in CRPC cell lines to localise at cryptic exon DNA, recruiting hnRNPF and subsequently U2AF65 to promote CE inclusion during splicing (Fan *et al.*, 2018). Similar observations using *in vitro* models were made with the demethylase KDM4B, where it was shown that phosphorylation by protein kinase A in androgen depleted conditions enables KDM4B to tether SF3B3 at cryptic exon loci (Duan *et al.*, 2019). The 2OG-dependent dioxygenase JMJD6 has multiple functions including histone arginine demethylation and lysine hydroxylation (Kwok *et al.*, 2017). Paschalis *et al.* performed an siRNA screen of putative AR-V7 splicing regulators based on expression changes using multiple datasets including in CRPC cell line LNCaP95 vs CSPC LNCaP. This identified JMJD6 as a mediator of AR-V7 expression which hydroxylates U2AF65 to enable its recruitment at AR pre-mRNA (Paschalis, Welti *et al.*, 2021).

In addition to influencing splicing by direct recruitment of splicing factors, chromatin structure, influenced by histone modifying enzymes such as those described above, has been shown to impact pre-mRNA splicing regulation. Nascent mRNAs, subject to splicing, are in close physical proximity to chromatin while transcription by RNA polymerase II takes place, and it is known that a large degree of splicing occurs cotranscriptionally before completion of the final mRNA transcript (Herzel *et al.*, 2017). Structural alterations to chromatin exerted by, for example, differences in histone methylation or acetylation, can influence rates of transcription by RNA polymerase II within genes to alter expression kinetics (Woo *et al.*, 2017; Stasevich *et al.*, 2014). Given there is evidence of substantial crosstalk occurring between splicing and transcriptional machinery, it is posited that rates of transcriptional elongation influence splicing protein recruitment at pre-mRNA (Shenasa and Bentley, 2023). The role of chromatin structure, histone modifications and cotranscriptional interactions in alternative splicing is an active area of research, and efforts have been made to uncover splicing-associated chromatin signatures linked with levels of exon inclusion/exclusion (Agirre *et al.*, 2021).

It is likely that alternative splicing of AR-V7 is influenced by a combination of the above factors, with splicing factors and RBPs, alternative polyadenylation and chromatin features all playing a role. The primary interest of this PhD project relates to splicing factors and RBPs,

however a summary of literature on the different categories of published factors is found in Table 1.2.

Gene/Protein	References
<b>Splicing factors/RNA-binding proteins (RBPs)</b>	
<b>SF3B1</b>	<i>Wang et al., 2017</i>
<b>SF3B2</b>	<i>Kawamura et al., 2019</i>
<b>hnRNPA1</b>	<i>Nadiminty et al., 2015</i>
<b>U2AF65</b>	<i>Liu et al., 2014; Paschalis et al., 2021</i>
<b>SRSF1</b>	<i>Liu et al., 2014</i>
<b>DDX39</b>	<i>Nakata et al., 2017</i>
<b>SFPQ</b>	<i>Takayama et al., 2017</i>
<b>Sam68</b>	<i>Stockley et al., 2015</i>
<b>RBM39</b>	<i>Melnyk et al., 2020</i>
<b>Alternative poly(A) selection</b>	
<b>CPSF1</b>	<i>Van Etten et al., 2017; Xia et al., 2022</i>
<b>CPSF3</b>	<i>Van Etten et al., 2017</i>
<b>CDK12</b>	<i>Sun et al., 2022</i>
<b>Chromatin-bound regulators</b>	
<b>KDM3A</b>	<i>Fan et al., 2018</i>
<b>KDM4B</b>	<i>Duan et al., 2019</i>
<b>JMJD6</b>	<i>Paschalis, Welti et al., 2021</i>

**Table 1.2 - Literature summary of currently identified AR-V splicing/processing regulators**

#### 1.6.4 Pharmacological inhibition of splicing

The central roles alternative mRNA splicing and processing play in AR-V generation raises the question of how it can be inhibited. Several splicing inhibitory compounds have been used as experimental cancer targeting agents to date, some of the earliest of which were pladienolides isolated from *Streptomyces platensis* (Mizui *et al.*, 2004). A derivative of pladienolide B, E7107, was subsequently demonstrated as a potent inhibitor of the SF3B complex, a major component of the U2 snRNP, and has since exhibited anti-tumour activity in a range of *in vitro* and *in vivo* cancer models (Kotake *et al.*, 2007; Folco, Coil and Reed, 2011; Aird *et al.*, 2019; Zhang *et al.*, 2020; Chan *et al.*, 2017). Spliceostatins *eg* spliceostatin A derived from *Pseudomonas* sp. have also been used preclinically to inhibit the SF3B complex, alter splicing patterns and downregulate expression of pro-proliferative genes (Kaida *et al.*, 2007; Corroniero, Miñana and Valcárcel 2011; Yoshimoto *et al.*, 2017; Yoshimoto *et al.*, 2021). Furthermore, another SF3B-targeting agent originating from *Streptomyces* sp., GEX1A, inhibits SF3B1-PHF5A interactions in the U2 snRNP complex and demonstrates preclinical efficacy in models of leukaemia (Hasegawa *et al.*, 2011; Sellin *et al.*, 2022).

Although a range of additional compounds that target the spliceosome have been used preclinically including sudemycins, meayamycin B and isoginkgetin (Thurman *et al.*, 2017; Makowski *et al.*, 2017; Wojtuszkiewicz *et al.*, 2014; O'Brien *et al.*, 2008), at present only two splicing modulators, E7107 and H3B-8800, have made it to Phase I trials. Both these pladienolide derivatives showed encouraging preclinical results (Kotake *et al.*, 2007; Seiler *et al.*, 2018), although E7107 Phase I trials in solid tumour patients were suspended due to ophthalmologic toxicities (Eskens *et al.*, 2013; Hong *et al.*, 2014; NCT00499499 and NCT00459823). H3B-8800 use in haematological malignancies showed better tolerability, however it failed to elicit even a partial response in any patients (Steensma *et al.*, 2019; Steensma *et al.*, 2021; NCT02841540).

Attempts at pharmacological splicing modulation have hitherto focused on core spliceosomal proteins and are likely to cause widespread alternative splicing perturbations. As outlined in Section 1.5 these processes are highly complex, with numerous accessory regulators dictating activity of the core spliceosome. A more nuanced understanding of proteins driving these processes will only benefit attempts to target aberrant alternative splicing. This PhD project will apply novel biotechnology to acquire increased knowledge of protein mediators that regulate AR-V splicing in CRPC.

## 1.7 CRISPR-Cas13 and proximity biotinylation

This work will combine two contemporary technologies: CRISPR-Cas13 and proximity biotinylation. A more detailed mechanistic and technical discussion of their use follows in Chapter 4, whilst this section will outline their fundamental concepts.

### 1.7.1 CRISPR-Cas13: RNA-targeting Cas nucleases

Applications of the prokaryotic adaptive immune system CRISPR (clustered regularly interspersed short palindromic repeats) have revolutionised molecular biology (Jiang and Doudna, 2017). Most work to date has been performed with CRISPR-Cas9 nuclease systems, where the ability to modify DNA at precise genomic loci as directed by a complementary guide RNA (gRNA) has enabled development of knockout models, gene insertions or corrections by homology-directed repair with a donor template, and epigenetic modifications through fusing catalytically inactivated ‘dead’ Cas9 (dCas9) to a range of enzymes (Doudna and Charpentier, 2014).

More recently discovered are RNA-targeting Cas nucleases, the Cas13 family, which are gaining traction as tools for biomedical research (Figure 1.21A). Since 2016 numerous Cas13 proteins (lettered Cas13a-d) have been identified in prokaryotes, all of which contain two Higher Eukaryotes and Prokaryotes Nucleotide-binding (HEPN) domains responsible for single-stranded RNase activity when activated by gRNA complementarity with a target RNA (Figure 1.21B). This mechanism has been applied in eukaryotic cells for potent, specific RNA knockdown. Furthermore and analogous to dCas9, targeted mutation of catalytic residues in the HEPN domains yields a ‘dead’ Cas13 (dCas13) that retains RNA-binding capability without subsequent transcript cleavage (Tambe and East-Seletsky *et al.*, 2018; Konermann *et al.*, 2018; Yan *et al.*, 2018; Smargon *et al.*, 2017; Abudayyeh *et al.*, 2016). A range of techniques in mammalian cells have been developed by fusing dCas13 with other functional proteins including base-editing and epitranscriptomic alterations, splicing modulation, live-cell imaging of RNA dynamics and characterisation of RNA-protein interactions (Figure 1.21C) (Cox, Gootenberg and Abudayyeh *et al.*, 2017; Wilson *et al.*, 2020; Li *et al.*, 2020; Du *et al.*, 2020; Konermann *et al.*, 2018; Yang *et al.*, 2019; Zhang *et al.*, 2020; Han *et al.*, 2020; Lin *et al.*, 2021).

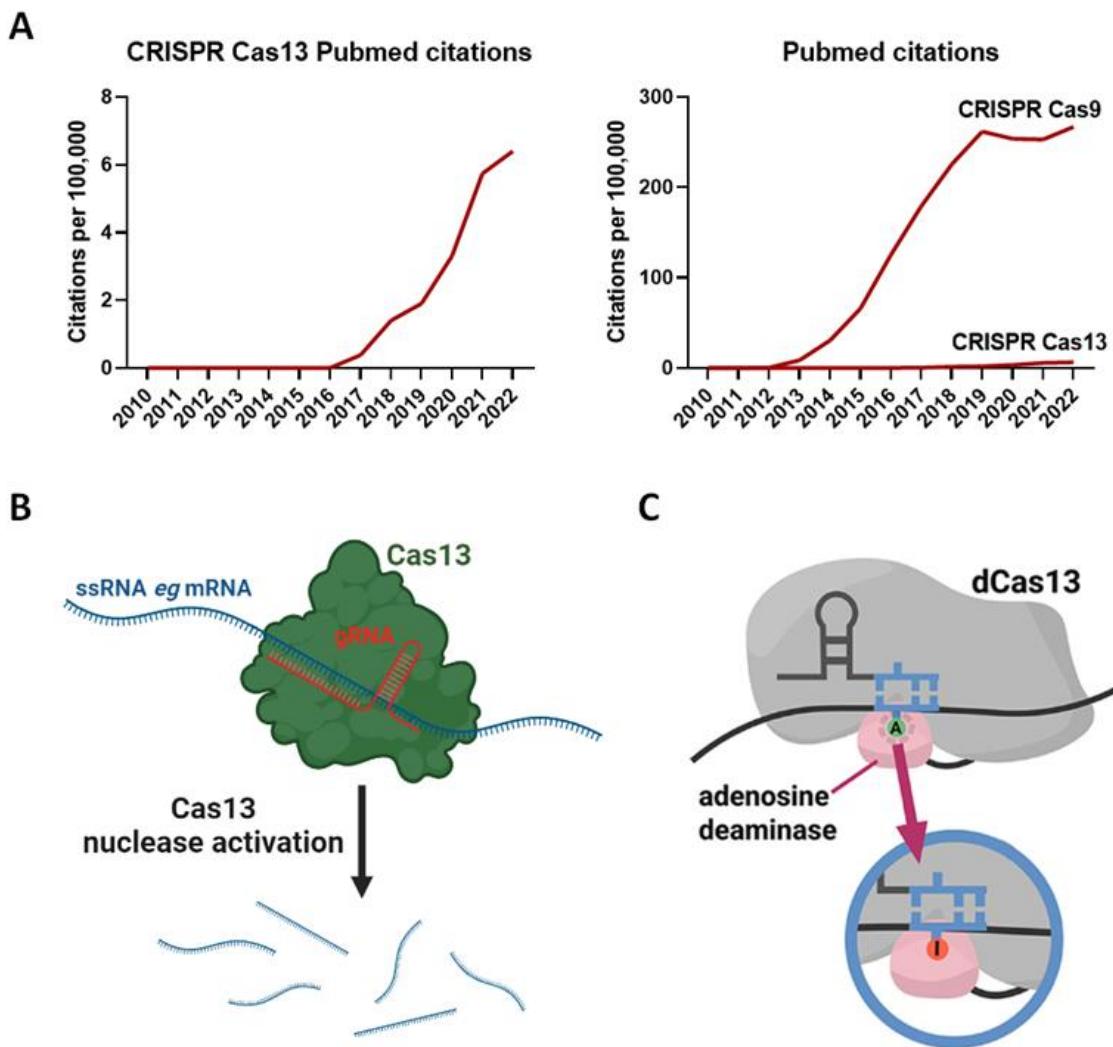


Figure 1.21 - Cas13 is a novel CRISPR nuclease usable for studying RNA biology

**A.** RNA-targeting Cas13 CRISPR proteins are a growing area of research, though their use still significantly lags behind that of Cas9 (Pubmed citation statistics collated using Pubmed by Year, accessible at <https://esperr.github.io/pubmed-by-year/>) **B.** The sequence-specific RNase activity of Cas13 HEPN domains is dictated by gRNA sequence complementarity with ssRNA. gRNA binding with ssRNA activates Cas13 nuclease activity resulting in RNA cleavage (Figure created using biorender.com) **C.** Catalytically inactive dCas13 proteins can be fused with functional partners to enable a range of RNA-targeting possibilities including selective base editing as performed in Cox, Gootenberg and Abudayyeh *et al.* (2017). Here, adenosine deaminase ADAR can be used to selectively convert adenosine to inosine on RNA. Inosine base pairs with cytosine, mediating changes to mRNA translation (Figure taken from Palaz *et al.*, 2021)

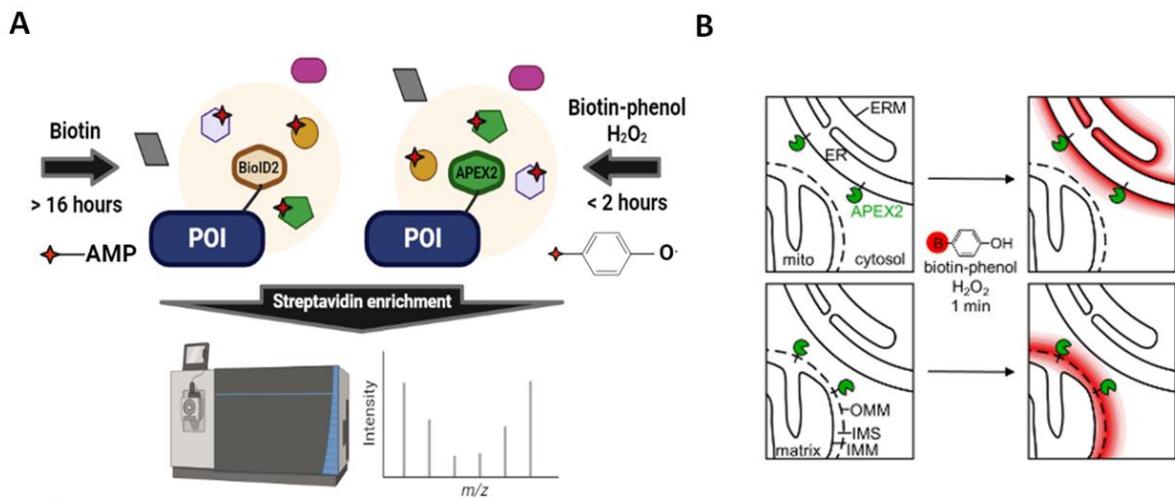
### 1.7.2 Proximity biotinylation unveils protein interactomes

A range of methods have been developed to ascertain cellular protein interaction networks. One widely used example is Rapid Immunoprecipitation Mass Spectrometry (RIME), where formaldehyde is used as a chemical crosslinker to bond proteins covalently, before a target protein is immunoprecipitated and binding partners identified by mass spectrometry (Mohammed *et al.*, 2016). However, given that formaldehyde will indiscriminately bond all

structures, there is high potential for false positives, as large, crosslinked complexes may be enriched that do not accurately reflect the spatial proximity of identified protein interactors. Furthermore, workflows are heavily dependent on the quality and specificity of the antibody used for immunoprecipitation (Qin *et al.*, 2021).

A powerful alternative is to fuse a protein of interest (POI) with a proximity biotinylation enzyme. The first application of this technique in mammalian cells was with BirA\*, a mutated form of the *E. coli* BirA biotin ligase that catalyses formation of a reactive biotinyl-AMP (bioAMP) from biotin and ATP (Roux *et al.*, 2012). bioAMP can react with primary amines of nearby proteins, covalently tagging them with biotin for subsequent cell lysis, streptavidin protein capture and mass spectrometric identification in a method termed BioID, later improved through using an *Aquifex aeolicus* biotin ligase (BioID2) (Roux *et al.*, 2018; Kim *et al.*, 2016).

Work performed by Alice Ting's lab to mutagenically engineer soybean ascorbate peroxidase APX resulted in creation of APEX (Martell *et al.*, 2012). In the presence of hydrogen peroxide ( $H_2O_2$ ), this monomeric enzyme catalyses oxidation of phenol compounds into short-lived phenoxyl radicals that react with electron-rich amino acids (eg Tyr, Trp, His and Cys), which when using biotin-phenol results in a covalent biotin tag on nearby proteins (Rhee and Zou *et al.*, 2013). Just as with BioID2, APEX was further improved using a yeast display screen of mutants and sorting by FACS for those with the highest enzymatic biotinylation activity, resulting in APEX2 (Lam *et al.*, 2015). APEX2 has significant time advantages over BioID2 as biotinylation and lysis is performed within 1-2 hours with the former rather than > 16 hours, enabling faster workflows and superior temporal resolution of interactions (Figure 1.22A) (Che and Khavari, 2017). APEX2 fusion proteins have since been applied for live-cell labelling and interaction proteomics in numerous settings including mitochondrial compartments, DNA damage responses and histone modifications (Figure 1.22B) (Han *et al.*, 2017; Gupta *et al.*, 2018; Li, Zhou, Zhao and Wen *et al.*, 2022; Hung *et al.*, 2016). Collectively, proximity biotinylation methods provide the significant benefit of leveraging streptavidin-biotin interactions, one of the strongest known in nature which is far in excess of ordinary protein-ligand binding affinity (Liu, Zhang and Mei, 2016). This enables use of rigorous post-capture wash steps to remove background contaminants, whilst also obviating requirements to source and optimise an appropriate antibody for pulldown.



**Figure 1.22 - Proximity biotinylation methods enable proteomic characterisation of interactomes**

**A.** Proximity biotinylation techniques enable live-cell covalent biotin tagging of protein interactors. Cell lysis and streptavidin pulldown allows enrichment of these biotinylated proteins for subsequent mass spectrometry. By fusing a POI to appropriate enzymes, the identity and abundance of POI interactome components can be ascertained. BiOID2 (left) and APEX2 (right) are two prominent methods that use different reaction chemistries. BiOID2 catalyses formation of a reactive biotinyl-AMP substrate from biotin and endogenous ATP, whilst APEX2 relies on exogenous biotin phenol and hydrogen peroxide to generate biotin-phenoxyl radicals. Both biotin substrates react with proximal amino acids to covalently tag proteins (biotin – red star) (Figure created using biorender.com) **B.** APEX2 fusion proteins can be used for a range of interactome studies. Hung *et al.* fused APEX2 to POIs in either the mitochondrial or endoplasmic reticulum outer membranes to proteomically map associated factors (Figure taken from Hung *et al.*, 2017)

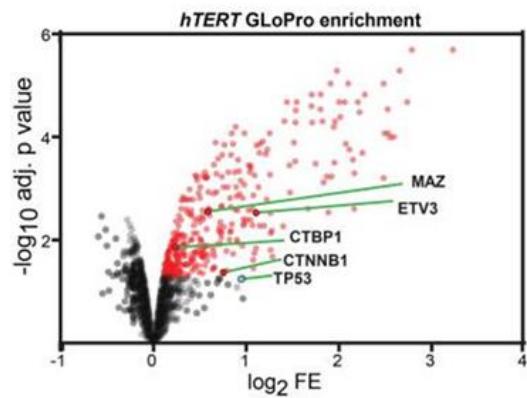
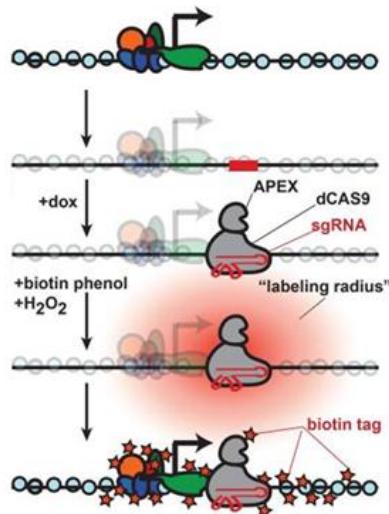
### 1.7.3 CRISPR and proximity biotinylation combine: proof of concept

Synergistic leverage of CRISPR and APEX2 has been achieved in eukaryotic cells previously, offering confidence that these methods are appropriate in the study of pathogenic AR-V splicing in CRPC. This was initially performed using dCas9 fusions, resulting in publication of methods termed GLoPro, C-BERST and CAPLOCUS (Myers *et al.*, 2018; Gao *et al.*, 2018; Gao, Rodriguez and Sontheimer, 2019; Qiu *et al.*, 2019). These applied gRNAs to target dCas9 at a genomic locus of interest before biotinylation of proximal interactors with APEX2. Next, streptavidin enrichment and mass spectrometric measurement of protein abundance was compared with a ‘non-targeting’ gRNA control to ascertain local proteins (Figure 1.23A).

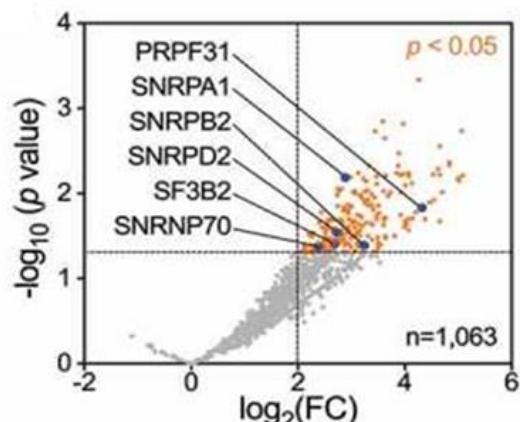
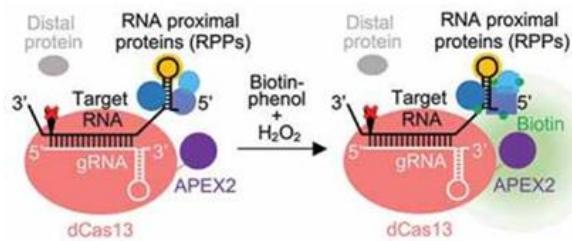
This has since been expanded to CRISPR-Cas13 systems, enabling identification of RNA interactomes using gRNA-directed dCas13-APEX2 fusions in workflows largely analogous to those with dCas9 (Figure 1.23B) (Lin *et al.*, 2021; Han *et al.*, 2020). However these publications have only used HEK293 cells as a technical proof of concept, and combining dCas13 with

APEX2 has not yet been achieved in models of cancer. This PhD project will employ these methods to reveal mechanistic insights into processes governing AR-V generation in CRPC.

**A**



**B**



**Figure 1.23 - Proximity biotinylation can combine with Cas9/Cas13 to elucidate sequence-specific DNA/RNA protein interactomes**

**A.** Myers *et al.* (2018) localised a DNA-binding dCas9-APEX2 fusion protein to the *hTERT* promoter with a complementary gRNA to reveal local protein interactors ( $\log_2 \text{FE}$  is  $\log_2$  fold protein enrichment of *hTERT*-targeting gRNA samples over non-targeting gRNA) (Figure taken from Myers *et al.*, 2018) **B.** Lin *et al.* (2021) adopted a similar approach with a dCas13-APEX2 fusion targeted to U1 snRNA ( $\log_2(\text{FC})$  is  $\log_2$  fold protein enrichment of U1 snRNA-targeting gRNA samples over non-targeting gRNA) (Figure taken from Lin *et al.*, 2021)

## Chapter 2 - Aims and objectives

Inhibition of AR signalling by depletion of circulating testosterone with ADT or direct receptor targeting using antiandrogens is the clinical standard for management of metastatic PCa. Whilst these can elicit substantial clinical responses, a significant number of cases will develop resistance. It is now recognised that this disease state, CRPC, continues to activate AR through numerous mechanisms.

One such mechanism is generation of AR-Vs. AR-Vs arise from alternative splicing of *AR* pre-mRNA to generate premature stop codons and ultimately a truncated AR protein without the LBD of AR-FL. The most prominently characterised AR-V, AR-V7, splices cryptic exon CE3 at its 3' end to enable CRPC progression by constitutive ligand-independent transactivation. Furthermore it lacks the binding site of all available antiandrogens, conferring resistance to their use, and attempts to directly inhibit AR-Vs have not yielded success.

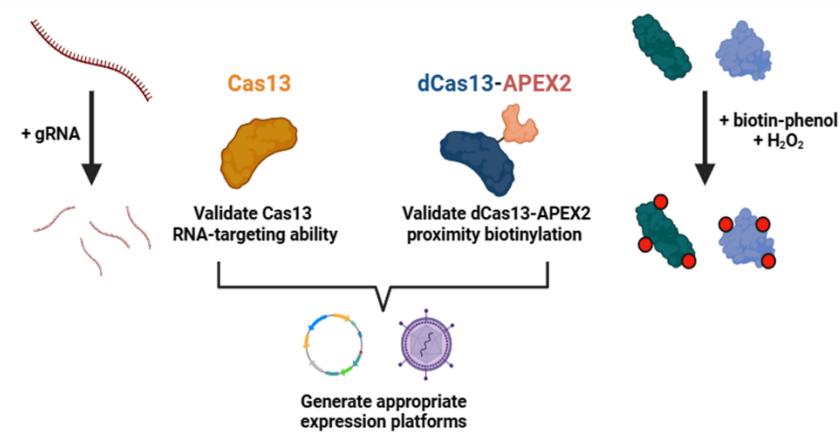
Another option is to target the cellular processes that create AR-Vs, however the alternative splicing mechanisms that dictate these are greatly complex and subject to extensive regulation. This work will aim to utilise novel RNA-targeting CRISPR-Cas13 and protein tagging with APEX2 proximity biotinylation to acquire further knowledge of these processes.

Therefore, the main objectives of this PhD project are:

- Development of CRISPR-Cas13 based approaches for targeting AR-V7 mRNA transcripts with an appropriate gRNA
- Optimisation of proximity biotinylation methods using APEX2 for protein biotin-tagging and affinity enrichment via streptavidin
- Combined application of the above methods with fusion protein dCasRx-APEX2 as a way to identify protein interactors of AR-V7 CE3 mRNA in CRPC cell line CWR22Rv1
- Utilisation of *in silico* approaches for selection of relevant proteins identified by the above techniques
- Validation of identified proteins as regulators of AR-V7 mRNA splicing through experimental manipulation

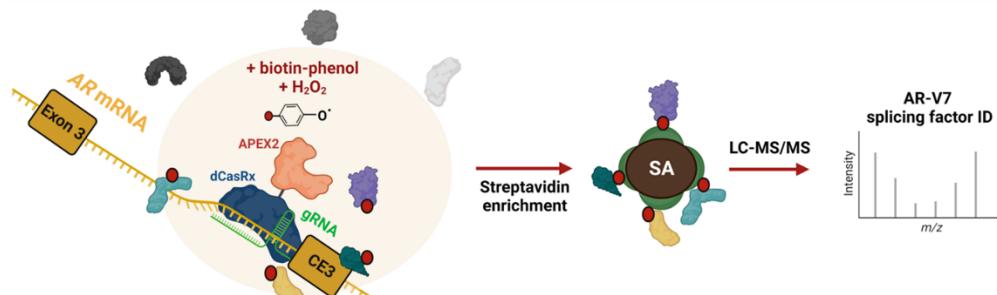
The overall aims and strategy of this PhD project, described above, are summarised visually in Figure 2.1. Figure 2.1 also denotes in which chapter number relevant results can be found.

Chapter 4  
Initial technical  
optimisation of methods



Chapter 5

Proteomic application in CRPC cell line CWR22Rv1  
for AR-V7 splicing factor discovery



Chapter 6

Analytical selection  
and validation of splicing factors

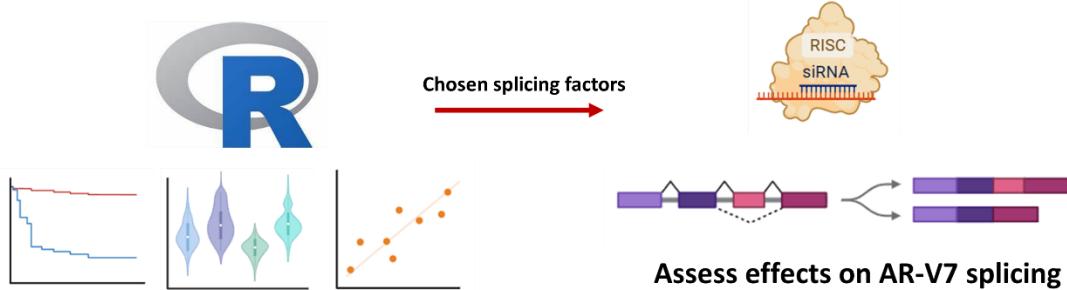


Figure 2.1 - Graphical summary of PhD project aims and objectives, with relevant chapter numbers noted

The overall project strategy is to achieve technical optimisation of Cas13 RNA-targeting and APEX2 proximity biotinylation approaches, followed by application of these methods in CRPC cell line CWR22Rv1 for identification of putative AR-V7 splicing factors. Finally, this list of candidate proteins will be refined by in silico analyses, before splicing factors are to be validated by siRNA knockdown and assessment of resulting effects on AR-V7 splicing.

## Chapter 3 - General materials and methods

All catalogue numbers are supplied at first reference to use of reagent in the document, after which they are omitted. Methods listed in this section are those that are applicable throughout the project. Methods specific to a particular results chapter are detailed in the relevant sections.

### 3.1 Cell lines and cell culture reagents

All cell lines were maintained in 5% CO<sub>2</sub> at 37°C. 10% foetal bovine serum (FBS) (Thermo A5256701) and penicillin-streptomycin (Sigma P0781) were added to culture media. Cells were tested quarterly for mycoplasma contamination. Cells were passaged using trypsin-EDTA solution (Sigma T4174) according to standard protocols for adherent cell lines. Specific media conditions for all cell lines used are listed in Table 3.1. Unless specified, all experiments were performed in media conditions listed above. For steroid depletion experiments, FBS was substituted for dextran-coated charcoal stripped FBS (VWR S181F-500). Enzalutamide (Selleckchem S1250) was dissolved in DMSO (Sigma D5879) and used at a final concentration of 10 µM.

Cell line	Media
CWR22Rv1 (ATCC CRL-2505)	RPMI-1640 (Sigma R8758)
VCaP (ATCC CRL-2876)	DMEM (Sigma D6171) supplemented with 2mM L-glutamine (Sigma G7513)
HEK293FT (Thermo R70007)	RPMI-1640 (Sigma R8758)

Table 3.1 - Culture media conditions for cell lines

### 3.2 Protein harvest and western blotting

Experiments performed in 6-well plates were lysed and harvested in 120 µl SDS sample buffer mixed 9:1 with β-mercaptoethanol (Sigma M3148). Lysates were boiled at 100°C for 10 minutes, before SDS-PAGE was performed by use of a stacking gel cast above a 10% acrylamide resolving gel. Samples were run alongside Spectra™ Multicolor Broad Range Protein Ladder (Thermo 26623). Proteins were transferred for 1 hour at 100 V or overnight at 30 V in transfer buffer onto Amersham™ Protran® nitrocellulose membrane (Sigma GE10600016). All steps were performed using a Mini-PROTEAN® electrophoresis and blotting system (Bio-Rad 1658029). Solutions used for PAGE can be found in Table 3.2.

Solution	Composition
SDS sample buffer	<b>125 mM Tris-HCl (pH6.8) (Fisher 10785341)</b> <b>5% SDS (Sigma L3771)</b> <b>10% glycerol (VWR 24386.298)</b> <b>0.01% bromophenol blue (Sigma B5525)</b>
<b>10% acrylamide resolving gel</b>	<b>3.33 ml 30% acrylamide (Sigma A3699)</b> <b>1.67 ml diH<sub>2</sub>O</b> <b>5 ml 2x buffer A (750 mM Tris-HCl (pH 8.8), 0.2% SDS)</b> <b>20 µl TEMED (Sigma T9281)</b> <b>100 µl APS (Sigma A3678)</b>
<b>Stacking gel</b>	<b>840 µl 30% acrylamide</b> <b>1.67 ml diH<sub>2</sub>O</b> <b>5 ml 2x buffer B (250 mM Tris-HCl (pH 6.8), 0.2% SDS)</b> <b>5 µl TEMED</b> <b>50 µl APS</b>
<b>Running buffer</b>	<b>25 mM Tris</b> <b>190 mM glycine (Fisher 10070150)</b> <b>0.1% SDS</b>
<b>Transfer buffer</b>	<b>25 mM Tris-HCl (pH8.3)</b> <b>150 mM glycine</b> <b>10% methanol (Fisher 10675112)</b>

Table 3.2 - Buffers used for protein harvest, PAGE and western blot transfer

Transfers were stained using Ponceau S solution (Sigma 7170) where indicated in figures. Ponceau was destained with 0.1 M NaOH (Sigma S5881). Membranes were blocked with 5% (w/v) non-fat milk (Marvel)/TBS (500 mM NaCl (Sigma S9888)), 20 mM Tris-HCl (pH 7.5)) for 1 hour at room temperature, before overnight incubation at 4°C in primary antibody diluted in 1% (w/v) non-fat milk/TBST (TBS + 0.1% Tween 20 (Sigma P1279)). Primary antibodies used for western blot are listed in Table 3.3. Membranes were then washed for 3 x 5 minutes in TBST and incubated for 1 hour at room temperature in 1:4,000 goat-anti mouse (Bethyl A90-516P) or swine anti-rabbit (Dako P0217) HRP-conjugated secondary antibody diluted in 1% (w/v) non-fat milk/TBST. Membranes were washed again for 3 x 5 minutes in TBST, before signal was developed using Clarity™ Western ECL Substrate (Bio-Rad 1705061) and imaged using a ChemiDoc™ system (Bio-Rad).

Antibody	Species	Supplier/catalogue no.	Dilution
HA	Mouse	Santa Cruz/sc-7392	1:1,000
AR NTD	Mouse	Dako/M3562	1:1,000
AR-V7	Rabbit	Abcam/ab198394	1:1500
AR-V7	Rabbit	RevMAb/31-1109-00	1:1,000
TRA2A	Mouse	Novus biologicals/H00029896-B01P	1:500
TRA2B	Rabbit	Abcam/ab31353	1:1,000
Biotin	Mouse	Santa Cruz/sc-101339	1:50
$\alpha$ -tubulin	Mouse	Sigma/T9026	1:4,000

Table 3.3 - Primary antibodies used for western blot

### 3.3 RNA extraction and RT-qPCR

RNA was extracted and purified by TRIzol™ reagent (Thermo 15596026) or GenElute™ mammalian total RNA miniprep kit (Sigma RTN350). For TRIzol™ extraction, manufacturer instructions were followed except for the following changes: i) during RNA precipitation with isopropanol (Sigma 34863), 2-3  $\mu$ l GlycoBlue™ coprecipitant (Thermo AM9515) was added to the aqueous phase/isopropanol mixture, ii) RNA was precipitated at -20°C overnight, and iii) RNA was washed 2-3 times in 1 ml 75% ethanol. For extractions performed using the GenElute™ method, manufacturer instructions for RNA extraction from 'Attached cell cultures' were followed. Concentration and purity of extracted RNA was measured on a NanoDrop® ND-1000 (Thermo) spectrophotometer.

cDNA was generated by reverse transcription of RNA in a 20  $\mu$ l reaction consisting of 4  $\mu$ l 5x MMLV-RT buffer (Promega M531A), 2  $\mu$ l 4mM dNTP mixture (Bioline BIO-39044), 1  $\mu$ l 100  $\mu$ g/ml oligo(dT)<sub>15</sub> (Promega C1101), 1  $\mu$ l 100  $\mu$ g/ml random primers (Promega C1181), 0.3  $\mu$ l MMLV-reverse transcriptase (Promega M170A), and 500 ng - 1  $\mu$ g RNA + nuclease-free water (Thermo 10977) to a final volume of 20  $\mu$ l. Reverse transcription reactions were performed for 1 hour at 37°C, before incubation at 100°C for 10 minutes. Resulting cDNA products were then diluted 1:5 - 1:10 in nuclease-free water.

qPCR analysis of cDNA was performed using 384-well plates on a QuantStudio™ 7 Flex Real-Time PCR System (Thermo 4485701). 10  $\mu$ l reactions were analysed consisting of 5  $\mu$ l 2x SYBR™ Green PCR Master Mix (Thermo 4309155) or 2x PowerTrack™ SYBR Green Master Mix (Thermo A46113), 0.4  $\mu$ l of each of forward and reverse 25 ng/ $\mu$ l primer, 2.2  $\mu$ l nuclease-free

water and 2 µl cDNA. For calculation of relative expression between samples, Ct values were exported and the Livak method was performed with the *ddCt* R package (v1.50.0, Zhang, Biczok and Ruschhaupt, 2022) using *RPL13A* as a housekeeping gene for normalisation (Livak and Schmittgen, 2001). Primers used for qPCR analysis are listed in Table 3.4.

Target gene	Forward primer sequence (5' - 3')	Reverse primer sequence (5' - 3')
<i>RPL13A</i>	CCTGGAGGAGAAGAGGAAAGAGA	TTGAGGACCTCTGTGTATTTGTCAA
<i>AR-FL</i>	AACAGAAGTACCTGTGCGCC	TTCAGATTACCAAGTTCTTCAG
<i>AR-V7</i>	AACAGAAGTACCTGTGCGCC	TCAGGGTCTGGTCATTTGA
<i>AR-V1</i>	AACAGAAGTACCTGTGCGCC	TGAGACTCCAAACACCCCTCA
<i>AR-V6</i>	AACAGAAGTACCTGTGCGCC	TATGACACTCTGCTGCCTTGC
<i>AR-V9</i>	AACAGAAGTACCTGTGCGCC	GCAAATGTCTCCAAAAGCAGC
<i>AR CE3 pre-mRNA</i>	GCCTGCTAGATACAAGCCCG	AGCCTTCTTCAGGGTCTGG
<i>TRA2A</i>	TTTGGAAACCCCTGATGGAC	AAAACAACCTCGAGGGCAGA
<i>TRA2B</i>	ATCCGTGAGCACTCCACTT	GCGTCACATCCGGTAGAGTT
<i>THRAP3</i>	GCCGATCTCTCTCGTTCA	TTGGGTGGTTCTTCTGTAA
<i>SART1</i>	AACGTGAACCTGGTGGATAAG	CTTCGTCATACTGGACAGGATAG
<i>TP53</i>	CAGCACATGACGGAGGTTGT	TCATCCAAATACTCCACACGC

Table 3.4 - Primers used for qPCR analysis

### 3.4 Plasmid transfection

Unless otherwise specified, all plasmids used throughout were reverse transfected at the amounts indicated with *TransIT*®-LT1 transfection reagent (MIR 2300) according to manufacturer instructions using a 3:1 (µl:µg) LT1:plasmid ratio. Opti-MEM™ I (Thermo 31985062) was used as serum-free media for creation of transfection complexes. Where indicated, expression of eGFP was assessed by imaging cells with a Nikon™ TE2000 fluorescence microscope (Nikon Corporation).

### 3.5 CasRx gRNA design

All CasRx gRNAs were designed using cas13design (available at <https://cas13design.nygenome.org/>) (Wessels, Méndez-Mancilla *et al.*, 2020) and ordered as single custom RNA oligos (Sigma) comprising the CasRx gRNA 30 nt direct repeat (DR)

sequence, 5' - AACCCCUACCAACUGGUCGGGUUUGAAC - 3', followed by a 22-23 nt spacer with complementarity to mRNA target sequence(s).

### **3.6 Statistical analyses and plotting of results**

Specific tests used for determination of statistical significance and parameters used for plotting of results are listed in the relevant figure legends. All statistical tests were performed using GraphPad Prism (v9.3.1) or R (v4.1.2, R Core Team, 2021) via the RStudio development environment (v2022.07.0, RStudio Team, 2022). All results, unless otherwise specified, were plotted using GraphPad Prism (v9.3.1) or R (v4.1.2, R Core Team, 2021) via the RStudio development environment (v2022.07.0, RStudio Team, 2022). The *tidyverse* software suite (v2.0.0, Wickham *et al.*, 2019) was used for data processing and plotting in R throughout. The *EnhancedVolcano* package (v1.12.0, Blighe, Rana and Myles, 2021) was used for creation of all volcano plots. The *pheatmap* package (v1.0.12, Kolde, 2019) was used for creation of all heatmaps.

## Chapter 4 - Initial technical optimisation and pilot experiments

### 4.1 Introduction and rationale

Development of Cas13 and proximity biotinylation techniques was performed using the Cas13 subtype RfxCas13d (CasRx) and engineered ascorbate peroxidase APEX2. The ability of the former to bind mRNA as directed by a complementary gRNA will be combined with APEX2 proximity biotinylation to covalently tag the local protein interactome of a desired sequence, in this case CE3 of AR-V7 mRNA.

CasRx was first isolated from *Ruminococcus flavefaciens* and presents numerous advantages over other Cas13 proteins for this project including higher potency RNA knockdown, more compact size, and the preexisting application of HEPN-mutated ‘dead’ CasRx (dCasRx) for gRNA-directed RNA binding without cleavage (Konermann *et al.*, 2018; Wessels, Méndez-Mancilla *et al.*, 2020). As a class 2 CRISPR protein, CasRx comprises a single-effector system whereby one protein is responsible for gRNA-mediated target binding and cleavage as opposed to class 1 CRISPR nucleases consisting of multiple Cas subunits (Konermann *et al.*, 2018; Zhang *et al.*, 2018; Shmakov *et al.*, 2017; Tang, 2019). CasRx has demonstrated applicability as a potent tool for RNA knockdown, with lower propensity for off-target effects when compared with RNAi (Konermann *et al.*, 2018). This has shown utility *in vitro* using a variety of eukaryotic cell types. Furthermore, the possibility of *in vivo* knockdowns by adenoviral delivery or embryonic injection has been demonstrated (Konermann *et al.*, 2018; Mahas, Aman and Mahfouz, 2019; Zhou *et al.*, 2020; Buchman *et al.*, 2020; Kushawah *et al.*, 2020).

There is however emerging evidence that despite high on-target ability, gRNA binding to target RNA mediates an activated CasRx nuclease conformation that is capable of cleaving nearby non-complementary RNAs in a so-called ‘bystander effect’. The resulting collateral damage to RNA can lead to widespread toxicity, with lethal effects *in vivo* (Ai, Liang and Wilusz, 2022; Li *et al.*, 2023). Given the aims of this project, off-target effects are not a prominent concern as CasRx will be used primarily as a tool for selecting the most appropriate gRNAs for proteomics workflows. Once optimised, dCasRx, shown to bind RNA without

cleavage activity, will be utilised to target APEX2 constructs to an appropriate mRNA region for biotinylation of protein interactors (Konermann *et al.*, 2018; Han *et al.*, 2020; Huynh *et al.*, 2020).

Proximity biotinylation using APEX2 is advantageous to this project for two primary reasons: i) its small size at 27 kDa is comparable to GFP, a protein widely used in fusion constructs and demonstrated as appropriate for multiprotein chimeras (Snapp, 2005); ii) rapid reaction kinetics, as the biotin-phenoxyl radical has a < 1 ms half-life in water compared to several minutes as for the biotinyl-AMP substrate generated by first generation enzyme BiOID2, thus reducing off-target tagging of distant proteins (Hung *et al.*, 2016; Chen and Perrimon, 2017). These characteristics make APEX2 an attractive choice for biotinylation of proteins in a tightly defined RNA region. Previous proximity biotinylation studies have often focused on protein enrichment in entire subcellular compartments, or all the cellular interactors of a specific protein (Hung *et al.*, 2017; Gillingham, Bertram *et al.*, 2019; Gupta *et al.*, 2018). However, for this investigation into CE3-specific splicing factors the spatial resolution offered by APEX2 is crucial.

APEX2 has been utilised for sequence-specific DNA and RNA interactome analyses previously, providing an informative proof of concept for this workflow. Myers *et al.* applied this to HEK293 cells using a dCas9-APEX2 fusion protein containing a T2A self-cleaving peptide upstream of GFP to select for cells expressing dCas9-APEX2. Targeting of dCas9APEX2 to the *MYC* promoter, followed by proximity biotinylation and subsequent analysis of protein enrichment over an untargeted control successfully captured multiple proteins previously shown to activate *MYC* gene transcription including HUWE1, RUVBL1 and ENO1 (Myers *et al.*, 2018). Gao *et al.* performed a similar workflow with dCas9-APEX2 in osteosarcoma cell line U2OS, where use of a telomere-targeting gRNA effectively captured a range of known telomere interactors (Gao *et al.*, 2018).

Analogous techniques have also shown success with dCas13 fusions, though only two such examples have been published. Lin *et al.* targeted a dCas13bAPEX2 fusion construct to the U1 snRNA, after which gene ontology (GO) analysis confirmed, as expected, a statistical enrichment of proteins related to spliceosomal processes (Lin *et al.*, 2021). Furthermore, Han *et al.* used workflows most comparable to that proposed here, in which dCasRx was fused with APEX2. By targeting their fusion construct to human telomerase RNA, they were able to

identify RNA N<sup>6</sup>-methyladenosine (m<sup>6</sup>A) demethylase ALKBH5 as a novel epitranscriptomic interactor that influences human telomerase RNA stability and enzymatic activity (Han *et al.*, 2020). These technical proof of concept experiments using dCas13APEX2 fusions were both performed in HEK293 cells, which uptake and express exogenous constructs with remarkable efficiency (Thomas and Smart, 2005). The aim of this PhD project is to answer, for the first time, mechanistic questions in an oncologically relevant cell line using a dCasRx-APEX2 fusion protein to identify *AR* CE3-interacting proteins that may regulate splicing fate decision in CRPC, ultimately generating AR-V7.

In order to apply these novel biotechnologies, an initial series of validation experiments was performed. The priorities here were: i) validation that expression of CasRx/dCasRx-APEX2 transgenes is detectable by western blotting and GFP microscopy for downstream cell selection; ii) testing of a published gRNA design algorithm for appropriate targeting of CasRx/dCasRx-APEX2 constructs to RNA(s) of interest; and iii) demonstration that the dCasRx-APEX2 fusion is capable of mediating live in-cell biotinylation of proteins for subsequent enrichment by streptavidin pulldown. Furthermore, development of an appropriate expression construct was undertaken to enable use of these techniques in the CRPC cell line CWR22Rv1, our cell line of choice due to high expression of both AR-FL and AR-Vs. To support this process, a range of protein structural prediction software was employed to provide confidence in fusion protein design.

## 4.2 Specific materials and methods

All catalogue numbers are supplied at first reference to use of reagent in the document, after which they are omitted.

### 4.2.1 Plasmids and molecular cloning

CasRx pXR001 (#109049) and dCasRx pXR002 (#109050) expression plasmids were purchased from Addgene. dCasRx-APEX2 expression plasmid pXR002-APEX2 was created by digestion and ligation of a *BamHI* 5'/3' flanked gBlock™ (IDT) encoding the APEX2 ORF into the *BamHI* site of pXR002, resulting in a N-dCasRx-APEX2-C fusion.

CasRx(VB) and dCasRx-APEX2(VB) were purchased from VectorBuilder™. CasRx(TLCV2) and dCasRx-APEX2(TLCV2) were created by digestion of Cas9 vector TLCV2 (Addgene #87360) with Anza™ restriction enzymes *BshTl* (Thermo IVGN0074) and *BamHI* (Thermo IVGN0056), followed by removal of the Cas9 ORF and gel extraction of the TLCV2 backbone using the Monarch® DNA Gel Extraction Kit (NEB T1020). CasRx(VB) and dCasRx-APEX2(VB) were also digested with Anza™ *BshTl* and *BamHI* enzymes, before CasRx and dCasRx-APEX2 ORFs were gel extracted as above and ligated into the TLCV2 backbone using T4 ligase (NEB M0202). Ligation reactions were subsequently transformed into *E. coli*. All bacterial transformations for cloning products were performed in Stbl3 prepared using the Mix and Go!™ *E. coli* Transformation Kit and Buffer Set (Zymo Research T3002). Bacterial colonies were propagated in LB with selection antibiotic before plasmid purification by miniprep (Thermo K210011) or maxiprep (Sigma PLEX15). Plasmid sequences were verified by Sanger sequencing (GENEWIZ).

### 4.2.2 Protein structure prediction and modelling

All listed protein structures were predicted based on the amino acid FASTA sequence from the relevant plasmids as input, which included all residues from the initiating methionine up to the glycine found at the PG of the T2A cleavage site. Structural prediction software I-TASSER, trRosetta, PrDOS, AlphaFold 2 and RoseTTAFold were run using default settings on the websites listed in Table 4.1. PDB file outputs were imported to NCBI iCn3D, available at <https://www.ncbi.nlm.nih.gov/Structure/icn3d/> (Wang *et al.*, 2022), for structure visualisation and conserved domain analysis.

Software	Authors	URL
I-TASSER	Roy, Kucukural and Zhang, 2010	<a href="https://zhanggroup.org/I-TASSER/">https://zhanggroup.org/I-TASSER/</a>
trRosetta	Du <i>et al.</i> , 2021	<a href="https://rosetta.bakerlab.org/">https://rosetta.bakerlab.org/</a>
PrDOS	Ishida and Kinoshita, 2007	<a href="https://prdos.hgc.jp/">https://prdos.hgc.jp/</a>
AlphaFold 2	Jumper <i>et al.</i> , 2021 (software) Kuták <i>et al.</i> , 2022 (web server)	<a href="https://catana.ait.ac.at/">https://catana.ait.ac.at/</a>
RoseTTAFold	Baek <i>et al.</i> , 2021	<a href="https://rosetta.bakerlab.org/">https://rosetta.bakerlab.org/</a>

Table 4.1 - Summary of software and corresponding websites used for protein structure prediction

#### 4.2.3 CasRx gRNA transfection in HEK293FT

HEK293FT were transfected in 6-well plates with 1 µg CasRx expression plasmid pXR001 and incubated for 48 hours before being transfected with CasRx gRNAs at the indicated concentrations using 0.2% lipofectamine™ RNAiMAX (Thermo 13778150) according to manufacturer instructions. Opti-MEM™ I was used as serum-free media for creation of transfection complexes. 48 hours after gRNA transfection, RNA was extracted and RT-qPCR was performed as described in Section 3.3. All gRNA sequences used for targeting of *TP53* mRNA in HEK293FT are listed in Table 4.2.

CasRx gRNA	Sequence (5' - 3')
non-targeting (NT)	AACCCCUACCAACUGGUUCGGGGUUUGAAACUCACCAGAACGUACCAUACUC
TP53 g1	AACCCCUACCAACUGGUUCGGGGUUUGAAACCCACACGCAAUUUCCUCCACU
TP53 g2	AACCCCUACCAACUGGUUCGGGGUUUGAAACGCAAACAUUUCGUUGAGGGCAG
TP53 g1	AACCCCUACCAACUGGUUCGGGGUUUGAAACCUGGGACGGAACAGCUUUGAGGU

Table 4.2 - CasRx gRNA sequences used for *TP53* mRNA targeting in HEK293FT

#### 4.2.4 dCasRx-APEX2 protein proximity biotinylation using transient plasmid expression

HEK293FT or CWR22Rv1 cells were transfected with pXR002-APEX2 and incubated for 48 hours. 500 mM biotin-phenol (Iris Biotech LS-3500) in DMSO was diluted to 500 µM in culture media and vortexed until fully dissolved. Cells were incubated in culture plates containing media + 500 µM biotin-phenol for the timeframes indicated at 37°C and 5% CO<sub>2</sub>, before

hydrogen peroxide (Sigma H1009) was added to a final concentration of 1mM and plates were gently swirled by hand for the timeframes indicated at room temperature. Media was then quickly aspirated and replaced with APEX2 quenching buffer (100 mM sodium ascorbate (Sigma A7631), 10 mM TROLOX (Sigma 238813) and 10mM sodium azide (VWR 786-299) in PBS). APEX2 quenching buffer was removed and replaced 3 times for a total of 4 x quenching washes before downstream processing.

#### **4.2.5 Biotinylated protein pulldown with streptavidin (whole-cell lysates)**

For whole-cell lysis, cells were scraped into APEX2 quenching buffer and spun at 500 x g for 10 minutes. Plates were then scraped and washed in additional APEX2 quenching buffer, which was mixed with cell pellets and spun again at 500 x g for 10 minutes. Pellets were either snap-frozen in LN<sub>2</sub> and stored at -80°C or lysed immediately. Cell lysis was carried out using RIPA buffer (50mM Tris, 150mM NaCl, 0.1% w/v SDS, 0.5% w/v sodium deoxycholate (Sigma D6750), 1% v/v Triton X-100, pH 7.5) supplemented with protease inhibitor cocktail (Sigma 5892791001). Pellets were lysed by resuspension in 250 – 500 µl RIPA buffer and rocking on ice for 30 minutes, before lysates were spun at 15,000 x g for 15 minutes at 4°C. The resulting supernatant containing soluble lysate was then transferred to a fresh tube for downstream applications.

Protein concentration in lysates was determined by Pierce™ 660nm protein assay (Thermo 22662) according to manufacturer 'microplate procedure'. Samples were diluted in PBS, and a pre-diluted set of BSA protein standards (supplied with kit) was used to generate a standard curve for calculation of unknown concentrations using  $y = mx + C$ . RIPA diluted 1:10 in PBS was used as a blank. 10 – 50 µg of each sample was retained as an input lysate, which was mixed 5:1 with 6x SDS loading buffer (10.5ml diH<sub>2</sub>O, 10.5ml 1M Tris-HCl (pH 6.8), 10.8ml glycerol, 3g SDS, 2.79g DTT (Sigma 1019777001), 3.6mg bromophenol blue) and boiled at 100°C for 10 minutes.

The remaining biotinylated protein lysates were enriched using Pierce™ streptavidin magnetic beads (Thermo 88817). Streptavidin beads and biotinylated protein lysates were used at a 1:1 µg bead:protein ratio. A volume of beads (stock concentration 10 µg/µl) appropriate to this ratio was taken and washed twice with 1 ml ice-cold RIPA buffer. Biotinylated protein lysates were then added to washed beads + 500 µl RIPA buffer supplemented with protease inhibitor

cocktail. Bead/lysate mixtures were rotated overnight at 4°C. The next day, beads were washed 7 x 1 ml in ice-cold RIPA buffer.

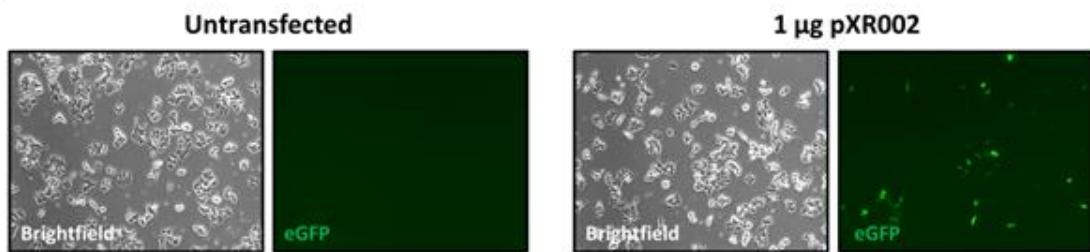
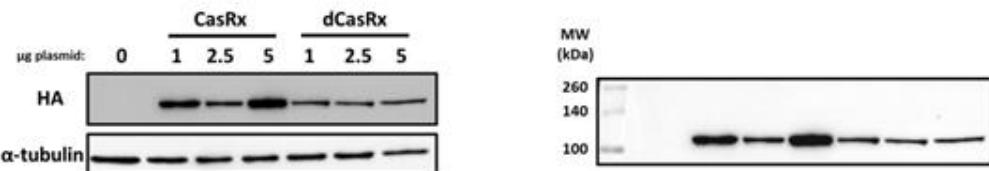
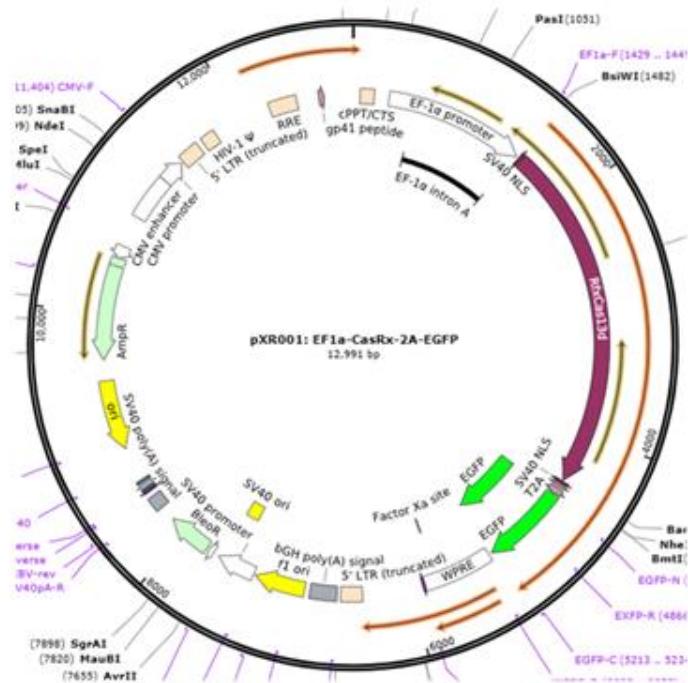
For western blot analysis of enriched proteins, washed beads were resuspended in 40 µl 3x SDS loading buffer (6x SDS loading buffer diluted 1:2 in diH<sub>2</sub>O) supplemented with 2 mM biotin (Sigma B4501) and 20 mM DTT and boiled at 100°C for 15 minutes. Boiled samples were vortexed, cooled on ice and eluted protein samples were transferred to fresh tubes. Where indicated, a portion of lysate was taken after overnight incubation with streptavidin beads, mixed 5:1 with 6x SDS loading buffer and boiled 100°C for 10 minutes. This represented the ‘flow-through’, or protein sample that did not bind with streptavidin. Eluted samples were subsequently run alongside inputs and/or flow-through by SDS-PAGE as detailed in Section 3.2.

SDS-PAGE samples were analysed for protein biotinylation either by methods detailed in Section 3.2 with an anti-biotin antibody (Table 3.3) or by use of streptavidin-HRP (Abcam ab7403). For the latter, protein-transferred membranes were blocked with 3% BSA/TBST for 1 hour at room temperature. Membranes were then probed for 1 hour at room temperature with 1:2,500 streptavidin-HRP in 3% BSA/TBST, washed for 4 x 5 minutes in TBST and imaged as in 3.2.

### 4.3 Development of CasRx/dCasRx-APEX2 expression constructs

Initial validation of CasRx/dCasRx expression plasmids pXR001 and pXR002 was undertaken. These plasmids contain a CasRx/dCasRx ORF driven by the EF1- $\alpha$  promoter, as well as N and C-terminal SV40 NLS sequences and a C-terminal HA tag for nuclear localisation and HA immunodetection, respectively (Figure 4.1A). pXR001 and pXR002 differ only by four mutations in the ORF of the latter that catalytically inactivate the HEPN domains: R239A, H244A, R858A, H863A.

CWR22Rv1 cells were transfected in 6-well plates with varying amounts of CasRx/dCasRx expression plasmids pXR001/pXR002 and protein level assayed by western blot. This revealed that increasing plasmid amounts up to 5  $\mu$ g had no appreciable effect on transgene expression as detected using an HA antibody (Figure 4.1B, left panel). CasRx/dCasRx protein was also detected at the expected molecular weight of 112 kDa (Konermann *et al.*, 2018) (Figure 4.1B, right panel). Furthermore, the T2A eGFP of this plasmid functions effectively, as live-cell fluorescence microscopy demonstrated detectable GFP signal in transfected cells. However transfection efficiencies were somewhat low and the extent of transfection was variable between cells (Figure 4.1C). Some cells demonstrated no plasmid uptake, whilst others express high levels of transgene based on eGFP levels. Due to this variability and relatively low uptake efficiency, as has been observed in CWR22Rv1 by others in our group, lentiviral transduction and creation of stable cell lines was explored.



**Figure 4.1 - Expression of CasRx/dCasRx plasmids is detectable by western blot and GFP microscopy**

**A.** Plasmid map of CasRx expression vector pXR001 highlighting key components, sourced from Addgene.org (restriction site/primer binding site annotations partially removed for display purposes) **B.** Varying amounts of CasRx/dCasRx expression vectors pXR001/pXR002 were transfected into CWR22Rv1 cells in 6-well plates. 72 hours later, protein lysates were harvested for analysis of HA tag expression by western blot.  $\alpha$ -tubulin was used as a loading control **C.** Live-cell GFP microscopy was used to assess plasmid transfection efficiency in CWR22Rv1

Multiple criteria were required in an appropriate lentiviral plasmid for use in CWR22Rv1. First, any such plasmid would need a puromycin resistance gene for selection of stably integrated cells. Second, tight expression control and doxycycline inducibility was desirable to prevent

any potential toxicities from excessive Cas13 accumulation as has been observed previously (Ai, Liang and Wilusz, 2022; Wu and Kapfhammer, 2021). Finally, retaining the T2A-eGFP ORF downstream from CasRx/dCasRx-APEX2, as present in pXR001/pXR002, would enable FACS selection of inducible cells after puromycin treatment. All other features of pXR001/pXR002 including SV40 NLS and an HA tag would be retained. The above criteria would collectively enable creation of stable CWR22Rv1 populations for applying Cas13/APEX2 techniques (Figure 4.2).

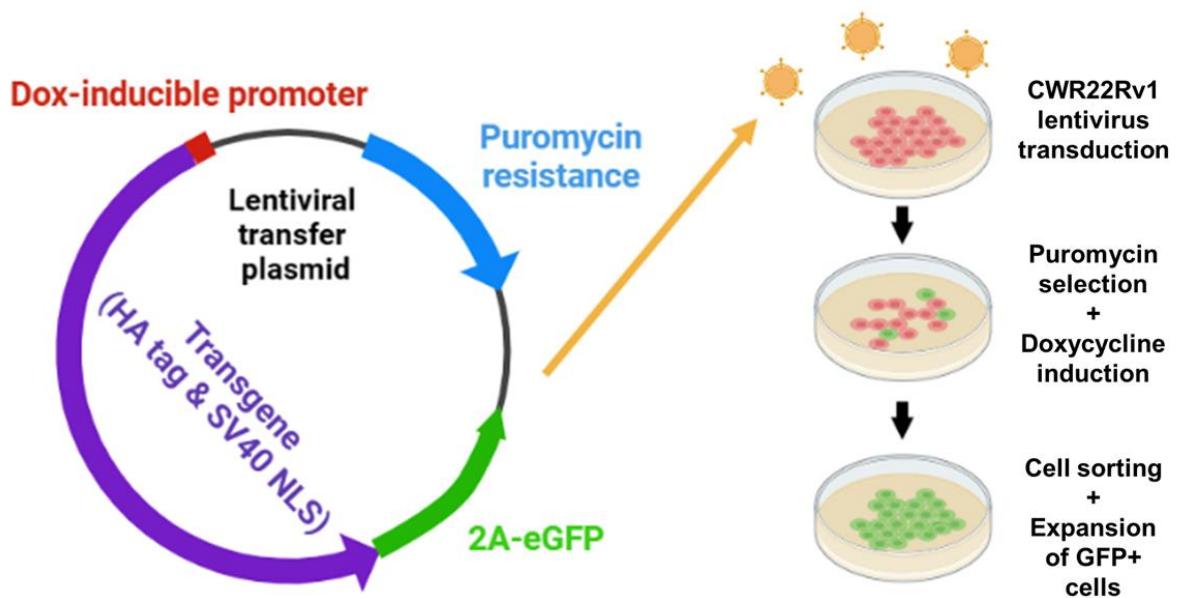


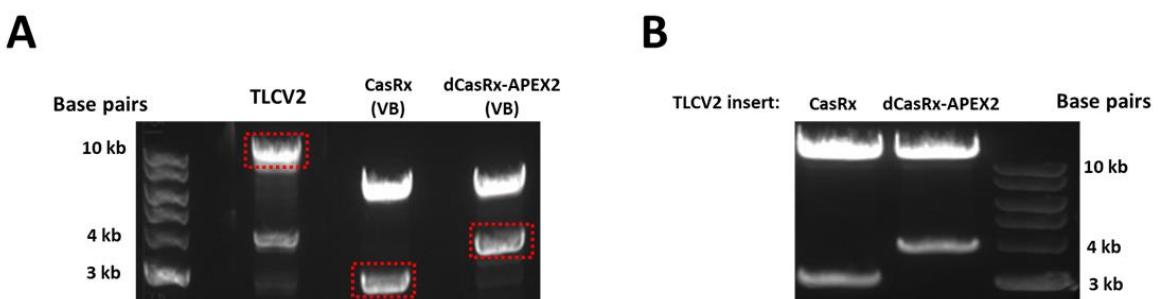
Figure 4.2 - Plasmid criteria and transgenic CWR22Rv1 generation strategy

Simplified summary of plasmid criteria for expression of transgenes in CWR22Rv1. Generation of transgenic cells would necessitate packaging of plasmids into lentivirus for CWR22Rv1 transduction, followed by puromycin selection and live-cell FACS to sort for successful doxycycline-inducible cells

Given these necessities, Cas9 expression vector TLCV2 was an appropriate choice and has been used by our lab previously to generate stable, doxycycline-inducible cells (data not shown). The Cas9 ORF of TLCV2 is flanked by *Age*I and *Bam*HI restriction sites enabling excision with these enzymes for insertion of novel transgenes into the plasmid. For our purposes, two novel transgenes would be introduced encoding. i) catalytically active CasRx, and ii) dCasRx-APEX2 fusion protein. Initially, a FLAG-APEX2 sequence was cloned into pXR002 downstream of the dCasRx C-terminus (performed by PI, Luke Gaughan, before start of project). This novel plasmid, pXR002-APEX2, would be applied in the initial optimisation of APEX2 proximity biotinylation workflows (discussed in Section 4.5).

However for insertion into TLCV2, pXR001 and pXR002-APEX ORFs for both CasRx and dCasRx-APEX2 required alteration due to internal *Age*I and *Bam*HI restriction sites. This was achieved by ordering expression plasmids with *Age*I/*Bam*HI-flanked ORFs from VectorBuilder™, in which redundant codons were used to remove internal *Age*I and *Bam*HI sites without any modification to amino acid sequence. These plasmids, CasRx(VB) and dCasRx-APEX2(VB) also enabled EF1- $\alpha$  promoter-driven expression with the same required features as described above, should an option for constitutive expression be desired.

First, TLCV2, CasRx(VB) and dCasRx-APEX2(VB) were digested with the *Age*I isoschizomer *Bsh*TI and *Bam*HI, and subsequently run on a 1% agarose gel supplemented with GelRed to enable size selection for DNA gel extraction. For subcloning, the TLCV2 backbone with Cas9 excised (~12.5 kb), CasRx ORF (~3 kb) and dCasRx-APEX2 ORF (~3.75 kb) were taken forward for ligation (Figure 4.3A). Following ligation, *Stbl3 E. coli* were transformed and colonies picked for screening by diagnostic digest with *Bsh*TI and *Bam*HI to confirm insertion of CasRx/dCasRx-APEX2 into TLCV2. This confirmed inserts were of the expected size and that digests yielded only two bands, the TLCV2 backbone and CasRx or dCasRx-APEX2 ORFs (Figure 4.3B). Presence of transgenes was subsequently verified by Sanger sequencing (not shown) thus yielding two novel plasmids, CasRx(TLCV2) and dCasRx-APEX2(TLCV2), which would be utilised in creation of transgenic CWR22Rv1 for CE3-proximal protein tagging workflows (discussed in Chapter 5).



**Figure 4.3 (previous page) - CasRx and dCasRx-APEX2 were successfully subcloned into TLCV2**

**A.** TLCV2, CasRx(VB) and dCasRx-APEX2(VB) were digested with *Bsh*TI/*Bam*HI, yielding the TLCV2 backbone and CasRx/dCasRx-APEX2 ORFs that were taken forward for ligation (bands highlighted in red) **B.** Novel plasmids CasRx(TLCV2) and dCasRx-APEX2(TLCV2) were digested with *Bsh*TI/*Bam*HI to confirm insert size

#### 4.4 Protein structure prediction modelling of dCasRx-APEX2 fusion protein

Paramount to the success of this project was effective functioning of dCasRx-APEX2 as a fusion protein, in which both dCasRx (112 kDa) and APEX2 (27 kDa) operate effectively for RNA binding and proximity biotinylation, respectively. In order to gain insight into the likely structural properties of our dCasRx-APEX2 fusion construct (which would be expressed by inducible plasmid dCasRx-APEX2(TLCV2), generation of which is described in Section 4.2), we leveraged a range of computational tools for protein structure prediction.

One of the primary considerations in a chimeric fusion protein is choice of inter-protein linker, which is crucial for achieving fusion partner separation and folding, and as such, independent functioning of the fusion moieties (Chen, Zaro and Shen, 2013). A range of linkers have been used in engineering of protein constructs, commonly involving small, non-polar amino acids such as glycine and serine that enable linker chain flexibility and freedom of protein movement/folding. Valuable information about linker design and properties can be derived from naturally occurring multidomain protein linkers. Analysis of these has revealed the amino acid preferences and average length (~6 - 15 aa) of these linker regions, aiding fusion construct design. This analysis revealed numerous amino acids that occur with comparable frequency to glycine and serine in protein linkers, presenting themselves as good candidates for selection in a dCasRx-APEX2 linker sequence (Argos, 1990).

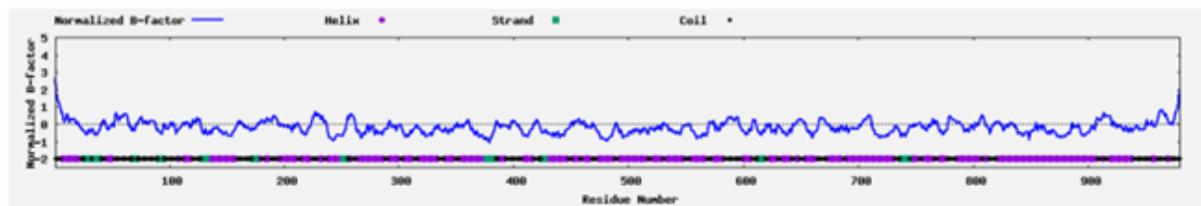
Next, a range of protein structure modelling tools were utilised in order to gauge secondary structure in dCasRx N/C-termini. Given that for many protein fusions a flexible linker is preferable to enable fusion partner independence (Chen, Zaro and Shen, 2013), fusing APEX2 to the N or C-terminus would have a greater chance of success if these termini had inherent flexibility and lacked secondary structure. A range of software were employed for prediction of disordered regions likely to be amenable for terminal fusions including RaptorX, I-TASSER, PrDOS, NetSurfP-2.0, IUPred3 and trRosetta (Wang *et al.*, 2016; Roy, Kucukural and Zhang, 2010; Ishida and Kinoshita, 2007; Klausen *et al.*, 2019; Erdős, Pajkos and Dosztányi, 2021; Du, Su *et al.*, 2021). These revealed the presence of disordered N and C dCasRx termini (Figure 4.4), of which a C-terminal fusion is favourable due to potential interference with cellular localisation signals that have been observed with N-terminal GFP fusions (Palmer and Freeman, 2004). Furthermore, numerous amino acids found near the C-terminus of dCasRx are found at a high frequency in natural linkers (Table 4.3).

Amino acid (3-letter code)	Natural linker propensity
Thr	1.55
<b>Ser</b>	<b>1.46</b>
Pro	1.35
<b>Gly</b>	<b>1.25</b>
<b>Asp</b>	<b>1.25</b>
<b>Lys</b>	<b>1.16</b>
Gln	1.13
<b>Asn</b>	<b>1.09</b>
Ala	1.05
<b>Val</b>	<b>1</b>
<b>Glu</b>	<b>0.87</b>
Arg	0.84
Ile	0.81
Tyr	0.75
Met	0.75
Phe	0.69
His	0.55
Cys	0.35
Trp	0.23
Leu	N/A

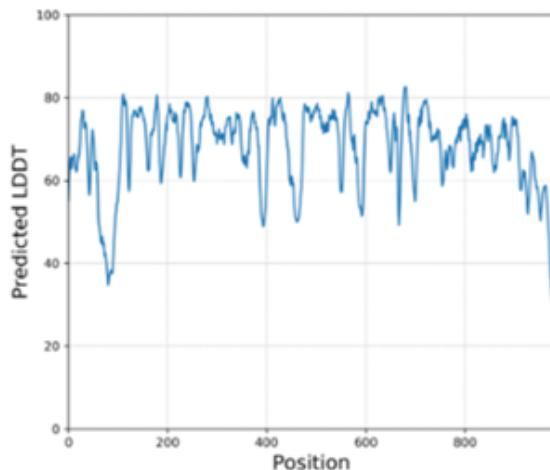
**Table 4.3 - Amino acid frequency of occurrence in naturally occurring multidomain protein linkers**

Analysis of amino acids found in natural multidomain protein linkers, undertaken by Argos (1990), provides an indication for fusion protein linker design. Values over 1 indicate higher than expected frequency. Amino acids in bold are found near the C-terminus of dCasRx

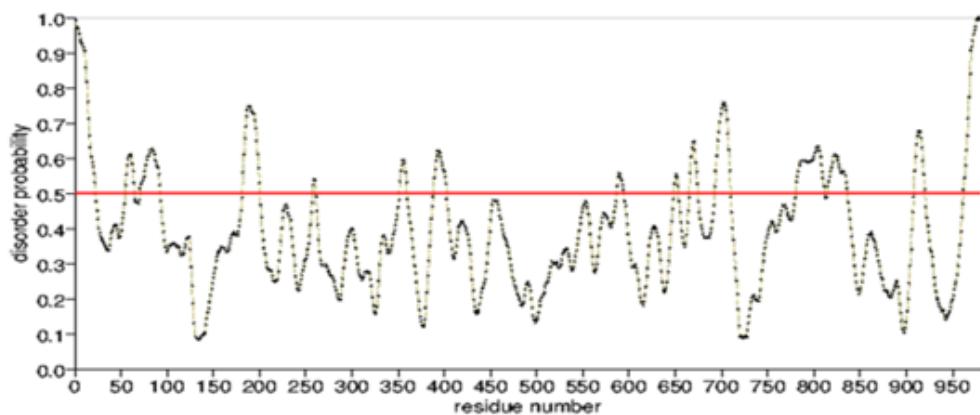
### I-TASSER (Roy, Kucukural & Zhang, 2010)



### trRosetta (Du, Su et al., 2021)



### PrDOS (Ishida & Kinoshita, 2007)



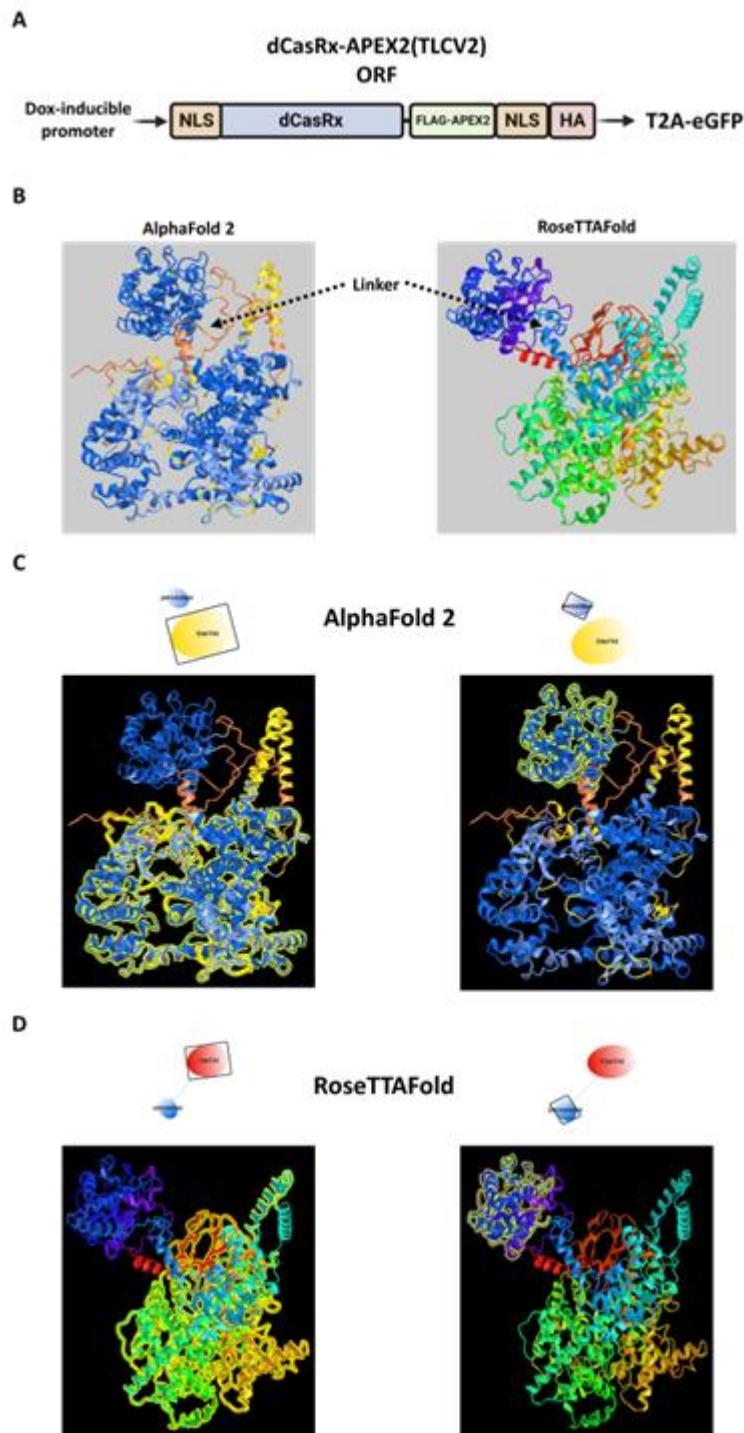
**Figure 4.4 - Protein disorder prediction of dCasRx reveals inherent N and C-terminal flexibility**

A range of protein modelling tools were used to assess likely secondary structures at dCasRx N and C-termini. Figures are imported directly from the relevant web servers. I-TASSER plot y-axis is normalized B-factor, a measure of inherent residue mobility. Values above 0 (dotted line) indicate greater amino acid mobility. trRosetta plot displays Predicted Local Distance Difference Test (pLDDT) values, where a lower value is indicative of a greater likelihood of disorder. PrDOS plot displays the likelihood of a region being disordered and lacking a fixed secondary structure, with values above 0.5 (red line) being higher confidence

In order to gain insight into the likely structural properties of our dCasRx-APEX2 fusion protein two further protein structure modelling tools were employed. AlphaFold 2 (AF2), developed by Google DeepMind and run here via CATANA server (Jumper *et al.*, 2021; Kuták *et al.*, 2022), and RoseTTAFold, created by the Baker Lab at the University of Washington, Seattle (Baek *et al.*, 2021). The former is widely considered the most accurate protein structure prediction tool to date, outperforming all rival software at the most recent Critical Assessment of Methods of Protein Structure Prediction (CASP) competition (Kryshtafovych *et al.*, 2021). Moreover, RoseTTAFold is the closest competitor to AF2 (Perrakis and Sixma, 2021).

The ORF of dCasRx-APEX2(TLCV2) contains a FLAG tag between dCasRx and APEX2 (Figure 4.5A), which provides additional function as a linker sequence given its sequence, DYKDDDDK, is rich in aspartate and lysine residues, both commonly found in natural linkers (Table 4.3). Visualisation of structural prediction Protein Data Bank (PDB) files from both AF2 and RoseTTAFold using NCBI iCn3D (Wang *et al.*, 2022) demonstrated a predicted separation between dCasRx and APEX2, with an unstructured linker region joining them as verified by the low pLDDT Predicted Local Distance Difference Test (pLDDT) score from AF2 which has been noted as a predictor of disorder and residue flexibility (Figure 4.5B) (Tunyasuvunakool *et al.*, 2021; Guo *et al.*, 2022; Goulet and Cambillau, 2022).

Further visualisation of conserved domains using Conserved Domain Database (CDD) analysis in iCn3D confirmed predicted ‘Cas13d’ and ‘peroxidase’ domains, consistent with the molecular functions of dCasRx and APEX2. Whilst highlighting of these domain predictions additionally demonstrates functional separation in our fusion construct (Figure 4.5C, D). Collectively, these analyses using the latest advances in protein structure prediction offer confidence in the likely functioning of dCasRx-APEX2 as a fusion protein, in which both proteins operate effectively.



**Figure 4.5 - Protein structure modelling of dCasRx-APEX2 indicates a separation of functional fusion protein domains**

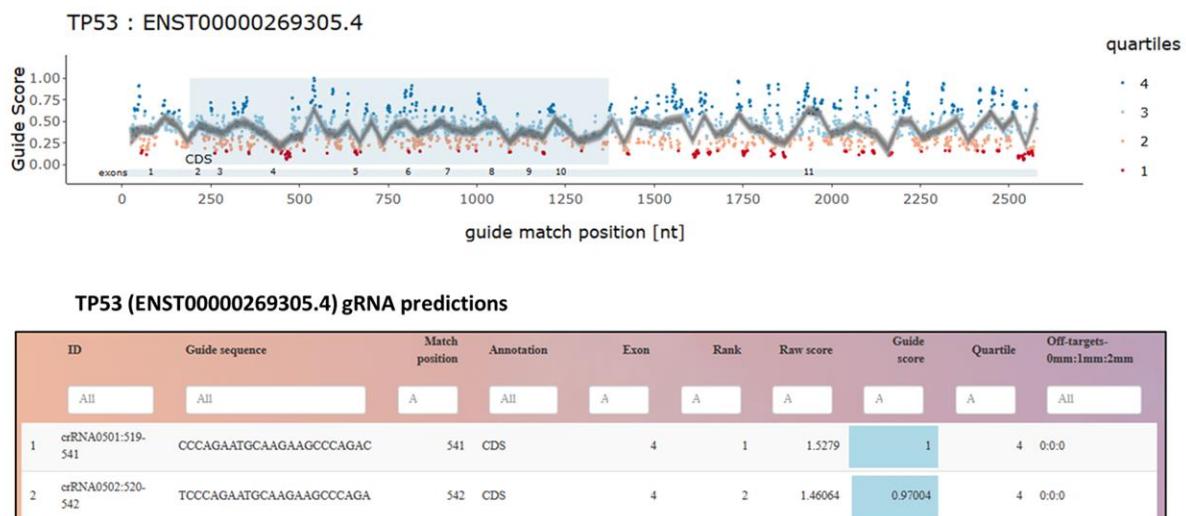
**A.** Visual summary of dCasRx-APEX2(TLCV2) plasmid ORF structure. N and C-terminal SV40 NLS flank the dCasRx-APEX2 ORF, which also contains a C-terminal HA tag. Protein structure predictions in B, C and D were based on the amino acid sequence from the initiating M to the PG of the T2A site, which is the site of separation between dCasRx-APEX2 and eGFP **B.** Visual representation of protein structure predictions from AF2 and RoseTTAFold, using iCn3D. AF2 predictions are coloured based on pLDDT scores, a measure of AF2 prediction confidence. Orange represents a low score ( $p\text{LDDT} < 50$ ). RoseTTAFold predictions are coloured from N (red) to C (violet) termini. Location of linker sequence is noted **C, D.** Highlighting (in yellow) of conserved Cas13d and peroxidase

domains of AF2 (C) and RoseTTAFold (D) structural models using iCn3d. The highlighted domain is indicated by a black box around the relevant domain above structure visualisations

## 4.5 Preliminary CasRx gRNA design algorithm testing

With creation of CasRx/dCasRx-APEX2 expression constructs appropriate for use in CWR22Rv1 cell experiments having been achieved, the next priority was design of gRNAs that would enable selective targeting of CasRx proteins to the CE3 region of *AR* mRNA. As a relatively new technology, algorithms developed for Cas13 gRNA design are few in number. This contrasts with Cas9, for which upwards of 15 published algorithms have been created (Chen and Wang, 2022). Whilst generation of plasmids was undertaken (described in Section 4.2), HEK293FT cells were used for testing of a gRNA design algorithm with CasRx expression vector pXR001.

Wessels and Méndez-Mancilla *et al* screened a lentiviral library of 7,500 gRNAs targeting GFP in HEK293 cells expressing CasRx and GFP, using FACS to sort gRNAs by knockdown efficacy (Wessels, Méndez-Mancilla *et al*, 2020). Their library contained a variety of sequence lengths, mismatches, and gRNA target locations across the GFP transcript, and in doing so the authors were able to ascertain requirements for gRNA designs using CasRx. After expanding this to additional endogenous transcripts CD44, CD55 and CD71 to increase prediction confidence, an accompanying computational design tool was released that enables users to input Ensembl transcript IDs and receive gRNA design predictions ranked on their likely efficacy (available at [cas13design.nygenome.org/](http://cas13design.nygenome.org/)) (Wessels, Méndez-Mancilla *et al*, 2020) (Figure 4.6).



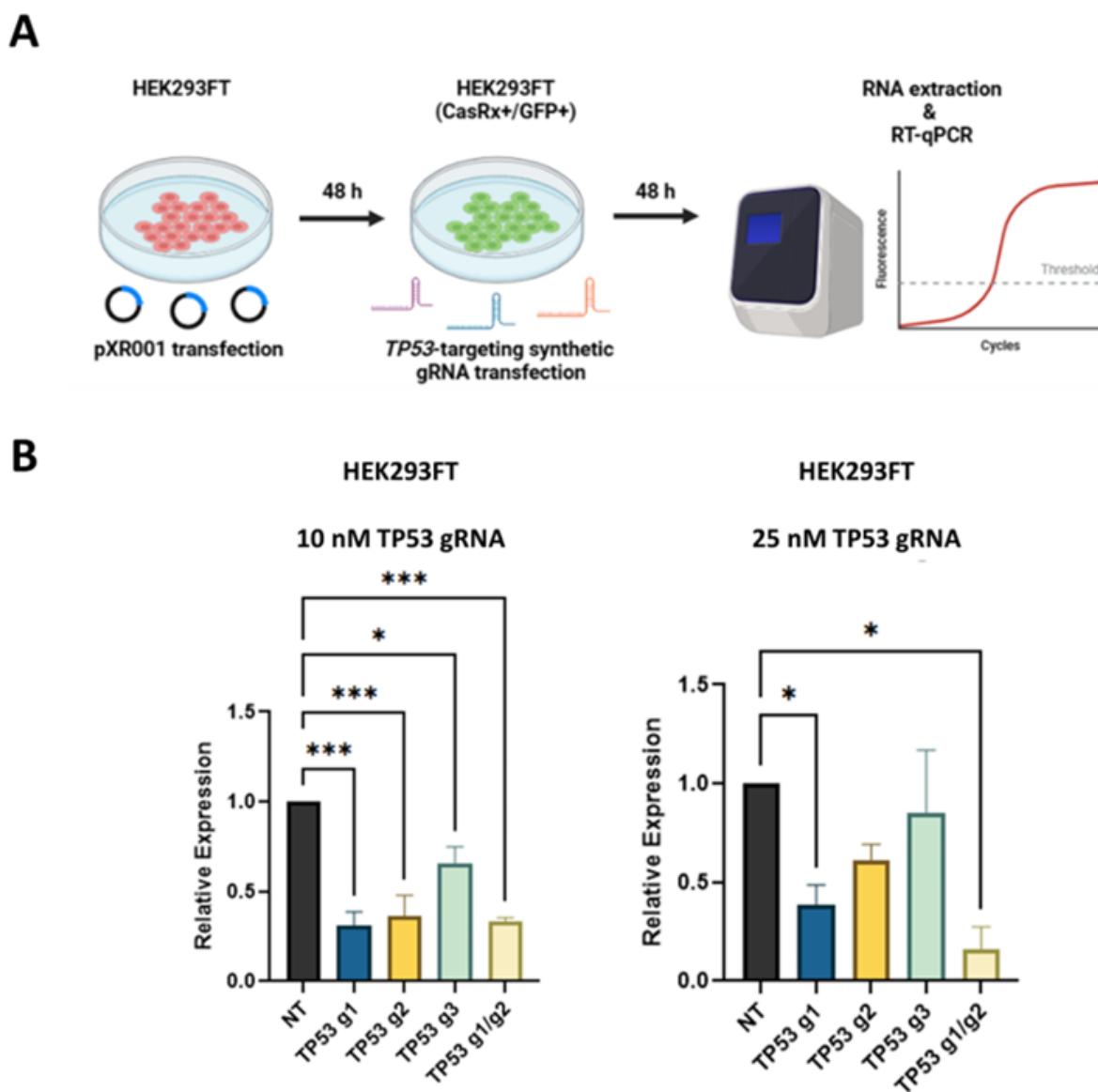
**Figure 4.6 - A published algorithm enables CasRx gRNA design for transcript targeting**

Wessels and Méndez-Mancilla *et al* undertook work to elucidate CasRx gRNA design requirements (Wessels, Méndez-Mancilla *et al*, 2020). Accompanying this publication is a computational tool, available at [cas13design.nygenome.org/](http://cas13design.nygenome.org/), enabling users to input Ensembl transcript IDs and receive CasRx gRNA sequences likely to be efficacious against a desired RNA, sorted into quartiles of predicted knockdown effect. Screenshots taken from [cas13design.nygenome.org/](http://cas13design.nygenome.org/) with targeting of *TP53* transcript ENST00000269305 as an example

For preliminary testing of this algorithm in HEK293FT, *TP53* mRNA was selected as a target due to its high abundance in HEK293FT cells and the availability of pre-validated qPCR primers sourced from a colleague at Newcastle University. Unlike the two component Cas9 gRNA which consists of a CRISPR RNA (crRNA) complementary to the target and a trans-activating CRISPR RNA (tracrRNA) which binds Cas9, CasRx gRNA consists of a single RNA oligomer (Scott *et al.*, 2019; Konermann *et al.*, 2018). This comprises an invariable 30 nt direct repeat (DR) stem loop region 5' - AACCCCUACCAACUGGUCGGGGUUUGAAC - 3' which 'loads' the gRNA onto CasRx protein, followed by a spacer exhibiting complementarity with target RNA which is most effective at lengths of 22-23 nt (Wessels, Méndez-Mancilla *et al*, 2020; Konermann *et al.*, 2018; Zhang *et al.*, 2018). *TP53* gRNAs were ordered as single synthetic 52-53 nt RNA oligos containing the 30 nt DR followed by spacer designs generated with the above algorithm.

HEK293FT were transfected in 6-well plates with 1 µg CasRx plasmid pXR001, incubated for 48 hours and transfected at final concentrations of either 10 nM or 25nM with three different synthetic gRNAs designed to target *TP53* mRNA, *TP53* g1, g2 and g3. Additionally, a 'non-targeting' (NT) gRNA spacer sequence designed by Konermann *et al.* was used as a control (Konermann *et al.*, 2018). After a further 48-hour incubation, RNA was harvested and RT-qPCR for *TP53* expression performed (Figure 4.7A). qPCR analyses of *TP53* expression demonstrated

effectiveness, to varying degrees, of all three *TP53* gRNAs. Furthermore, the higher gRNA concentration of 25 nM did not increase knockdown efficacy over 10 nM (Figure 4.7B). gRNA concentrations greater than 25 nM were not tested as ultimately the purpose here was to screen gRNA design algorithm effectiveness, and for future upscale in CWR22Rv1 proteomics experiments concentrations of, for example, 50 nM would raise costs/required gRNA quantity considerably. Additionally, this experiment provided valuable indications that use of synthetic gRNAs as single 52-53 nt RNA oligos can mediate CasRx targeting to mRNA.



**Figure 4.7 - Application of a CasRx gRNA design algorithm in HEK293FT elicits TP53 knockdown by transfection with synthetic gRNA and pXR001**

A. HEK293FT, which uptake plasmids with exceptional efficiency, were transfected with CasRx expression plasmid pXR001 and incubate for 48 hours, followed by transfection with synthetic gRNA oligos designed as described above. After a further 48-hour incubation, RNA was extracted and RT-qPCR was performed to assess

knockdown efficacy **B.** TP53-targeting gRNAs g1, g2 and g3 were transfected either alone or using a g1/g2 combination (5 nM/12.5 nM of each in the latter case). An NT gRNA negative control was also used. RT-qPCR was subsequently used to assess knockdown efficacy. qPCR data comprises  $n = 3$  independent biological replicates, plotted as mean  $\pm$  SEM. Unpaired t-test was used for determination of statistical significance. Only results significant at  $\alpha = 0.05$  or lower have significance denoted (\* =  $p < 0.05$ , \*\*\* =  $p < 0.001$ ).

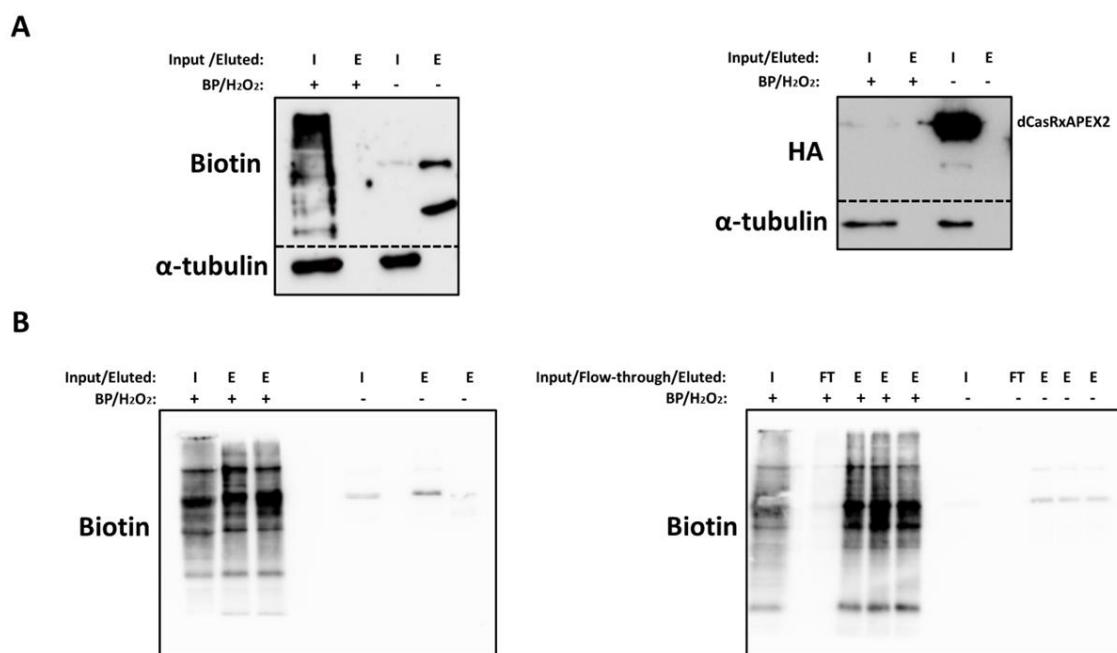
#### 4.6 Validation of dCasRx-APEX2 utility for protein proximity biotinylation

Use of synthetic gRNAs in HEK293FT to target *TP53* mRNA garnered confidence that CasRx can be targeted to specified mRNA using the aforementioned algorithm. The second novel biotechnology requiring validation was that our dCasRx-APEX2 fusion protein could enable proximity biotinylation of proteins in the presence of biotin-phenol (BP) and  $H_2O_2$ .

For initial testing of proximity biotinylation, dCasRx-APEX2 construct pXR002-APEX2, which has the same amino acid ORF as dCasRx-APEX2(TLCV2), was employed. Initial plasmid transfection in HEK293FT followed by incubation with BP and  $H_2O_2$  for 30 minutes and 1 minute, respectively, as employed previously (Hung *et al.*, 2016), demonstrated successful protein biotin-labelling as exhibited by biotin signal at a range of molecular weights. Whilst cells transfected with pXR002-APEX2 that were untreated with BP/ $H_2O_2$  failed to undergo biotinylation (Figure 4.8A, left panel). Moreover, the presence of dCasRx-APEX2 by HA western blot was confirmed, while also demonstrating the expected loss of detection after biotinylation. This is a consequence of compromised HA antibody-epitope engagement upon biotinylation due to a high percentage of tyrosine residues in the HA tag (YPYDVPDYA), which are highly reactive to biotin-phenoxyl radicals (Figure 4.8A, right panel) (Hung *et al.*, 2016). However, no biotinylated protein could be detected in fractions that were pulled down with magnetic streptavidin beads and eluted despite high levels of biotinylation in the corresponding lysates. Intriguingly, two protein bands were detectable in the eluate of enriched protein from untreated cells (Figure 4.8A, left panel).

Pilot experiments had hitherto been performed with existing lab reagents for demonstration of dCasRx-APEX2 functionality. Consequently, numerous alterations were made to our protocol due to the lack of biotinylated protein enrichment. These alterations were made to exactly match protocols of the Ting lab, the original authors of APEX2 methods (Hung *et al.*, 2016). Changes included: i) higher concentrations of biotin-phenoxyl radical quenchers sodium ascorbate, TROLOX and sodium azide in post-biotinylation wash steps; ii) the

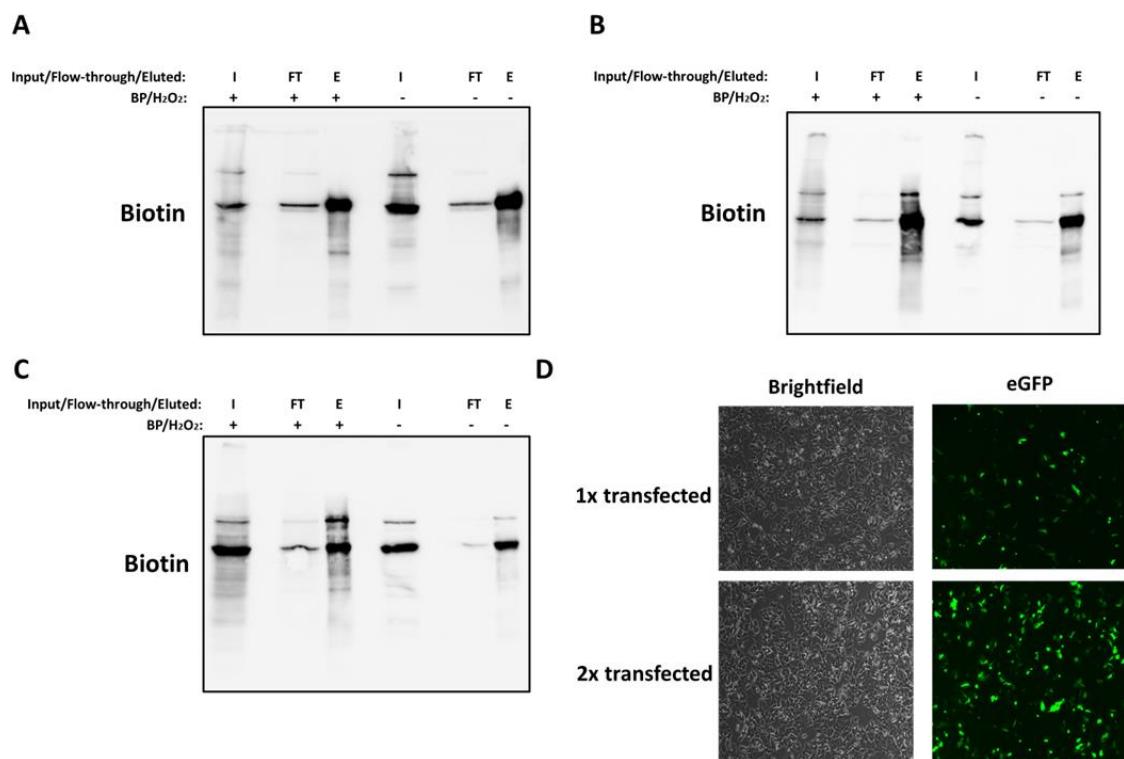
concentration of protein in lysates was ascertained in order to ensure protein lysates and streptavidin beads were being used at the correct ~1:1 µg ratio; iii) Pierce™ streptavidin beads were used as opposed to New England Biolabs, the former being used throughout the literature, and iv) streptavidin-bound samples were eluted in loading buffer supplemented with 2 mM biotin and 20mM DTT. These alterations led to significant improvements in biotinylated protein enrichment, regardless of which streptavidin washing protocols were utilised, of which 7 x 1 ml washes with ice-cold RIPA buffer was taken forward for future use (Figure 4.8B). Furthermore, the extent of biotinylated protein in the ‘flow-through’, or the lysate content that failed to bind with streptavidin, was analysed. This detected a negligible signal suggesting that the ~1:1 µg bead:protein ratio utilised is appropriate for enriching the majority of biotinylated protein (Figure 4.8B, right panel). Additionally, there appears to be presence of endogenously biotinylated proteins detected in HEK293FT samples untreated with BP/H<sub>2</sub>O<sub>2</sub> (Figure 4.8B).



**Figure 4.8 - dCasRx-APEX2 plasmid pXR002-APEX2 enables live-cell protein biotinylation and streptavidin enrichment from lysates in HEK293FT**

**A.** HEK293FT were transfected with pXR002-APEX2 and incubated for 48 hours before proximity biotinylation was induced with BP and H<sub>2</sub>O<sub>2</sub> for 30 minutes and 1 minute, respectively. Cells transfected with pXR002-APEX2 but untreated with BP/H<sub>2</sub>O<sub>2</sub> were used as a negative control. Protein lysates were extracted and incubated with magnetic streptavidin beads, before being washed and eluted. Biotin and HA signal was determined by western blot of input (I) and eluted (E) fractions **B.** Experiments were repeated and performed as in (A), with alterations made to streptavidin enrichment and elution protocols as described in text. The extent of ‘flow-through’ (FT), or protein components that did not bind to streptavidin, was also assessed. The multiple eluted fractions represent testing of different streptavidin wash buffer combinations

Next, verification was sought that proximity biotinylation was possible in the CWR22Rv1 CRPC cell line. pXR002-APEX2 was employed again whilst generation of transgenic CWR22Rv1 was undertaken (Chapter 5). Transfection with pXR002-APEX2 followed by proximity biotinylation and streptavidin enrichment as performed successfully in HEK293FT (Figure 4.8B) led to disappointing protein enrichment as evidenced by a lack of distinction between +/- BP/H<sub>2</sub>O<sub>2</sub> experimental arms (Figure 4.9A). Consequently, incubations with BP and H<sub>2</sub>O<sub>2</sub> were increased to 2 hours and 2 minutes, respectively, as has been performed by others and noted as providing greater cellular BP availability (Gupta *et al.*, 2018; Tan *et al.*, 2020). These changes improved proximity biotinylation efficacy, as evidenced by a greater distinction in eluted biotin signal between +/- BP/H<sub>2</sub>O<sub>2</sub> experimental arms (Figure 4.9B). However, labelling efficiency could not be substantially enhanced beyond this using these methods in CWR22Rv1. More plasmid was introduced through means of a double-transfection, whereby pXR002-APEX2 was transfected into cells for a second time 48 hours after the first transfection, before cells were incubated for a further 48 hours. This led to a negligible improvement in distinction between +/- BP/H<sub>2</sub>O<sub>2</sub> samples (Figure 4.9C), despite significantly higher plasmid uptake with this approach (Figure 4.9D).



**Figure 4.9 - Endogenous biotinylated proteins negatively impact APEX2 effectiveness in CWR22Rv1**

**A.** CWR22Rv1 were transfected with pXR002-APEX2 and incubated for 48 hours, before proximity biotinylation was induced and protein lysates were extracted as performed in HEK293FT previously (Figure 4.8). Lysates were enriched with magnetic streptavidin and eluted, also as performed in HEK293FT previously (Figure 4.8). Input (I), eluted (E) and flow-through (FT) fractions were analysed by biotin western blot **B.** The same experiment was performed as in (A), with an alteration that BP and H<sub>2</sub>O<sub>2</sub> incubations were increased to 2 hours and 2 minutes, respectively **C.** CWR22Rv1 were double-transfected, consisting of a second pXR002-APEX2 transfection 48 hours after the first. 48 hours after the second transfection, cells were incubated with BP/H<sub>2</sub>O<sub>2</sub>, enriched with streptavidin and eluted as in (B) **D.** Live-cell fluorescence microscopy was used to assess levels of pXR002-APEX2 plasmid uptake into CWR22Rv1 that had undergone single (1x) or double (2x) transfection

## 4.7 Discussion

The primary goal of the experiments presented in this initial results chapter was to build preliminary data and confidence in CasRx and dCasRx-APEX2 methodologies. Results and analyses here provide an invaluable proof of principle that:

- i) Expression of CasRx and dCasRx-APEX2 transgenes is detectable in CWR22Rv1 using western blotting for its HA epitope tag and through fluorescence microscopy of downstream T2A-eGFP. These features were subsequently taken forward and subcloned for both ORFs into a doxycycline-inducible expression construct that enables puromycin selection, creating novel plasmids CasRx(TLCV2) and dCasRx-APEX2(TLCV2). These will subsequently be utilised in the creation of transgenic CWR22Rv1.
- ii) The latest advances in protein structure prediction modelling software, namely AlphaFold 2 and RoseTTAFold, can be applied to simulate the structural properties of the dCasRx-APEX2 fusion protein. These provided encouraging predictions suggesting our chimeric protein is likely to fold into two separate subunits connected by a flexible linker region. It must be emphasised that these are predictions only, and dCasRx-APEX2 function will be empirically assessed in coming chapters.
- iii) Implementation of a published CasRx gRNA design algorithm enables targeted transcript degradation through transfection of synthetic gRNA oligonucleotides, as demonstrated by targeting of *TP53* mRNA in HEK293FT.
- iv) APEX2 is capable of mediating live-cell proximity biotinylation when fused with dCasRx, and biotinylated proteins are amenable to enrichment with streptavidin.

APEX2 approaches require refinement in our chosen AR-V7 expressing model of CRPC, the CWR22Rv1 cell line, due to an apparently high abundance of endogenously biotinylated proteins (Figure 4.9) which are likely to saturate streptavidin binding availability. Their presence will impose limitations on the capacity of streptavidin to bind our proteins of interest including potential splicing factors labelled by APEX2, thus in theory lowering successful protein identifications by mass spectrometry.

Contamination by endogenous biotinylated protein species has previously been noted as a concern in a variety of methods utilising streptavidin-biotin chemistries, leading to false

positives and greater background (Papageorgiou, Demmers and Strouboulis, 2013; Tytgat et al., 2015; Grant et al., 2019). For example analysis of carboxylases, known to be endogenously biotinylated, show them to reside in the cytoplasm or mitochondria in higher eukaryotes (Chapman-Smith and Cronan Jr, 1999). Therefore, alternative means of cell lysis and streptavidin enrichment will be explored. Given that splicing is a nuclear process (Han et al., 2011), and if endogenously biotinylated contaminants were restricted to extranuclear cellular compartments in CWR22Rv1, then isolation of nuclei may be an effective means of both reducing contamination and providing greater resolution in studying splicing-specific mechanisms.

It seems that other relevant studies utilising APEX2 have not encountered this difficulty, as they have routinely generated material for streptavidin enrichment by whole-cell lysis (Gao et al., 2018; Myers et al., 2018; Han et al., 2020; Lin et al., 2021; Han et al., 2017). The majority of such APEX2 work has been performed in HEK293 cells, in which we observed significantly lower biotinylated protein in BP/H<sub>2</sub>O<sub>2</sub>-untreated cells than in CWR22Rv1 (Figure 4.8). Efforts that will be made in this project to selectively isolate nuclear proteins thus represent a novel approach to subcellular proximity biotinylation that may significantly reduce unwanted contamination.

Furthermore, to enhance intracellular biotinylation and subsequent enrichment of modified proteins, nuclear extraction will be performed in a doxycycline-inducible CWR22Rv1 derivative, generation of which will be explored in Chapter 5. This would be projected to express the dCasRx-APEX2 fusion consistently across cells and between biological replicate samples. It is prudent to perform optimisation of APEX2 methods in these cells as they will subsequently be used for proteomics, and any lessons learnt from transient plasmid transfection approaches may not apply to a lentivirally transduced cell line as we hope to develop.

Having preliminarily validated these techniques, the focus of the following chapter will be development and application of these in CWR22Rv1 for the ultimate goal of identifying CE3-proximal proteins as putative AR-V7 splicing factors.

# Chapter 5 Development of transgenic CWR22Rv1 and CE3-targeted proteomics workflows

## 5.1 Introduction and rationale

Having achieved demonstrable CasRx and dCasRx-APEX2 functionality through transient transfection of HEK293FT and CWR22Rv1, focus was now shifted to optimisation of techniques in CWR22Rv1. As previously shown (Figure 4.1), transient transfection efficiency of plasmids in CWR22Rv1 is unlikely to provide the consistency or level of expression necessary for larger scale proteomics experiments, therefore use of lentivirus for creation of stable cell lines was explored. APEX2 workflows have largely adopted stable expression approaches, either with the latter method of lentiviral transduction using constitutive or doxycycline-inducible plasmids, or through knocking-in an APEX2 ORF at a gene of interest to fuse endogenous protein with APEX2 (Gao *et al.*, 2018; Myers *et al.*, 2018; Han *et al.*, 2020; Gupta *et al.*, 2018). Endogenous protein fusion-tagging presents an advantage that the level of APEX2 fusion will not be excessive which may introduce off-target proximity biotinylation, furthermore APEX2 fusion proteins will not need to compete with native species. However given that this project requires introduction of a novel, non-mammalian protein, a lentiviral expression approach is necessitated.

Paramount to project success is the selection of an appropriate CasRx gRNA to target dCasRx-APEX2 to AR CE3 in CWR22Rv1. Here, a previously established and tested CasRx gRNA design algorithm will be used along with synthetic gRNAs as done for *TP53*-targeting in HEK293FT (Figure 4.7). This is in contrast to publications using combined CRISPR-APEX2 approaches which express gRNA using U6-promoter driven constructs, either lentivirally or through transient transfection (Gao *et al.*, 2018; Myers *et al.*, 2018; Han *et al.*, 2020). We decided to utilise synthetic gRNAs, as our lab has previously demonstrated siRNA and gRNA transfection into CWR22Rv1 to occur at a far greater efficiency than seen using plasmids. Additionally this approach enables faster testing of multiple gRNA sequences, obviating the need to transduce and create a separate cell population for every gRNA of interest.

Proximity biotinylation in CWR22Rv1 will also be optimised prior to commencement of proteomics experiments. Evidence from transient transfection with pXR002-APEX2 (Figure

4.9) demonstrates that increasing incubation times with BP/H<sub>2</sub>O<sub>2</sub> to 2 hours/2 minutes, respectively, leads to improved biotin-labelling in this cell type. Moreover, a double transfection method also resulted in subtle improvements to labelling efficiency. The hypothesis here is that longer incubation times, a higher proportion of cells expressing dCasRx-APEX2 through lentiviral methods, combined with application of nuclear protein isolation, will collectively permit application of this method in CWR22Rv1.

Since the advent of APEX2 proximity biotinylation of proteins, approaches have been developed for APEX2 biotin-labelling of RNA. This can then be combined with streptavidin pulldown and RNA-Seq of enriched mRNAs, termed APEX-Seq, and has been used to interrogate spatial transcript localisation across a diverse range of subcellular compartments (Fazal *et al.*, 2019; Padrón, Iwasaki and Ingolia, 2019; Wu *et al.*, 2021). Improvements to this approach have been made through screening a range of biotin-conjugated aromatic compounds to identify biotin-aniline (BAn), a compound capable of high efficiency APEX2-catalysed RNA-labelling superior to that offered by BP (Zhou *et al.*, 2019; Huang *et al.*, 2021; Li *et al.*, 2022). The authors also utilised this method to validate APEX2 location with an APEX2-qPCR assay, a powerful method that can be leveraged for this project as a means to confirm dCasRx-APEX2 binding to its desired target transcript. This presents a crucial advantage over conventional RNA immunoprecipitation (RIP) assays to confirm target RNA binding in that it will utilise APEX2 biotinylation chemistries, which should more accurately reflect the biochemical conditions that will occur in proteomics experiments.

The primary focus of this research is identification of novel splicing factors involved in AR-V7 splicing in CRPC. Numerous publications have identified such splicing regulators (Table 1.2), primarily from candidate-based studies, which provide an informative resource to confirm the effectiveness of our dCasRx-APEX2 workflow. A wide variety of possible analysis workflows and normalisations could be applied to the data we will generate (Välikangas, Suomi and Elo, 2018; Graw *et al.*, 2020), therefore knowing the identities of confirmed CE3-interacting proteins is crucial for analysis validation.

Thus, the main aims of this chapter are:

- i) Development of transgenic CWR22Rv1 that express CasRx or dCasRx-APEX2 under the control of a doxycycline-inducible promoter

- ii) Design of an appropriate synthetic gRNA to target CasRx/dCasRx-APEX2 to AR CE3 mRNA in the CWR22Rv1 cell line
- iii) Demonstration that CWR22Rv1 cells, expressing dCasRx-APEX2, can elicit live-cell proximity biotinylation of proteins and RNA through use of aromatic biotin conjugates and H<sub>2</sub>O<sub>2</sub>
- iv) Confirmatory validation, through APEX2 proximity biotinylation of RNA, that dCasRx-APEX2 is localizing to its expected mRNA target
- v) Application of developed techniques and upscaling of material for proteomic identification of CE3-interacting proteins in CWR22Rv1.

## 5.2 Specific materials and methods

All catalogue numbers are supplied at first reference to use of reagent in the document, after which they are omitted.

### 5.2.1 Lentiviral production, cell line transduction and cell sorting

$6 \times 10^6$  HEK293FT were seeded in 100 mm dishes in 10 ml media and incubated overnight. The next day, media was replaced and cells were transfected, using *TransIT*<sup>®</sup>-LT1, with 7.5  $\mu$ g lentiviral packaging plasmid psPAX2 (Addgene #12260), 5  $\mu$ g VSV-G plasmid pMD2.G (Addgene #12259) and 2  $\mu$ g lentiviral transfer plasmid. 48 hours post-transfection, HEK293FT media was harvested and stored at 4°C, and culture media was replaced. Media was harvested again the next day, and 48/72-hour lentiviral media harvests were combined, spun at 500  $\times$  g for 5 minutes and supernatant was filtered through 0.45  $\mu$ m syringe filters. Lentivirus aliquots were subsequently frozen in cryovials at -80°C.

For lentiviral transduction of CWR22Rv1,  $1 \times 10^6$  cells were seeded in 10 cm dishes with 1 ml lentivirus to a final plate volume of 10 ml. 1 ml media was used in place of 1 ml lentivirus as a non-transduced control for puromycin selection. Polybrene (Sigma TR-1003-G) was added to transduction mixes at a final concentration of 10  $\mu$ g/ml. Cells were incubated for 48 hours before lentiviral transduction media was removed and replaced with fresh media containing 2  $\mu$ g/ml puromycin (Sigma P7255). Cells were maintained in 2  $\mu$ g/ml puromycin until 100% cell death occurred in non-transduced controls.

Successful integration and expression of CasRx(TLCV2) and dCasRx-APEX2(TLCV2) in CWR22Rv1 was screened post-puromycin selection by FACS. Cells were treated with 1  $\mu$ g/ml doxycycline hydiate (Sigma D5207) for 72 hours, with culture media and doxycycline being refreshed after 48 hours, and sorted using a FACSaria<sup>™</sup> Fusion (BD) to retain cells with  $\sim 10^2$  fold greater GFP fluorescence than uninduced cells. FACS-selected CWR22Rv1 were then taken forward for further expansion and expression was confirmed by western blot as described in Section 3.2. Expression of eGFP was also used as a surrogate for transgene expression by imaging cells with a Nikon<sup>™</sup> TE2000 fluorescence microscope.

### 5.2.2 CasRx gRNA transfection in CWR22Rv1(CasRx) and CWR22Rv1(dCasRx-APEX2)

CWR22Rv1(CasRx) and CWR22Rv1(dCasRx-APEX2) cells were reverse transfected with gRNAs at the indicated concentrations using 0.2% lipofectamine™ RNAiMAX (Thermo 13778150) according to manufacturer instructions. Opti-MEM™ I was used as serum-free media for creation of transfection complexes. At the point of transfection, expression of CasRx/dCasRx-APEX2 was induced using 1 µg/ml doxycycline hyclate. Culture media and doxycycline was refreshed after 48 hours. Unless otherwise specified, 72 hours post-gRNA transfection, samples were utilised for the relevant experiments as indicated in results. All gRNA sequences used for targeting of AR mRNA in CWR22Rv1(CasRx) and CWR22Rv1(dCasRx-APEX2) are listed in Table 5.1.

CasRx gRNA	Sequence (5' - 3')
non-targeting (NT)	AACCCCUACCAACUGGUUCGGGGUUUGAAACUCACCAGAACGUACCAUACUC
AR g1	AACCCCUACCAACUGGUUCGGGGUUUGAAACGUAAAAGUGAACUGAUGCAGCU
AR g2	AACCCCUACCAACUGGUUCGGGGUUUGAAACAGAUGCUUGCAAUUGCCAACCCG
AR g3	AACCCCUACCAACUGGUUCGGGGUUUGAAACAAAGUAAUAGUCAAUGGGCAAAA

Table 5.1 - CasRx gRNA sequences used for AR mRNA targeting in CWR22Rv1(CasRx) and CWR22Rv1(dCasRx-APEX2)

### 5.2.3 dCasRx-APEX2 protein proximity biotinylation using CWR22Rv1(dCasRx-APEX2)

CWR22Rv1(dCasRx-APEX2) were induced with doxycycline for 72 hours as described in Section 5.2.1, before APEX2 proximity biotinylation was performed by incubation with 500 µM biotin-phenol and 1mM H<sub>2</sub>O<sub>2</sub> for 2 hours and 2 minutes, respectively, as described in Section 4.2.4. Proximity biotinylation was terminated using APEX2 quenching buffer washes as detailed in Section 4.2.4, before nuclear-cytoplasmic fractionation was performed.

### 5.2.4 Biotinylated protein pulldown with streptavidin (nuclear-cytoplasmic fractionation)

For nuclear-cytoplasmic fractionation, after 4 x APEX2 quenching buffer washes were performed, an additional 4 x PBS washes were carried out. Cells were then trypsinised, pelleted and the NE-PER™ nuclear and cytoplasmic extraction kit (Thermo 78833) was used according to manufacturer instructions with the following changes: i) protease inhibitor cocktail tablets (as used for RIPA whole-cell lysis) were made up to a 100x stock concentrate slurry in 100 µl PBS, which was added 1:100 to CER I and NER; ii) vortex and incubation times for all lysis steps performed using CER I and CER II were doubled; iii) after centrifugation of

nuclei, the resulting pellet was washed once in each of 1 ml ice-cold PBS and 200 µl CER I, ensuring complete removal of all supernatant before addition of NER; iv) during nuclear lysis with NER, vortex times were doubled and at the 0, 20 and 40-minute lysis timepoints, nuclei were sonicated 2x 30 seconds on/off on a low setting using a Bioruptor® (Diagenode) sonicator.

All subsequent steps involving determination of protein concentration, streptavidin pulldown, washing of beads and SDS-PAGE were performed as described in Section 4.2.5, with the exceptions that NE-PER™ NER or CER I lysis reagents were used as blanks in place of RIPA for assaying of protein concentration, and NER was added to overnight lysate/streptavidin mixtures in place of RIPA.

### **5.2.5 Immunofluorescence**

Cells were grown on 22mm x 22mm glass coverslips (VWR 631-0125) sterilised with 70% ethanol. Upon the indicated timepoints/treatments, media was removed and cells were washed with PBS, before being fixed in 4% paraformaldehyde (Fisher 11400580) in PBS for 20 minutes at room temperature. Cells were washed 2 x 5 minutes in PBS, and permeabilised with 0.1% triton X-100 (Sigma X100) in PBS for 10 minutes at room temperature. Cells were washed again for 2 x 5 minutes in PBS before being blocked in 4% BSA (Sigma 810033) in PBS for 30 minutes at room temperature.

The relevant primary antibody was diluted 1:50 in 4% BSA/PBS, and blocked cells on coverslips were stained at 4°C overnight. Cells were then washed for 3 x 5 minutes in PBS and incubated for 1 hour at room temperature using donkey anti-mouse Alexa Fluor™ Plus 647 secondary antibody (Thermo A32787) diluted 1:200 in 4% BSA/PBS. Cells then underwent another 3 x 5 minutes wash in PBS, before being mounted on glass slides (VWR 631-0108) using mounting medium with DAPI (Abcam ab104139). Samples were imaged on a Leica™ DM6 widefield microscope (Leica Biosystems).

Colocalisation analysis of immunofluorescence images was performed using Fiji (v2.1.0) plugins JACoP (v2.1.1) and Colocalization Finder (v1.6) (Schindelin et al., 2012; Bolte and Cordelières, 2006; Carl et al., 2004), using default settings.

### 5.2.6 Multiple structure alignment analysis of AlphaFold 2 structural predictions

PDB structural predictions were generated from dCasRx and dCasRx-APEX2 amino acid FASTA sequences using AlphaFold 2 run via CATANA server as described in Section 4.4. Multiple structure alignment analysis comparing dCasRx and dCasRx-APEX2 structures was carried out using TM-score available at <https://zhanggroup.org/TMscore/> (Zhang and Skolnick, 2004) and iCn3D, available at <https://www.ncbi.nlm.nih.gov/Structure/icn3d/> (Wang et al., 2022). Both software were run using default settings.

### 5.2.7 Protein immunoprecipitation

CWR22Rv1(dCasRx-APEX2) induced with doxycycline or HEK293FT transfected with pXR002-APEX2 were crosslinked in 150 mm dishes using 0.2% formaldehyde (Sigma F1635) in PBS for 15 minutes at room temperature, before formaldehyde was removed and crosslinking reactions quenched in 125 mM glycine in PBS for 10 minutes at room temperature. Cells were scraped into 5 ml ice-cold PBS and centrifuged at 1,000 x g for 10 minutes at 4°C. A further 5 ml of ice-cold PBS was used to scrape and wash out remaining cells from dishes, which was combined with cell pellets and spun again at 1,000 x g for 10 minutes at 4°C. All supernatant was removed and cell pellets were snap-frozen in LN<sub>2</sub>.

1.2 mg magnetic protein A Dynabeads™ (Thermo Fisher 10746713) were washed in 400 µl 0.5% BSA in PBS. Beads were magnetised and supernatant was replaced with fresh 700 µl 0.5% BSA in PBS, plus 1 – 5 µg of relevant antibody for protein(s) of interest. An appropriate mouse (Thermo 15297367) or rabbit (Diagenode C15410206) IgG isotype control was also included. Bead/antibody mixtures were rotated overnight at 4°C to allow antibody to couple with beads. The next day, snap frozen cell pellets were lysed by resuspension in 250 – 500 µl RIPA buffer supplemented with protease inhibitor cocktail, 1mM DTT and 400 µM ribonucleoside vanadyl complex (VRC) (NEB S1402) and rocking on ice for 30 minutes, before lysates were spun at 15,000 x g for 15 minutes at 4°C. The soluble protein supernatant was then transferred to a fresh tube and protein concentration was determined by Pierce™ 660nm protein assay as in Section 4.2.5. 50 µg of sample was taken as an input sample and boiled for 10 minutes with 6x SDS loading at 10 minutes at 100 °C.

Antibody-conjugated dynabeads were washed with 4 x 1 ml ice-cold NT<sub>2</sub> buffer (50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1 mM MgCl<sub>2</sub> (Sigma M8266), 0.05% IGEPAL (Sigma I8996)), followed by resuspension in 800 µl NT<sub>2</sub> supplemented with 1mM DTT, 400 µM VRC and 15mM

EDTA (Sigma E6758). 500 µg – 1 mg of protein lysate was added to the antibody-conjugated beads/NT<sub>2</sub> mixture, which was rotated overnight at 4°C. The next day, beads were washed with 7 x 1 ml ice-cold NT<sub>2</sub> and boiled in 50 µl SDS sample buffer mixed 9:1 with β-mercaptoethanol for 15 minutes at 100°C. Input and immunoprecipitated samples were subsequently analysed by SDS-PAGE for protein(s) of interest as described in Section 3.2.

### **5.2.8 RNA immunoprecipitation**

For RNA immunoprecipitation, RIPA and NT<sub>2</sub> buffers were made up in nuclease-free water. HEK293FT transfected with pXR002-APEX2, or 22Rv1(dCasRx-APEX2) induced with doxycycline, were transfected with the relevant gRNAs as described in results, before being crosslinked with formaldehyde, harvested, snap-frozen, and dynabeads blocked/coupled with antibody as described in Section 5.2.7. Snap-frozen cells were lysed using RIPA buffer supplemented with supplementary reagents as detailed in Section 5.2.7, with the addition of 100 U/ml RNaseOUT™ (Thermo 10777019). A 5 - 10% volume of cleared RIPA lysate was taken and stored at -80°C as an input sample. Antibody-conjugated dynabeads were washed in 4 x 1 ml ice-cold NT<sub>2</sub>, and resuspended along with remaining RIPA lysates and NT<sub>2</sub> to a final bead/RIPA/NT<sub>2</sub> mixture volume of 1 ml, which was supplemented with final concentrations of 1mM DTT, 400 µM VRC, 15mM EDTA and 200 U/ml RNaseOUT™ and rotated overnight at 4°C.

The next day, beads were washed with 7 x 1 ml NT<sub>2</sub>, and RNA was eluted/proteins were digested by resuspending in 100 µl NT<sub>2</sub>, 4.8 µl 5 M NaCl and 200 µg proteinase K (Qiagen 1019497), before being incubated at 42°C for 1 hour followed by 55°C for 1 hour with regular vortexing. Simultaneously, input lysates were mixed with 100 µl NT<sub>2</sub>, 4.8 µl 5 M NaCl and 40 µg proteinase K, before being incubated at 55°C for 1 hour followed by 65°C for 1 hour with regular vortexing. 1 ml TRIzol™ was then added to beads/input samples, RNA was extracted and RT-qPCR analysis was performed as detailed in Section 3.3. Enrichment of target RNA(s) between mRNA-targeting vs. NT gRNA samples was compared by normalising samples to their respective input and calculating relative fold enrichment. *RPL13A* was used as a housekeeping gene. Details of RT-qPCR enrichment analysis are explained in more depth in Section 5.2.9.

### **5.2.9 dCasRx-APEX2 RNA biotinylation and streptavidin-pulldown assay**

5 x 10<sup>6</sup> CWR22Rv1(dCasRx-APEX2) cells were reverse transfected in 150 mm dishes with 25 nM of either NT or AR g2 gRNA using 0.2% RNAiMAX as described in Section 5.2.2, and

expression of dCasRx-APEX2 was induced with doxycycline. After 72 hours, 500 mM biotin-aniline (Iris Biotech LS-3970) dissolved in DMSO was diluted to 500  $\mu$ M in culture media and vortexed until fully dissolved. gRNA-transfected cells expressing dCasRx-APEX2 were incubated in culture media + 500  $\mu$ M biotin-aniline for 2 hours at 37°C and 5% CO<sub>2</sub>, before H<sub>2</sub>O<sub>2</sub> was added to a final concentration of 1mM and plates were gently swirled by hand for 2 minutes at room temperature. Media was quickly aspirated and replaced with APEX2 quenching buffer. APEX2 quenching buffer was removed and replaced 3 times for a total of 4 x quenching washes, which was followed by 4 x PBS washes. Cells were subsequently lysed in an appropriate volume of TRIzol™ reagent.

Total RNA was extracted using TRIzol™ according to manufacturer instructions, with the exception of changes to precipitation/washing as detailed in Section 3.3. 2  $\mu$ g of RNA was taken and stored at -80°C for use as an input sample, whilst biotinylated RNA in the remaining sample was pulled down using Pierce™ streptavidin magnetic beads. Streptavidin beads and total RNA were used at a 1:2.5 ( $\mu$ l: $\mu$ g) bead:RNA ratio. RNA was stored at -80°C whilst beads were washed and blocked. An appropriate volume of beads for the aforementioned ratio was washed 2 x 1 ml in ice-cold nuclease-free NT<sub>2</sub> buffer, made up in nuclease-free H<sub>2</sub>O. Washed beads were resuspended in 200  $\mu$ l 0.1 M NaOH/0.05 M NaCl in nuclease-free H<sub>2</sub>O supplemented with 4% nuclease-free NT<sub>2</sub> and incubated for 2 minutes at room temperature, which was performed twice. Beads were then resuspended in 200  $\mu$ l 0.1 M NaCl in nuclease-free H<sub>2</sub>O supplemented with 4% nuclease-free NT<sub>2</sub> and incubated at room temperature for a further 2 minutes. Streptavidin beads were subsequently blocked by overnight rotation at 4°C in 500  $\mu$ l RNA-blocking buffer (1 mg/ml BSA, 1 mg/ml yeast tRNA (Fisher 11508736) in nuclease-free H<sub>2</sub>O, supplemented with 4% nuclease-free NT<sub>2</sub>).

After streptavidin blocking, total RNA was thawed on ice. Beads were washed in 2 x 1 ml nuclease-free NT<sub>2</sub> buffer and resuspended in 500  $\mu$ l 0.05 M NaCl in nuclease-free water, and thawed RNA was added. Bead/RNA mixtures were rotated for 90 minutes at 4°C followed by 30 minutes at room temperature, before being washed 7 x 1 ml in RIPA buffer made up in nuclease-free H<sub>2</sub>O. Washed beads were resuspended in 100  $\mu$ l digest/elute buffer (20 mM DTT, 5 mM biotin, 200  $\mu$ g/100  $\mu$ l proteinase K, 200 U/ml RNaseOUT™, 0.2 M NaCl, made up in nuclease-free water) and incubated for 1 hour at 42°C, followed by 1 hour at 55°C with regular vortexing. 1 ml TRIzol™ was then added to beads/buffer mixture and vortexed

thoroughly. RNA was isolated as previously for input samples, with the final RNA resuspension being done in 15 µl nuclease-free water due to expected low yields.

Input samples were thawed and equal amounts of input/enriched RNA were reverse transcribed and analysed by qPCR as described in Section 3.3. Target RNA enrichment between gRNA samples was calculated as below, with *RPL13A* used as a housekeeping gene. For this worked example, AR-V7 is assumed as the target RNA. First,  $\Delta Ct$  was calculated for both input and streptavidin-enriched samples:

**Input  $\Delta Ct$ :** mean  $Ct_{(AR-V7)}$  input – mean  $Ct_{(RPL13A)}$  input

**Enriched  $\Delta Ct$ :** mean  $Ct_{(AR-V7)}$  enriched – mean  $Ct_{(RPL13A)}$  enriched

The above calculations of input/enriched  $\Delta Ct$  were performed for both NT and AR g2 gRNA samples. Next,  $\Delta\Delta Ct$  was calculated for each gRNA against respective inputs:

**$\Delta\Delta Ct$  (NT):** Enriched  $\Delta Ct_{(NT)}$  - Input  $\Delta Ct_{(NT)}$

**$\Delta\Delta Ct$  (AR g2):** Enriched  $\Delta Ct_{(AR\ g2)}$  - Input  $\Delta Ct_{(AR\ g2)}$

Finally,  $\Delta\Delta Ct$  for each gRNA arm was converted to linear scale, and fold enrichment in AR g2 over NT samples was determined:

**Fold enrichment:**  $2^{-\Delta\Delta Ct_{(AR\ g2)}} / 2^{-\Delta\Delta Ct_{(NT)}}$

For confirmation of RNA biotinylation by dot blot, RNA was labelled with biotin-aniline and enriched with streptavidin as above. 500 ng RNA was then spotted onto a BrightStar™ Plus positively charged nylon membrane (Thermo AM10102) and crosslinked for 2 x 30 seconds using a SpectroLinker™ XL-1000 UV Crosslinker (Spectronics Corporation), before being air dried for 15 minutes. The membrane was then blocked in PBS + 10% SDS and 1 mM EDTA for 20 minutes at room temperature, before being incubated with 1:2,500 streptavidin HRP for 1 hour at room temperature. Membranes were then washed for 2 x 10 minutes in each of PBS + 10% SDS, PBS + 1% SDS and PBS + 0.1% SDS, before being imaged for chemiluminescence as in Section 3.2. Methylene blue stain (0.4 M acetic acid (Fisher A/0400/PB17), 0.4 M sodium acetate (Sigma S2889), 0.2% (w/v) methylene blue (Sigma M9140)) was then added to membranes for 30 minutes, and samples were washed with diH<sub>2</sub>O before imaging.

### **5.2.10 dCasRx-APEX2 proteomics sample preparation**

CWR22Rv1(dCasRx-APEX2) cells were utilised in proteomics experiments. Three experimental arms were used, consisting of 1x unlabelled control, 1x NT gRNA and 1x AR g2 gRNA transfected sample.  $5 \times 10^6$  cells in 1 or 2 150-mm dishes were reverse transfected with 25 nM of the relevant gRNA (gRNA transfection was omitted for unlabelled control samples) and induced with 1  $\mu$ g/ml doxycycline hydclate as outlined in Section 5.2.2. 72-hours post-gRNA transfection/doxycycline induction, proximity biotinylation and quenching washes were performed as described in Section 5.2.3 (unlabelled controls omitted biotin-phenol and hydrogen peroxide), followed by nuclear protein isolation using an NE-PER™ extraction kit and assaying of protein concentrations as detailed in Section 5.2.4.

Next, an input sample was taken to confirm presence or absence of biotin-labelling by SDS-PAGE and western blot, and an appropriate volume of streptavidin beads for a 1:1  $\mu$ g bead:protein ratio was washed with RIPA buffer as described in Section 5.2.4. Biotinylated nuclear lysates/washed beads were mixed and mixtures made up to a final volume of ~500  $\mu$ l by addition of NE-PER™ NER supplemented with protease inhibitor cocktail, and beads were mixed overnight at 4°C. The next day, streptavidin beads were washed by vigorous pipetting using 7 x 1 ml in ice-cold RIPA buffer. Beads were then left magnetised on racks and gently washed with 4 x 1 ml ice-cold PBS, followed by 3 x 1 ml ice-cold 50 mM ammonium bicarbonate (Sigma 09830), taking care not to disrupt magnetic binding of beads. All washing steps were performed using filter tips in a class II laminar flow hood to minimise keratin contamination. After washing, streptavidin beads were resuspended in 100  $\mu$ l 50 mM ammonium bicarbonate, transferred to a fresh eppendorf tube, pelleted by centrifugation and stored at -80°C. Beads were then sent on dry ice by courier to Glasgow Polyomics for sample preparation and mass spectrometry.

### **5.2.11 Label-free mass spectrometry sample preparation and submission**

Streptavidin-bound protein samples in 100  $\mu$ l 50 mM ammonium bicarbonate received by Glasgow Polyomics were prepared for LC-MS/MS by tryptic digest and drying via a filter aided sample preparation (FASP) protocol. Briefly, samples were denatured in 4% (w/v) SDS, 100 mM Tris/HCl (pH 7.6), 0.1 M DTT and mixed with UA (8 M urea (Sigma, U5128) in 0.1 M Tris/HCl (pH 8.5)). Samples were then alkylated using 0.05 M iodoacetamide in UA. Streptavidin-bound proteins were digested into peptides using trypsin at 37°C overnight,

which were then mixed with 10% acetonitrile, acidified in  $\text{CF}_3\text{COOH}$  and dried down. Dried peptides were solubilised in 20  $\mu\text{l}$  5% acetonitrile with 0.5% formic acid, before being sampled using the auto-sampler of a RSLCnano nanoflow uHPLC system (Thermo), followed by electrospray ionisation and tandem mass spectrometry (MS/MS). This was initially performed without quantification of peptide injection masses for initial  $n = 4$  experiments. A 5  $\mu\text{l}$  volume of solubilised peptide was desalted and concentrated on a trap column in 1% acetonitrile and 0.1% formic acid, and peptide separation was performed on a PepMap™ C18 reversed phase column using a formic acid/acetonitrile solvent gradient. LC elutes underwent electrospray ionisation using Sharp Singularity emitters (Fossil Ion Tech), followed by peptide ion detection on an Orbitrap™ Elite mass spectrometer (Thermo).

After initial proteomics experiments, a second  $n = 4$  was performed in which digested peptide concentrations were assayed. Dried, digested peptides were solubilised in 100  $\mu\text{l}$  100 mM triethylammonium bicarbonate (TEAB) and their concentration was determined using a DeNovix® DS-11 spectrophotometer against a standard curve of HeLa cell digested peptide standards (Thermo 88328). 5  $\mu\text{g}$  of each peptide sample was then dried down, before all subsequent steps were performed as for initial  $n = 4$  experiments as described above.

### 5.2.12 Proteomics data analysis

Steps described in Section 5.2.11 were all performed by Glasgow Polyomics, who subsequently provided Thermo Orbitrap™ RAW file outputs. MaxQuant (v2.0.3.0) (Cox and Mann, 2008) was used for peptide and protein-level intensity quantification of mass spectrometry outputs, with the inbuilt Andromeda search engine (Cox *et al.*, 2011) run against a *Homo sapiens* Uniprot (Swiss-Prot and TrEMBL) FASTA proteome (release version 2021\_04, accessed 14/01/2022). MaxQuant was run using label-free intensity based absolute quantification (iBAQ) with default settings for instrument type 'Orbitrap' selected. Peptide-spectrum match and protein-level false-discovery rate thresholds were set at 1%, and the MaxQuant match between runs algorithm was enabled.

Protein-level iBAQ values were subsequently read into R (v4.1.2, R Core Team, 2021) via the RStudio development environment (v2022.07.0, RStudio Team, 2022) for data processing. Contaminant and reverse sequences were removed, and only proteins identified by  $\geq 2$  unique peptides were retained. Individual replicates were batch processed separately. First, iBAQ values in the corresponding unlabelled control sample were subtracted from iBAQ

values in NT and AR g2 samples, and any resulting negative values were set to 0. For the initial  $n = 4$  experiment, proteins that had positive iBAQ values in at least 2 out of 4 replicates for either gRNA arm were retained for analysis. Values of 0 were imputed with half the minimum iBAQ value within that replicate. iBAQ values were converted to relative iBAQ (riBAQ) by dividing each protein iBAQ value by the sum of iBAQ values within that sample. riBAQ values were then log2 transformed. For calculation of protein enrichment, a moderated t-test with a paired design was implemented using the *limma* package (v3.50.3, Ritchie *et al.*, 2015) for comparison of log2 riBAQ values between AR g2 and NT gRNA arms.

For the second, optimised  $n = 4$  proteomics experiment, data was processed as above with the following changes: i) proteins were only retained if they had a positive iBAQ value in at least 3 out of 4 replicates for either gRNA arm; ii) log2 iBAQ, rather than riBAQ, values were used as inputs for calculation of protein enrichment with *limma*.

STRING (Szklarczyk *et al.*, 2022) was used for functional analysis of proteins enriched at the specified significance thresholds. Rotation gene set testing and barcode plots were implemented using *limma*.

### 5.3 CasRx(TLCV2)/dCasRx-APEX2(TLCV2) enable creation of transgenic CWR22Rv1 derivatives for use in CasRx and APEX2 workflows

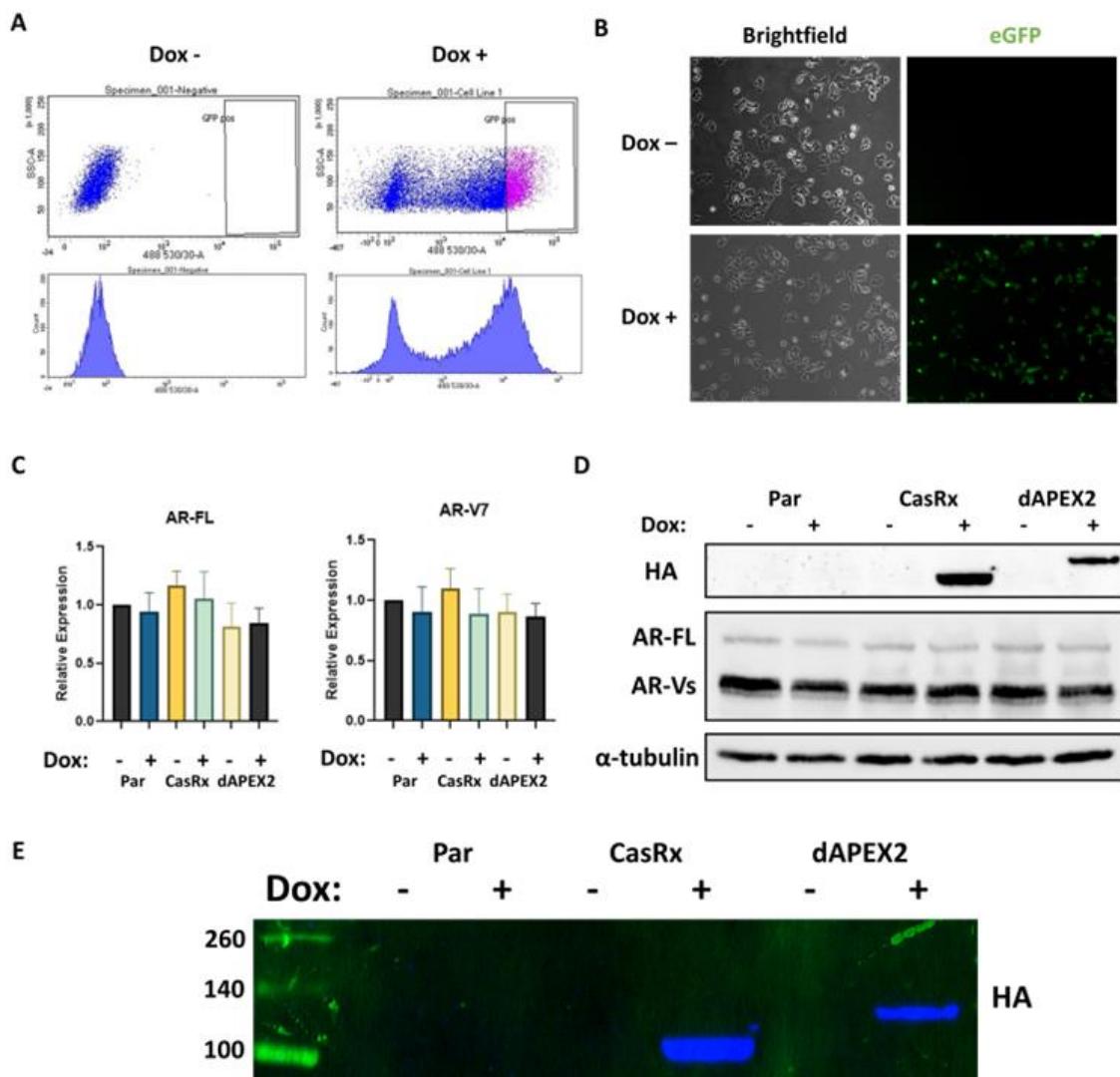
In order to apply these biotechnologies in CWR22Rv1, novel plasmids CasRx(TLCV2) and dCasRx-APEX2(TLCV2) (creation of which is discussed in Section 4.2) were utilised. CWR22Rv1 cells were transduced using lentivirus packaged with CasRx(TLCV2) or dCasRx-APEX2(TLCV2) before undergoing puromycin selection. Subsequently, live-cell FACS gated for GFP expression was used after 72 hours of doxycycline induction to select a polyclonal population of doxycycline-inducible CWR22Rv1 cells expressing our CasRx proteins based on surrogate expression of T2A-eGFP downstream of the CasRx/dCasRx-APEX2 ORFs.

The optimal cellular quantity of dCasRx-APEX2 to appropriately target *AR* CE3 mRNA is challenging to determine. Given the high level of *AR* transcription in CWR22Rv1, there are likely to be significant amounts of *AR* pre-mRNA, AR-FL and AR-V transcripts in the nucleus at any given time. However, if dCasRx-APEX2 expression is too high relative to the number of available transcripts, then greater levels of off-target biotinylation are probable due to a pool of unbound fusion protein. There is discordance in the literature with regards to preferred transgene expression levels (Myers *et al.*, 2018; Han *et al.*, 2020). Therefore, a polyclonal population of inducible cells was selected (Figure 5.1A), as we hypothesised that this mixed approach would collectively encompass suitable dCasRx-APEX2:*AR* CE3 mRNA ratios.

In addition, our goal was to express an engineered fusion protein for proteomics methods rather than elicit or study a particular phenotype, rendering monoclonal selection unnecessary and circumventing a need to repeat experiments in multiple individual clonal lines. Interestingly, we observed a significant population of cells surviving puromycin selection that failed to undergo GFP induction with doxycycline treatment (Figure 5.1A), highlighting the value of constructs containing a 2A-fluorescent protein ORF for transduction workflows. Subsequent live-cell fluorescence microscopy of the polyclonal population confirmed successful FACS gating of inducible cells, with a far higher percentage of GFP-positive cells than previously observed with transient plasmid transfections (Figure 5.1B).

For studying *AR* mRNA splicing it was imperative that levels of AR-FL and AR-V7 were retained in our transgenic cells, including under doxycycline exposure. In order to test this we compared expression of these transcripts in parental CWR22Rv1 with transgenic polyclonal

populations CWR22Rv1(CasRx) and CWR22Rv1(dCasRx-APEX2), both with and without a 72-hour doxycycline treatment. Reassuringly, there were no appreciable expression differences in AR-FL or AR-V7 mRNA between cell populations (Figure 5.1C). Additionally, we used western blot to confirm expression of HA-tagged CasRx and dCasRx-APEX2 occurred in the presence of doxycycline only, whilst also determining that the maintenance of AR-FL/AR-V7 seen by qPCR was observed at the protein level (Figure 5.1D). Furthermore, CasRx and dCasRx-APEX2 were expressed at the predicted molecular weights of ~112 and 139 kDa, respectively (Figure 5.1E). Thus, we successfully created two novel doxycycline-inducible cell populations for downstream use, CWR22Rv1(CasRx) and CWR22Rv1(dCasRx-APEX2).



**Figure 5.1 - Lentiviral transduction with CasRx(TLCV2) and dCasRx-APEX2(TLCV2) enables creation of doxycycline-inducible CWR22Rv1 derivatives**

**A.** CasRx(TLCV2) and dCasRx-APEX2(TLCV2) plasmids were packaged into lentivirus and used to transduce CWR22Rv1. Cells were then puromycin selected and induced with 1 µg/ml doxycycline for 72 hours before being

sorted by live-cell FACS. x-axis represent GFP signal intensity, events in purple were cells retained for expansion in culture **B**. Live-cell fluorescence microscopy of cells sorted in (A) **C**. Parental CWR22Rv1, CWR22Rv1(CasRx) and CWR22Rv1(dCasRx-APEX2) were treated with or without doxycycline for 72 hours, before RNA was extracted and RT-qPCR was used to measure expression of AR-FL and AR-V7. Par = parental CWR22Rv1, CasRx = CWR22Rv1(CasRx), dAPEX2 = CWR22Rv1(dCasRx-APEX2). qPCR data comprises n = 3 independent biological replicates, plotted as mean ± SEM **D**. Western blot analysis of the same treatment arms as in (C) was performed to measure protein levels of AR-FL, AR-Vs and CasRx/dCasRx-APEX2 (HA tagged).  $\alpha$ -tubulin was used as a loading control. Par = parental CWR22Rv1, CasRx = CWR22Rv1(CasRx), dAPEX2 = CWR22Rv1(dCasRx-APEX2). Western blot is representative of n = 2 biological replicates **E**. The molecular weights of HA-tagged CasRx and dCasRx-APEX2 as detected by western blot are as indicated. Numbers to left of blot represent molecular weight, in kDa, of ladder markers. Par = parental CWR22Rv1, CasRx = CWR22Rv1(CasRx), dAPEX2 = CWR22Rv1(dCasRx-APEX2)

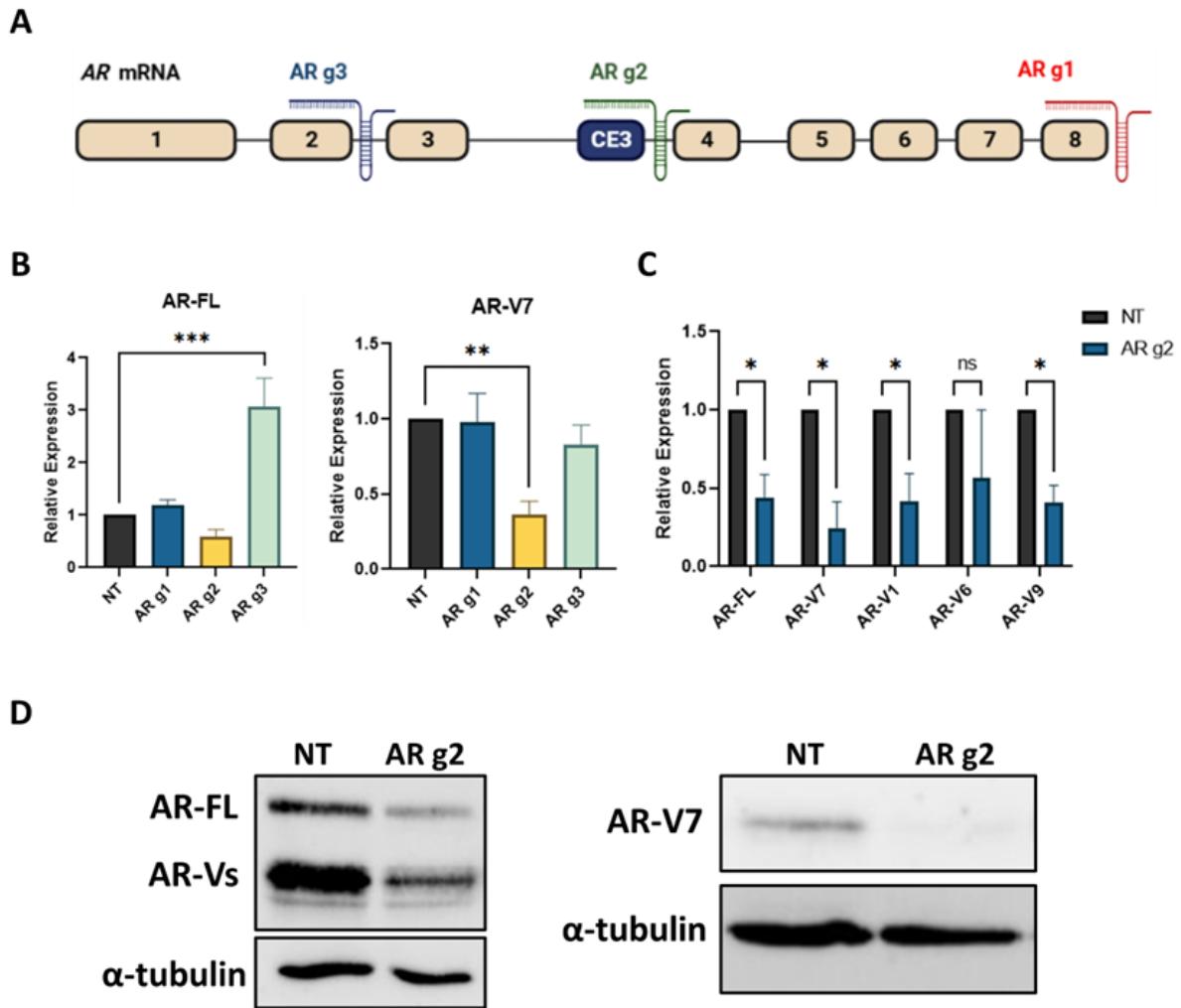
## 5.4 Synthetic gRNAs enable targeting of AR CE3 mRNA in CWR22Rv1(CasRx)

Having successfully generated doxycycline-inducible CWR22Rv1 derivatives CWR22Rv1(CasRx) and CWR22Rv1(dCasRx-APEX2), attention now turned to targeting proteins local to AR CE3 mRNA. We leveraged the same gRNA design algorithm and synthetic RNA oligo approach used previously to target TP53 mRNA in HEK293FT (Chapter 4.4) (Wessels, Méndez-Mancilla *et al.*, 2020) to design synthetic gRNAs directed at AR mRNA, resulting in three gRNAs denoted AR g1, g2 and g3. Of these, AR g2 was complementary to our region of interest for proteomics, CE3 (Figure 5.2A), whereas AR g1 and AR g3 are predicted to bind at AR exons 8 and 2, respectively. CWR22Rv1(CasRx) were subsequently utilised to screen gRNAs for selection, with the assumption made that if CasRx is capable of recognising, binding to and cleaving its mRNA target using a synthetic gRNA, then that gRNA would likely be amenable for mRNA-targeted proteomics. This presents a key advantage in that gRNAs can be efficiently screened based on knockdown efficacy as analysed by qPCR or western blot, as opposed to performing more time and resource-intensive RNA-pulldown assays with dCasRx-APEX2. For initial screening of AR g1-3, CWR22Rv1(CasRx) were transfected with gRNA and CasRx expression was induced with 1  $\mu$ g/ml doxycycline. Cells were incubated for 72 hours and RNA was harvested for RT-qPCR to quantify levels of AR-FL and AR-V7 transcript.

As expected, transfection with CE3-targeting AR g2 mediated a significant reduction in AR-V7 mRNA (Figure 5.2B). AR-FL transcripts, which do not contain CE3, also underwent a decrease with this gRNA although to a lesser extent than AR-V7. This seemingly contradictory result would be explained by CasRx activity in the nucleus due to its C and N-terminal NLS. As CE3 arises from an intronic region, CasRx is capable of targeting this CE in nascent AR transcripts to deplete overall AR pre-mRNA that could become any mature CDS, including AR-FL. Given

that splicing is coupled with transcription (Herzel *et al.*, 2017), there is only a short temporal window in which CE3 could be bound by CasRx before being spliced into AR-V7 (or out of AR-FL). Thus, if CasRx is mediating nuclear cleavage events there is an expectation that both AR-FL and AR-V7 levels would decrease, the latter to a greater extent than the former. To further probe this expectation, we assessed the effect of AR g2 on multiple AR-V species expressed in CWR22Rv1, AR-V1, AR-V6 and AR-V9. Consistent with our model, AR g2 mediated reduction of all mRNAs assayed (Figure 5.2C). We confirmed this result at the protein level by western blot, demonstrating comparable reductions to AR-FL and AR-Vs as seen by qPCR (Figure 5.2D, left panel). Moreover, an AR-V7 specific antibody demonstrated a strong reduction of AR-V7 protein using AR g2 in CWR22Rv1(CasRx) (Figure 5.2D).

Interestingly, AR-FL levels consistently underwent an approximately 3-fold increase with AR g3 (Figure 5.2B). This gRNA targets *AR* exon 2, whereas our qPCR primers for detection of AR-FL bind in exons 3 and 4. Therefore, CasRx cleavage may apparently mediate unpredictable effects on remaining *AR* exons, the mechanism of which is unclear. It was decided not to investigate this further given that effects of AR g2 appeared consistent with CE3-targeting capability, therefore this gRNA would be taken forward for future experiments.



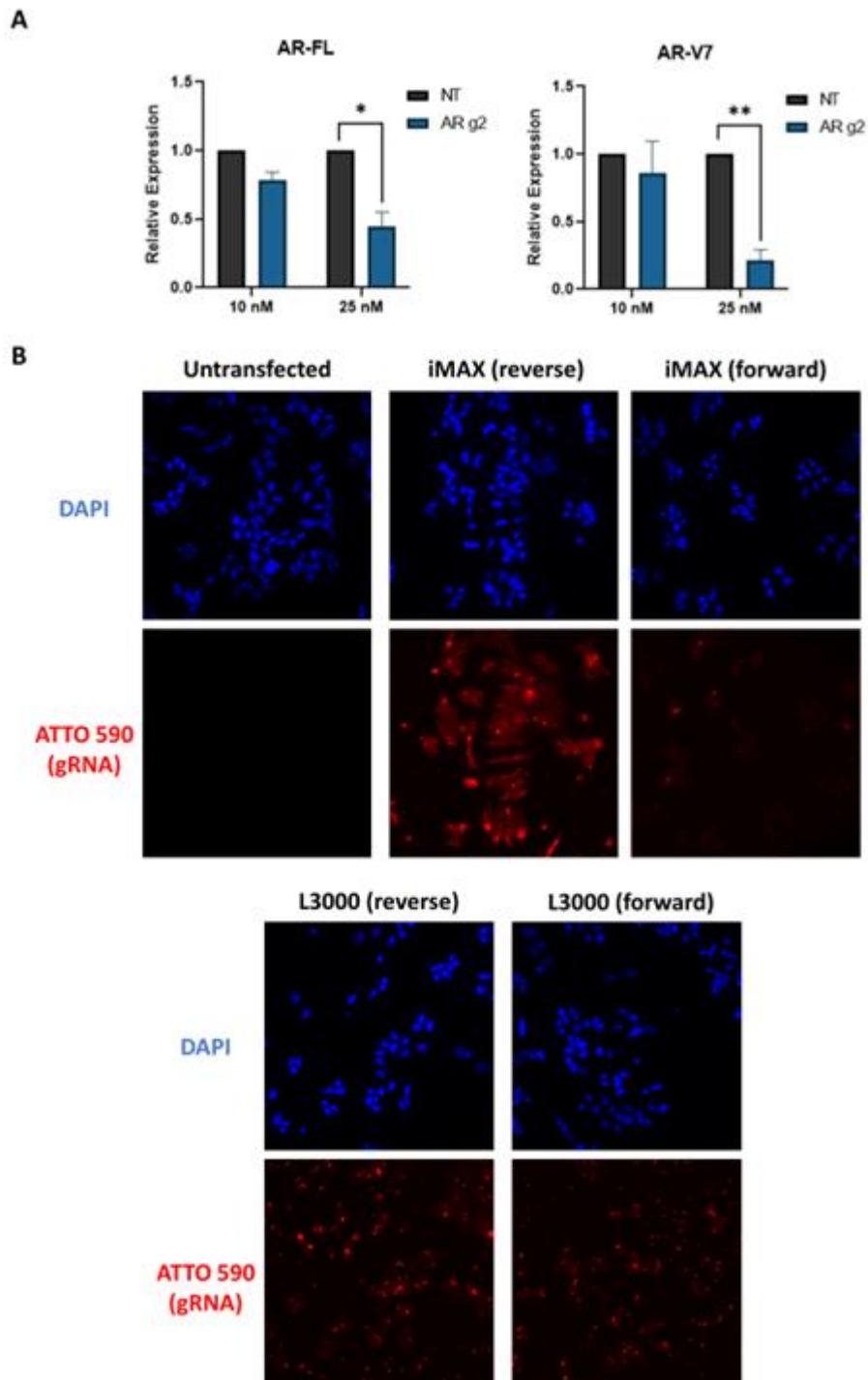
**Figure 5.2 - Use of synthetic gRNA in CWR22Rv1(CasRx) elicits AR mRNA knockdown**

**A.** A published CasRx gRNA design algorithm (Wessels, Méndez-Mancilla *et al.*, 2020), utilised previously for targeting TP53 mRNA in HEK293FT, was used to design synthetic gRNAs against AR transcripts. These gRNAs, AR g1, g2 and g3 target AR exons 8, CE3 and 2, respectively **B.** CWR22Rv1(CasRx) were transfected with the indicated gRNAs at 25 nM and induced with 1  $\mu$ g/ml doxycycline. Cells were then incubated for 72 hours before RNA was extracted for RT-qPCR and levels of AR-FL/AR-V7 mRNAs were analysed. qPCR data comprises  $n = 3$  independent biological replicates, plotted as mean  $\pm$  SEM. Unpaired t-test was used for determination of statistical significance. Only results significant at  $\alpha = 0.05$  or lower have significance denoted (\* =  $p < 0.05$ , \*\* =  $p < 0.01$ , \*\*\* =  $p < 0.001$ ) **C.** The experimental setup in (B) was repeated, with only NT and AR g2 gRNA utilised in this instance. Knockdown of the indicated AR transcripts by AR g2 was analysed. qPCR data comprises  $n = 3$  independent biological replicates, plotted as mean  $\pm$  SEM. Unpaired t-test was used for determination of statistical significance (\* =  $p < 0.05$ , ns =  $p > 0.05$ ) **D.** The experimental setup in (C) was repeated and protein levels of AR-FL, AR-Vs and AR-V7 were analysed.  $\alpha$ -tubulin was used as a loading control. Western blot is representative of  $n = 3$  independent biological replicates

Having determined AR g2 would be appropriate for proteomics workflows due to its ability to target CE3, our gRNA delivery strategy was validated. 25 nM gRNA was used throughout experiments in C 22Rv1(CasRx) (Figure 5.2). This was chosen despite prior demonstration in HEK293FT that 10 nM gRNA resulted in comparable or superior knockdown of *TP53* mRNA than 25 nM gRNA (Figure 4.7). It was reasoned that due to their exceptionally high uptake and expression of plasmid transgenes, HEK293FT would produce significantly more CasRx than CWR22Rv1(CasRx). Were this the case, there is a higher likelihood of any CasRx protein finding its cognate gRNA, even at a lower gRNA concentration. We hypothesised that gRNA quantity may prove a rate-limiting step in CWR22Rv1(CasRx), whereby a higher concentration would increase the likelihood of CasRx finding its gRNA. In order to test this, AR g2 knockdown efficacy at 10 nM and 25 nM concentration was compared in CWR22Rv1(CasRx). This comparison revealed superior knockdown of both AR-FL and AR-V7 with 25 nM AR g2 (Figure 5.3A). Due to a 25 nM concentration proving efficacious and comparable to CasRx knockdown percentages seen in the literature, greater concentrations were not pursued. Significant upscale of material would be required for subsequent proteomics experiments, thus it was reasoned that increasing gRNA amounts further would reach the point of diminishing returns for considerably greater costs.

Having demonstrated 25 nM gRNA to be appropriate for targeting of *AR* CE3 mRNA, the efficiency of our chosen gRNA delivery reagent, RNAiMAX, was determined. To this end a fluorescent NT CasRx gRNA was designed, in which a molecule of ATTO 590 dye was appended to the 3' end to enable fluorescence imaging assessment of gRNA uptake. CWR22Rv1 were transfected with gRNA using lipofectamine™ RNAiMAX or lipofectamine™ 3000 (L3000), both by reverse and forward transfection. Cells were incubated for 72 hours post-transfection before being fixed, permeabilised and imaged for ATTO 590 and DAPI. Cells were also treated with doxycycline as previously (Figure 5.2) in order to assess gRNA stability and retention when present with CasRx at the 72-hour timepoint, as it has been noted in CRISPR-Cas9 systems that synthetic gRNA may be afforded protection from exonuclease degradation when complexed with Cas9 (Allen, Rosenberg and Hendel, 2021). Appraisal of gRNA uptake by fluorescence microscopy revealed greater gRNA transfection efficiency using RNAiMAX, and that a reverse transfection protocol as has been performed in knockdown experiments is preferable to forward transfection (Figure 5.3B). Although cells transfected with L3000 did

exhibit efficient gRNA uptake, this was accompanied by extracellular fluorescent foci that may represent a degree of cellular debris caused by toxicity (Figure 5.3B).



**Figure 5.3 - Transfection of 25 nM CasRx gRNA using RNAiMAX mediates efficient cellular uptake in CWR22Rv1(CasRx)**

A. 22Rv1(CasRx) were transfected with 10 nM or 25 nM AR g2 and induced with 1  $\mu$ g/ml doxycycline. After 72 hours, RNA was harvested and analysed by RT-qPCR for levels of AR-FL and AR-V7 transcripts. Data comprises n

= 3 independent biological replicates, plotted as mean  $\pm$  SEM. Unpaired t-test was used for determination of statistical significance. Only results significant at  $\alpha$  0.05 or lower have significance denoted (\* =  $p < 0.05$ , \*\* =  $p < 0.01$ ) **B.** NT CasRx gRNA, conjugated with a 3' ATTO 590 fluorophore, was transfected into CWR22Rv1(CasRx) at 25 nM using a range of methods, and cells were induced with doxycycline as in (A). At the 72-hour timepoint, cells were fixed, permeabilised and imaged by fluorescence microscopy at 40x magnification (RNAiMAX = lipofectamine™ RNAiMAX, L3000 = lipofectamine™ 3000)

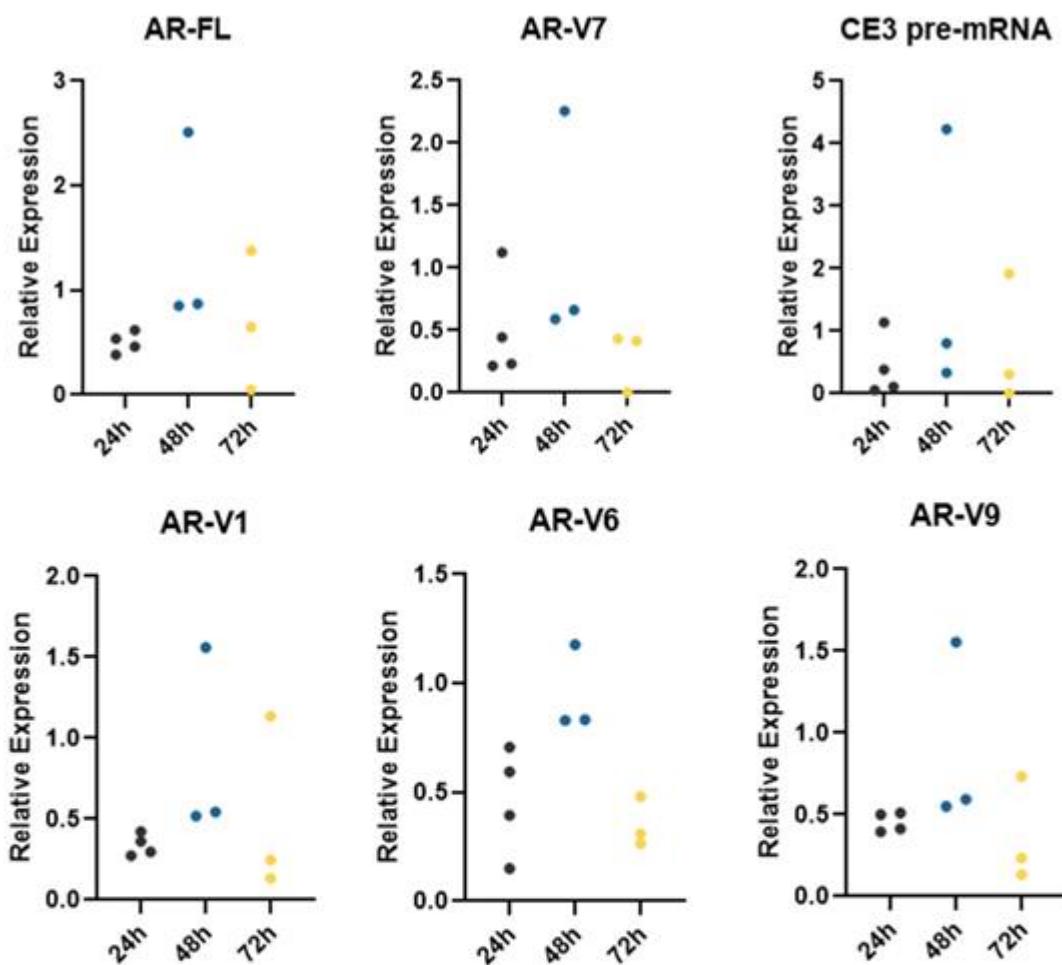
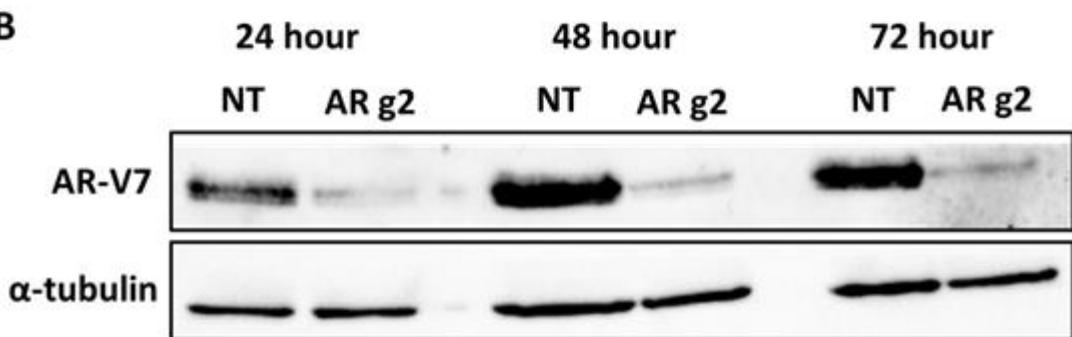
Although fluorescence imaging results suggest retention of gRNAs is maintained at the 72-hour timepoint, it is feasible that the oligonucleotide itself may be degraded by exonucleases whilst observed signal is attributable to remaining ATTO 590 fluorophores. Similarly, studies have aimed to enhance intracellular stability of synthetic CasRx gRNAs, resulting in use of chemical modifications such as 3x uridine residues appended to the 3' end of gRNAs by a phosphorothioate bond, or a 3' inverted thymidine. These modifications were demonstrated to significantly improve target knockdown duration beyond the 24-hour timepoint in HEK293FT cells expressing CasRx, purportedly due to added protection from 3' - 5' RNA exonucleases (Méndez-Mancilla *et al.*, 2022).

In order to test knockdown persistence with synthetic AR g2 in CWR22Rv1(CasRx), a time-course experiment was established in which cells were transfected with 25 nM NT or AR g2 gRNA and CasRx expression was induced with doxycycline as previously, before RNA was harvested at 24, 48 and 72-hour timepoints for subsequent qPCR knockdown analysis. Analysis across a range of primer sets demonstrated no clear temporal trend in knockdown of AR transcripts, including qPCR primers designed for CE3 found in AR pre-mRNA (Figure 5.4A). Knockdown efficiencies were variable (therefore individual datapoints were plotted), and although the 24-hour timepoint led to somewhat more consistent knockdown across replicates, the overall knockdown percentages displayed no clear inverse relationship between time post-transfection and knockdown efficiency, as observed by others (Figure 5.4A) (Méndez-Mancilla *et al.*, 2022). Intriguingly, samples harvested 48-hours post gRNA transfection/doxycycline induction generally exhibited poorer transcript reduction than 24 or 72-hour timepoints (Figure 5.4A). Nevertheless, should previously published results demonstrating synthetic gRNA degradation over time occur in our system, the expectation would be of a linear reduction in knockdown percentage from 24-72 hours, which we did not observe.

Work on CasRx synthetic gRNA stability performed by others used fluorescent protein intensity as a readout, demonstrating a near 4-fold reduction in protein knockdown between 24 and 72-hour timepoints for unmodified synthetic gRNAs that lack 3' modifications, reducing knockdown to only ~20% vs NT controls (Méndez-Mancilla *et al.*, 2022). Thus, our results were confirmed at the protein level by western blot. The same time-course, here measuring AR-V7 protein levels, revealed no apparent differences in protein reduction between 24 and 72-hour timepoints (Figure 5.4B). This direct comparison over time confirmed our previous screening of synthetic gRNAs which were all harvested at 72 hours (Figure 5.2). Given these results and that CWR22Rv1(CasRx)/CWR22Rv1(dCasRx-APEX2) were sorted for GFP positivity after a 72-hour doxycycline induction, we proceeded with this timeframe in subsequent work. It is intriguing that CasRx knockdown efficacy of AR-V7 measured by qPCR (Figure 5.4A) does not directly reflect protein level readouts (Figure 5.4B). Variability in qPCR results was apparent, and the latter readout shows a comparatively greater level of reduction. CasRx as a knockdown tool is still a relatively novel technique, and it is conceivable that CasRx cleavage may result in irregular mRNA fragments that are detectable by qPCR but unable to undergo translation into protein, possibly explaining this inconsistency.

Differences in gRNA delivery method may explain discrepancies between our results and those of others. Our choice, RNAiMAX, is a cationic lipid-based formulation, which has been noted to shield synthetic RNA oligonucleotides from exonuclease degradation for several hours (Judge *et al.*, 2006; Zuris *et al.*, 2015). This contrasts with delivery by electroporation, which offers no such protection (Kelley *et al.*, 2016). Given that nucleofection, an electroporation based method, was used in the sole publication to date that has studied synthetic CasRx gRNA stability (Méndez-Mancilla *et al.*, 2022), our results suggest that cationic lipids are preferred for these workflows. It is likely that this approach sufficiently protects gRNAs until CasRx is expressed, at which point complexing with the latter further enhances guide preservation.

Collectively, this has demonstrated that an appropriately designed synthetic gRNA can be used to selectively degrade *AR* transcripts in CWR22Rv1(CasRx). We have shown that optimisation of delivery method, gRNA concentration and timeframe elicits effective targeting of *AR* mRNA, which will be invaluable for targeting dCasRx-APEX2 to CE3 in subsequent proteomics experiments.

**A****B**

**Figure 5.4 - Knockdown using synthetic gRNAs in CWR22Rv1(CasRx) is maintained over 72 hours**

**A.** CWR22Rv1(CasRx) were transfected with 25 nM NT or AR g2 and CasRx expression was induced with 1  $\mu$ g/ml doxycycline as previously. RNA was harvested at 24, 48 and 72-hour timepoints for RT-qPCR knockdown analysis of the indicated transcripts. qPCR data is normalised to a corresponding NT sample at each timepoint, plotted as individual replicate knockdown vs NT control. Data comprises  $n = 3$  independent biological replicates ( $n = 4$  for 24-hour timepoint due to RNA extraction issues with 48/72-hour). **B.** The experimental setup in (A) was repeated and protein was harvested for western blot analysis of AR-V7 levels.  $\alpha$ -tubulin was used as a loading control. Western blot experiment included a corresponding NT gRNA sample at each timepoint. Data comprises  $n = 1$

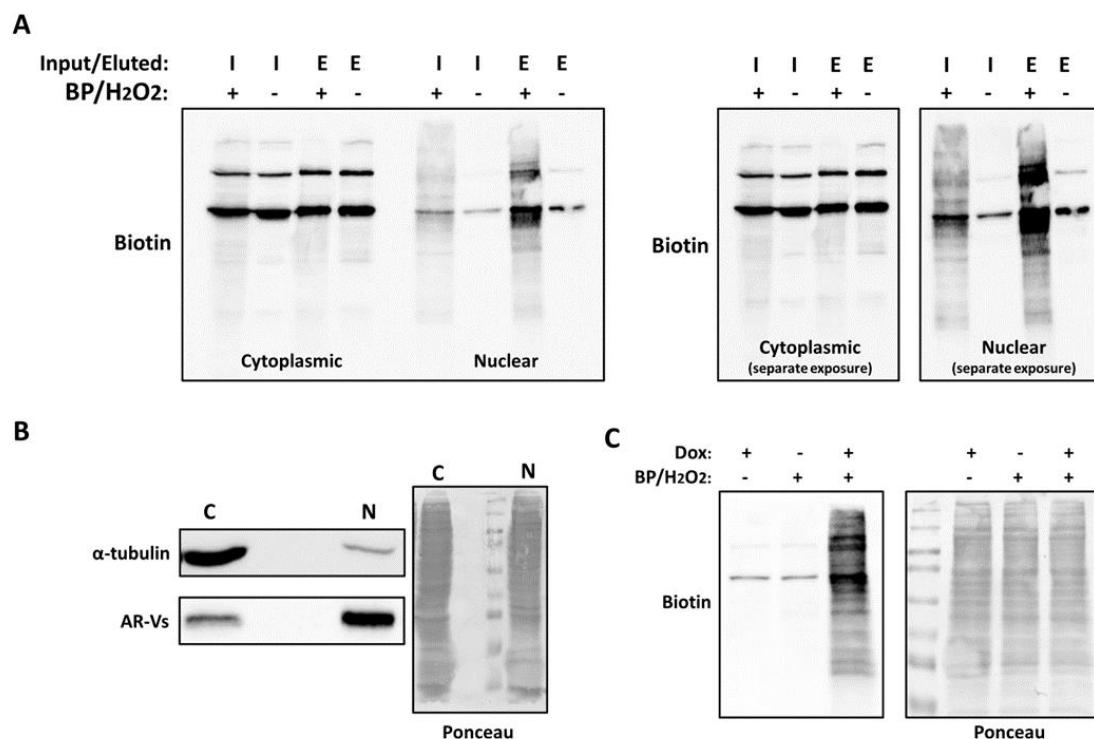
## 5.5 CWR22Rv1(dCasRx-APEX2) enable live-cell proximity biotinylation of proteins

Now that selection and delivery of a synthetic gRNA was validated for CE3-targeted proteomics by using CWR22Rv1(CasRx-TLV2), focus turned to CWR22Rv1(dCasRx-APEX2). Previous results using transient plasmid transfection revealed that although CWR22Rv1 appear capable of live-cell proximity biotinylation by dCasRx-APEX2, alterations to cell lysis and streptavidin enrichment were necessary, ostensibly due to the presence of endogenously biotinylated proteins (Figure 4.9). Confidence that our dCasRx-APEX2 fusion protein was functioning effectively was acquired by experiments in HEK293FT (Figure 4.8), and we now possessed an effective means to express it in CWR22Rv1(dCasRx-APEX2) (Figure 5.1).

Due to the previously speculated presence of endogenously biotinylated proteins in CWR22Rv1, and that these were likely to be cytoplasmic or mitochondrial (Section 4.7), a nuclear isolation protocol was explored. This may have a dual advantage of both depleting these contaminants and selectively enriching for proteins in the nucleus, where alternative splicing takes place. CWR22Rv1(dCasRx-APEX2) were induced with 1 $\mu$ g/ml doxycycline for 72 hours, shown to enable effective dCasRx-APEX2 expression (Figure 5.1), before undergoing proximity biotinylation by a 2-hour and 2-minute incubation with BP and H<sub>2</sub>O<sub>2</sub>, respectively. These incubation times were demonstrated to improve biotinylation in CWR22Rv1 during preliminary experiments with transient pXR002-APEX2 plasmid transfection (Figure 4.9). Post-labelling, cells were washed with APEX2 quenching buffers as performed previously, washed with PBS to remove quenchers and a nuclear-cytoplasmic fractionation was performed. Western blot analysis revealed significant improvements in distinction between +/- BP/H<sub>2</sub>O<sub>2</sub> experimental arms, as well as a clear ability to enrich biotinylated proteins with streptavidin (Figure 5.5A). Crucially, these improvements were only seen in nuclear protein samples thus vindicating earlier hypotheses regarding the source of contamination. No detectable difference in biotin signal was observed between +/- BP/H<sub>2</sub>O<sub>2</sub> samples for cytoplasmic fractions (Figure 5.5A).

Next, subcellular protein markers were used to assess bleed-through between compartments. Cytoskeletal component  $\alpha$ -tubulin was used as a cytoplasmic protein marker, whilst we exploited constitutive nuclear localisation of AR-Vs as a surrogate for nuclear content (Hohmann and Dehghani, 2019; Kim *et al.*, 2022). Western blotting demonstrated

greatly elevated  $\alpha$ -tubulin protein content in the cytoplasmic compared to nuclear fraction whilst the converse was true of AR-Vs, despite equal total protein as verified by ponceau stain (Figure 5.5B). These results are consistent with an effective separation of cytoplasmic and nuclear proteins. Our fractionation method does not enable the total absence of cytoplasmic proteins, as evidenced by a small amount of  $\alpha$ -tubulin signal in nuclear fractions, which may also explain the persistence of biotin signal in unlabelled nuclei (Figure 5.5A). Finally, we sought to demonstrate that both dCasRx-APEX2 induction with doxycycline and incubation with BP/H<sub>2</sub>O<sub>2</sub> were required to enact proximity biotinylation in CWR22Rv1(dCasRx-APEX2). Cells were either induced and not treated with BP/H<sub>2</sub>O<sub>2</sub>, treated and not induced, or both induced and treated. Subsequent western blot analysis of nuclear proteins emphasizes that both of these steps are required for effective proximity biotinylation (Figure 5.5C).



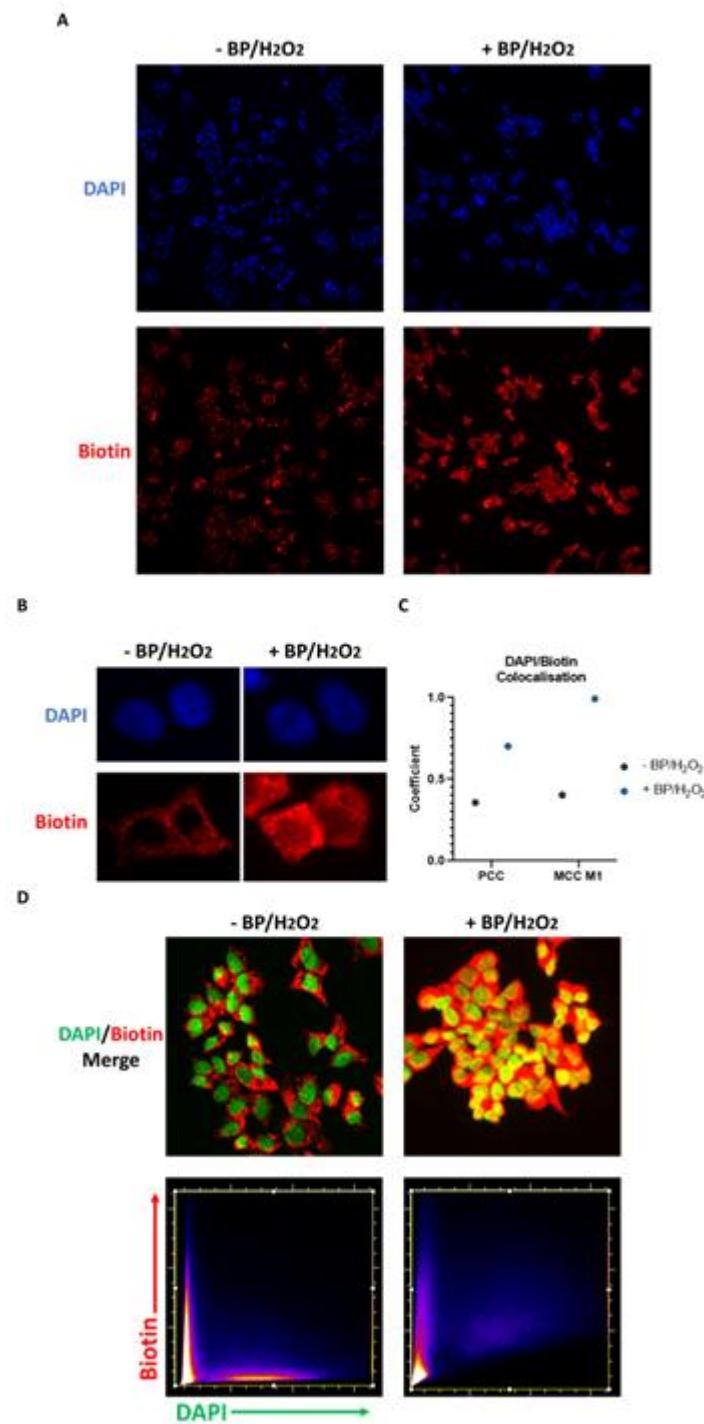
**Figure 5.5 - Nuclear/cytoplasmic fractionation enables effective enrichment of biotinylated proteins in CWR22Rv1(dCasRx-APEX2)**

**A.** 22Rv1(dCasRx-APEX2) were induced with 1  $\mu$ g/ml doxycycline for 72 hours before proximity biotinylation was performed with BP and H<sub>2</sub>O<sub>2</sub> for 2 hours and 2 minutes, respectively. After washes with quenching buffer, cells were subject to cytoplasmic-nuclear fractionation and separate subcellular fractions were enriched with streptavidin and eluted as optimised previously. Western blot analysis of biotin signal in input (I) and eluted (E) fractions of each subcellular compartment was performed. An unlabelled control induced with doxycycline but untreated with BP/H<sub>2</sub>O<sub>2</sub> was included **B.**  $\alpha$ -tubulin and AR-Vs were used as markers of cytoplasmic (C) and nuclear (N) protein content in input lysate samples from (A). Ponceau was used to confirm protein loading **C.** 22Rv1(dCasRx-APEX2) were induced with doxycycline and untreated with BP/H<sub>2</sub>O<sub>2</sub>, treated and not induced,

or both induced and treated as in (A) before nuclear extraction was performed and biotinylated protein content was analysed by Western blot. Ponceau was used to confirm protein loading

These results were further substantiated by immunofluorescence, as abundant cytoplasmic signal is observed in CWR22Rv1(dCasRx-APEX2) untreated with BP/H<sub>2</sub>O<sub>2</sub>, whereas only biotinylation using these reagents resulted in significant nuclear biotin staining (Figure 5.6A, B). Quantification of overlap between nuclear DAPI and biotin signal across all cells in Figure 5.6A was performed using the Fiji (Schindelin *et al.*, 2012) plugin JACoP (Bolte and Cordelières, 2006) to calculate Pearson's correlation coefficient (PCC) and Manders' M1 (MCC M1), both commonly utilised methods of fluorophore colocalisation analysis (Dunn, Kamocka and McDonald, 2011). This demonstrated a marked increase in both PCC and MCC M1 coefficients when APEX2 proximity biotinylation is performed with BP/H<sub>2</sub>O<sub>2</sub> (Figure 5.6C). The MCC M1 coefficient, or the fraction of detectable DAPI signal overlapping with biotin, increased to a near-total colocalisation, scoring 0.992 (1 represents complete signal overlap) (Figure 5.6C).

In order to further visualise colocalisation, DAPI and biotin were coloured green and red respectively. Qualitative assessment of merged images demonstrably illustrates an increase in yellow signal (representing combined green and red) in nuclei when cells were biotin-labelled (Figure 5.6D). An additional Fiji plugin, Colocalization Finder (Carl *et al.*, 2004), was used to graphically display spectral overlap. Pixel scatterplots produced by this software indicate an increase in signal overlap between DAPI (x-axis) and biotin (y-axis) with BP/H<sub>2</sub>O<sub>2</sub> treatment, seen as an upward shift of signal intensity away from the x-axis to a greater correlation with biotin (Figure 5.6D). Signal intensity on the y-axis does not see the same magnitude of shift, representing the fact that large amounts of cytoplasmic biotin will remain spatially separate from DAPI (Figure 5.6D).



**Figure 5.6 (previous page) - Immunofluorescence visualises nuclear biotinylation in CWR22Rv1(dCasRx-APEX2)**

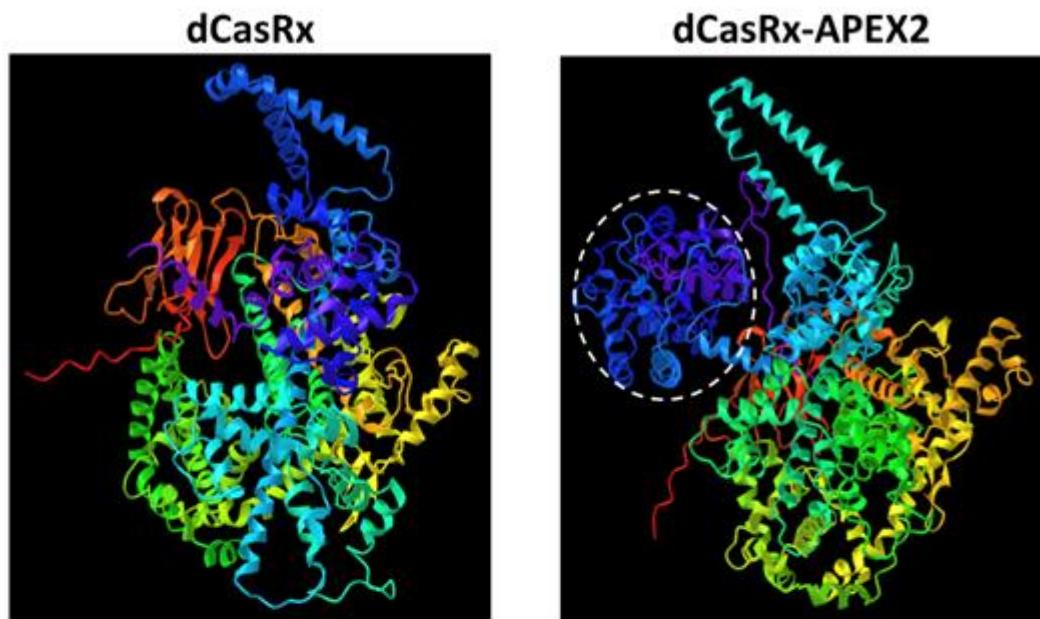
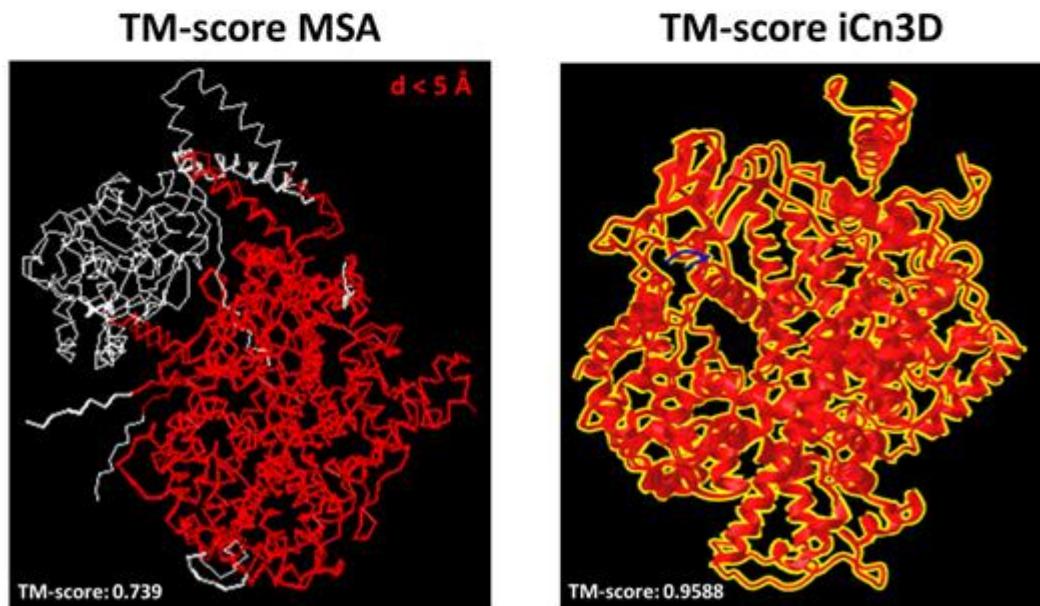
A. 22Rv1(dCasRx-APEX2) were induced with 1  $\mu$ g/ml doxycycline for 72 hours before proximity biotinylation was performed with BP and H<sub>2</sub>O<sub>2</sub> for 2 hours and 2 minutes, respectively. Cells were then fixed, permeabilised and analysed by DAPI staining and biotin immunofluorescence at 20x magnification B. Enlarged image of example cells from (A) C. Quantification of DAPI/biotin overlap was performed using ImageJ plugin JACoP, as measured by both Pearson's correlation coefficient (PCC) and Manders' M1 (MCC M1) D. Qualitative visualisation of spectral overlap between biotin and DAPI was performed. Pixel scatterplots produced by ImageJ plugin Colocalization Finder were used to quantify the increase in signal overlap

It is therefore apparent that inadequate APEX2 biotinylation and enrichment seen in CWR22Rv1 previously (Figure 4.9) was due to contamination with endogenously biotinylated proteins that reside in the cytoplasmic fraction. Extraction of nuclei vastly improved both resolution between +/- BP/H<sub>2</sub>O<sub>2</sub> treated arms and the ability to enrich proteins using streptavidin. Furthermore this observation was validated by immunofluorescence. Abundant cytoplasmic biotin signal was observed in the absence of APEX2 biotinylation, whereas only incubation with BP/H<sub>2</sub>O<sub>2</sub> resulted in appreciable nuclear staining. Nuclear extraction will thus be performed for subsequent proteomics experiments, providing a dual advantage of reducing assay background and enriching for proteins in the nucleus where alternative splicing occurs. The successfully developed transgenic CWR22Rv1(dCasRx-APEX2) will therefore enable us to leverage proximity biotinylation in a model of CRPC that expresses AR-V7.

## 5.6 dCasRx-APEX2 is targetable to CE3 mRNA

The final stage of method development before proteomics workflows could commence was to demonstrate dCasRx-APEX2 gRNA-directed targeting of CE3. This was a crucial piece of validation, as although our previous work has shown effective targeting with AR g2 based on knockdown readouts in CWR22Rv1(CasRx) (Figure 5.2), it is possible that catalytically inactive dCasRx exhibits different properties when present in our dCasRx-APEX2 fusion. We again employed AF2 structural predictions as before (Figure 4.5) to model both dCasRx and dCasRx-APEX2 (Figure 5.7A). The unfused former has previously been shown to bind mRNA in a gRNA-directed manner without cleavage due to two targeted mutations in each HEPN domain that abrogate catalytic activity (Konermann *et al.*, 2018; Du *et al.*, 2020; Cao *et al.*, 2023).

Multiple structure alignment (MSA) analyses were performed on AF2 PDB outputs to assess predicted structural similarity between native and APEX2-fused forms of dCasRx. We used TM-score (Zhang and Skolnick, 2004; Xu and Zhang, 2010) and iCn3D (Wang *et al.*, 2022) for calculation of MSA. Both programs gave significant predicted similarity between protein structures, reflected by Template modeling scores (TM-score) of 0.739 and 0.9588 respectively, whilst TM-score visualisation shows the majority of each structure to have  $< 5 \text{ \AA}$  difference between predictions (Figure 5.7B). A TM-score of 1 represents a perfect match, while scores  $> 0.5$  are indicative of proteins that significantly share folding patterns (Xu and Zhang, 2010). The higher TM-score (TM-score is a structural similarity metric, as well as a specific program as used above by Zhang and Skolnick, 2004) derived from iCn3D is likely due to this program calculating scores based on a core of aligned residues rather than all, therefore the APEX2 fusion is excluded from calculations (Figure 5.7B). The difference in score between programs recapitulates these differing metrics, as APEX2 (27 kDa) has approximately  $\frac{1}{4}$  the mass of dCasRx (112 kDa), it follows that any MSA including APEX2 would result in a TM-score  $\sim 25\%$  lower, as we observe ( $0.739/0.9588 = 77\%$ ). MSA analyses therefore provide some indication that fusion with APEX2 may not unduly perturb dCasRx function.

**A****B**

**Figure 5.7 - Multiple structure alignment analyses demonstrate high structural similarity between dCasRx and dCasRx-APEX2**

**A.** Visualisation of dCasRx and dCasRx-APEX2 AF2 structural prediction PDB files in iCn3D. Structures are coloured from red (N-terminus) to violet (C-terminus). APEX2 protein is highlighted by white dotted oval in dCasRx-APEX2 structure image **B.** Multiple structure alignment (MSA) was performed using TM-score (left, Zhang and Skolnick, 2004) and iCn3D (right, Wang *et al.*, 2022) to compare AF2-predicted dCasRx structure when native or fused with APEX2. TM-score MSA (left) displays wireframe model of dCasRx and dCasRx-APEX2 overlaid. Regions in red are those which differ by  $< 5 \text{ \AA}$  between structures. White region to the left of structure is APEX2, which has no analogous structure in unfused dCasRx. iCn3D visualisation (right) only displays residues and structural similarity of dCasRx, excluding APEX2 therefore resulting in a higher TM-score

We next sought to experimentally validate dCasRx-APEX2 functionality and ability to bind CE3 via an RNA pulldown assay. This was initially attempted by use of RNA immunoprecipitation (RIP). Here, HA-tagged dCasRx-APEX2 expression would be induced in CWR22Rv1(dCasRx-APEX2) alongside transfection with gRNAs AR g2 or NT. Cells would subsequently be formaldehyde treated to crosslink proteins with RNA, lysed to release cellular contents, lysates immunoprecipitated with an HA antibody and bound RNA extracted using TRIzol. More dCasRx-APEX2 is expected to be found at CE3 in cells transfected with AR g2 than NT gRNA, therefore RT-qPCR analysis should demonstrate an enrichment for CE3 RNA in AR g2-transfected cells (Figure 5.8).

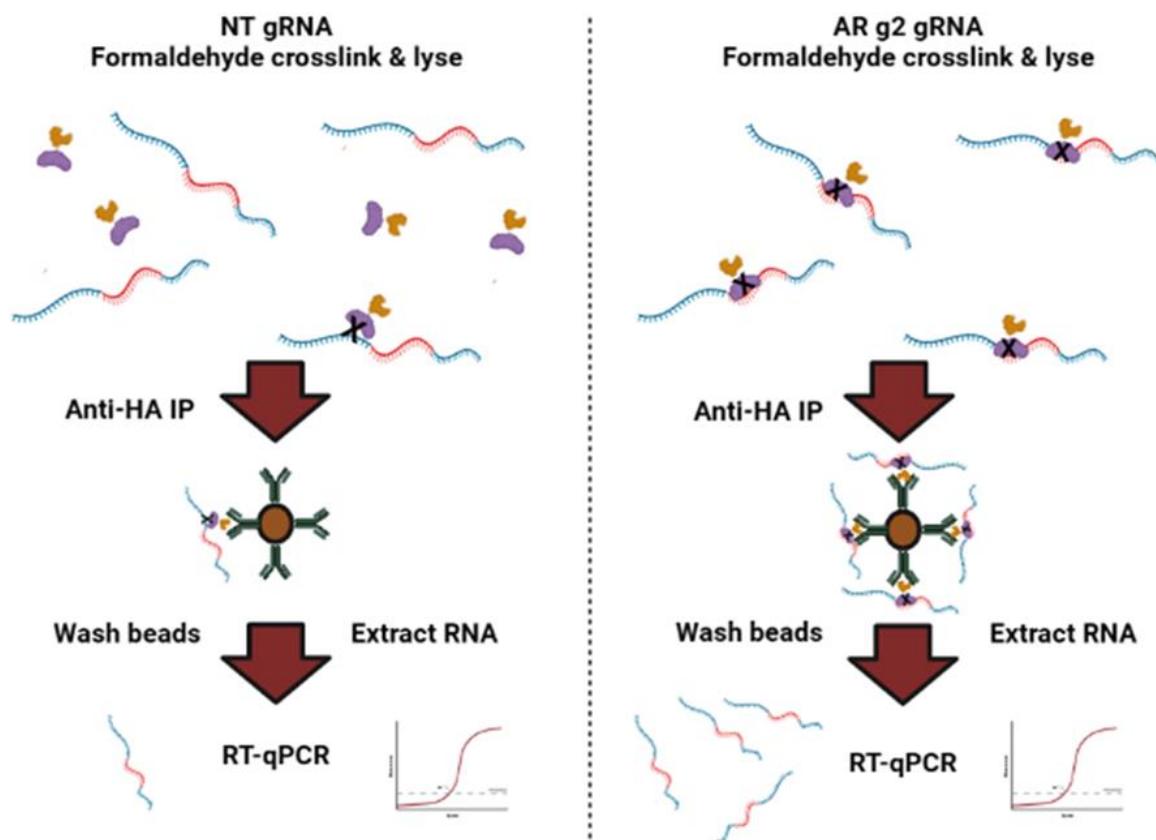


Figure 5.8 - Schematic of RNA immunoprecipitation assay to verify dCasRx-APEX2 binding with CE3 mRNA

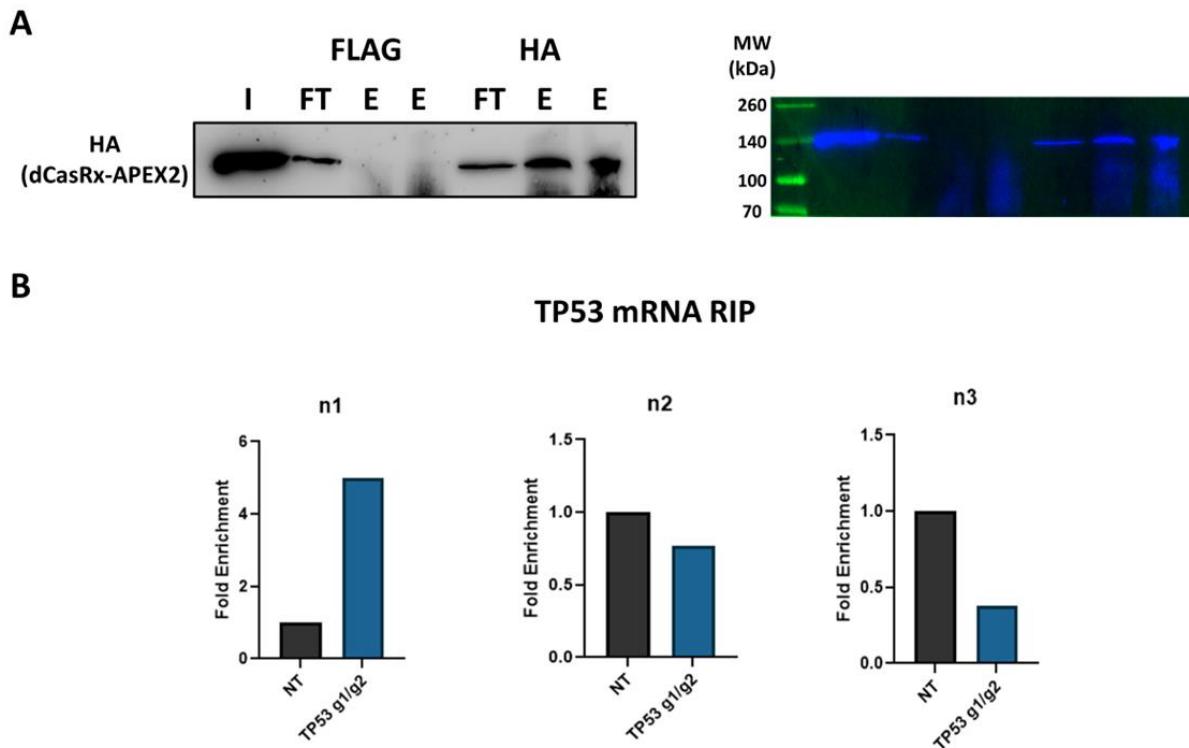
CWR22Rv1(dCasRx-APEX2) will in theory have a higher proportion of dCasRx-APEX2 (purple/orange) localised at CE3 (red) mRNA when transfected with AR g2 than NT gRNA. Formaldehyde crosslinking (black X) will stabilise these interactions, before cell lysis and immunoprecipitation (IP) of HA-tagged dCasRx using antibody-conjugated magnetic beads is performed to enrich dCasRx-APEX2-bound RNA. Subsequent washing of beads, extraction of RNA and RT-qPCR can then be used to demonstrate enrichment of CE3 mRNA and thus its interaction with dCasRx-APEX2 (Figure created using biorender.com)

The decision to follow a formaldehyde-crosslinked approach was made based on previous publications utilising dCas13 RIP, in which formaldehyde was used to fix dCas13 gRNA-mRNA interactions before lysis and immunoprecipitation (Han *et al.*, 2020; Abudayyeh *et al.*, 2016). Furthermore, this fixation of interactions was vital to validate that dCasRx-APEX2 was associating with CE3 at the point of APEX2 biotinylation. It has been shown that in performing RIP without crosslinking, *ie* native RIP, it is possible that proteins and RNAs may associate during *in vitro* mixing of lysates/antibody in ways that would not otherwise occur in their native cellular context (Mili and Steitz, 2004; Gagliardi and Matarazzo, 2016). Thus we would be unable to confirm whether dCasRx-APEX2 is interacting with CE3 in live cells or if it is an artefact of *in vitro* antibody incubations. Given that our goal is to capture live-cell protein interactions with CE3 it is essential we capture this spatial information at what would be, during proteomics experiments, the point of treating cells with BP/H<sub>2</sub>O<sub>2</sub>.

First, confirmation was required that dCasRx-APEX2 could be immunoprecipitated from cell lysates by its HA epitope tag. As initial validation, we verified that pXR002-APEX2 transfected HEK293FT lysates could be effectively immunoprecipitated, reasoning that success in CWR22Rv1(dCasRx-APEX2) is unlikely if we are unable to pull down dCasRx-APEX2 in cells expressing a far higher degree of transgene. Immunoprecipitation (IP) of dCasRx-APEX2 from HEK293FT lysates was trialled using both an HA and FLAG antibody. Two independent combinations of IP wash buffers were also tested. This demonstrated capability to pulldown dCasRx-APEX2 using an HA antibody, whereas the FLAG tag present on the fusion protein (Figure 4.5A) failed to enable IP (Figure 5.9A). A potential degree of protein degradation is apparent in the form of smearing in these blots (Figure 5.9A). Given this is only observed in immunoprecipitated samples it suggests this is likely to be a result of the IP process, possibly due to prolonged incubation at 4°C during antibody pulldown. It is notable however that a similar degree of smearing was seen in FLAG immunoprecipitated samples, despite this antibody's failure to enrich protein. Both HA and FLAG antibodies are raised in mouse, so this effect could also, rather than degradation, merely be a detection artefact from using anti-mouse secondary antibodies, which may be detecting fragmented mouse antibody/bead complexes.

Next, transfection with pXR002-APEX2 was combined with 25 nM NT gRNA or a TP53 g1/g2 combination previously shown to enact effective *TP53* mRNA depletion using CasRx plasmid

pXR001 (Figure 4.7B), followed by RIP and RT-qPCR as described above (Figure 5.8). Initial results were encouraging, with a ~5-fold enrichment for *TP53* mRNA seen in HEK293FT transfected with *TP53* g1/g2 over NT gRNA (Figure 5.9B, left panel). However, we were unable to reproduce this enrichment in subsequent replicates (Figure 5.9B middle/right panels).



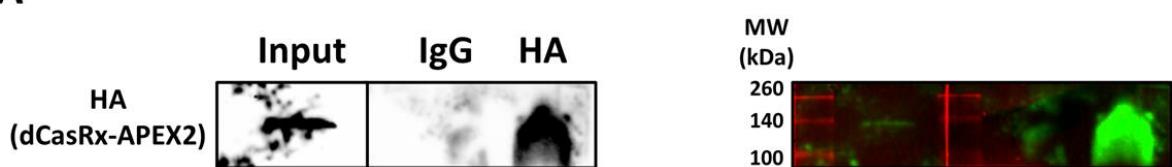
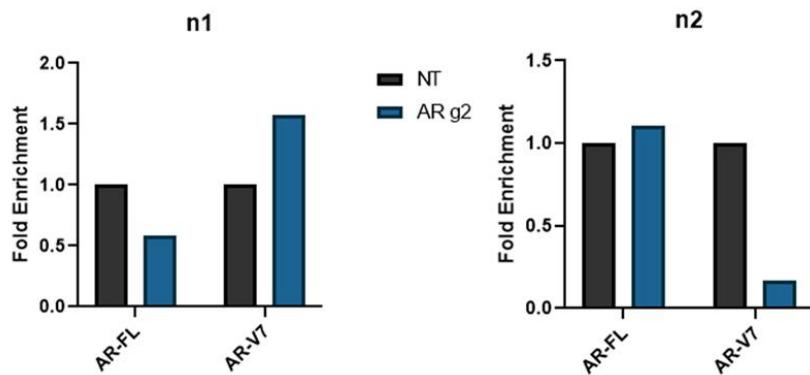
**Figure 5.9 - RNA immunoprecipitation of dCasRx-APEX2 targeted at *TP53* mRNA fails to consistently enrich *TP53* transcripts in HEK293FT**

**A.** HEK293FT were transfected with dCasRx-APEX2 expression plasmid pXR002-APEX2 and incubated for 72 hours, before being formaldehyde-crosslinked and lysed. Protein lysates were immunoprecipitated with antibodies against FLAG or HA epitope tags. Western blot of input (I), flow-through (FT) and eluted (E) fractions was performed with an HA antibody. Two lanes of eluate are present for each immunoprecipitation as two different bead-washing protocols were tested **B.** HEK293FT were transfected with pXR002-APEX2 along with either NT or *TP53* g1/g2 gRNA and incubated for 72 hours. RIP was performed as summarised in Figure 5.8, and RT-qPCR was used to compare *TP53* mRNA enrichment between NT and *TP53* g1/g2 transfected cells

As noted prior, HEK293FT express remarkably high levels of plasmid transgenes. Consequently there is potential for an amount of cellular dCasRx-APEX2 protein to exist far in excess of available *TP53* target transcripts. Accordingly, there may be significant amounts of HA-immunoprecipitated protein without an associated *TP53* transcript bound due to mRNA saturation. As we have observed by appreciable protein bands in our flow-through fraction (Figure 5.9A, 'FT'), the amount of dCasRx-APEX2 expressed may be above the limit of immunoprecipitation, therefore an element of stochasticity exists as our target RNA-bound

protein may be in unbound fractions. This could therefore explain the inconsistent nature of our results in HEK293FT although results did show some promise based on a successful n1 replicate.

Titration of input protein lysate:antibody ratios may negate this problem, however as RIP assays are ultimately being used here for confirmation of CE3 binding in CWR22Rv1(dCasRx-APEX2) it would be prudent to optimise RNA pulldown assays in this cell line. Experiments in HEK293FT have at a minimum confirmed that dCasRx-APEX2 can be immunoprecipitated by its HA tag, which could now be applied to our transgenic model of CRPC in which lower levels of dCasRx-APEX2 are expressed. To test this approach in CWR22Rv1(dCasRx-APEX2), cells were induced with doxycycline before being lysed and lysates immunoprecipitated using an HA antibody. Comparison with an isotype-matched IgG control demonstrated effective IP of dCasRx-APEX2 (Figure 5.10A). This was then performed alongside transfection with 25 nM NT or AR g2 gRNA, before RIP and RT-qPCR were performed as done previously in HEK293FT; in this instance using primers to quantify AR-FL or AR-V7 mRNA where the latter should be enriched using AR g2. Despite somewhat encouraging initial enrichment showing ~1.5-fold over NT when assaying for AR-V7 (Figure 5.10B, left panel), once more consistent enrichment was not achieved (Figure 5.10B, right panel).

**A****B**

**Figure 5.10 - RNA immunoprecipitation of dCasRx-APEX2 targeted at CE3 mRNA fails to consistently enrich AR-V7 transcripts in CWR22Rv1(dCasRx-APEX2)**

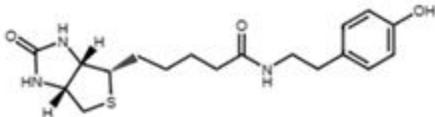
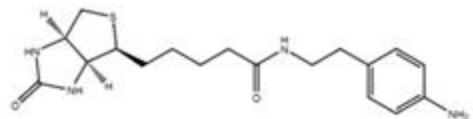
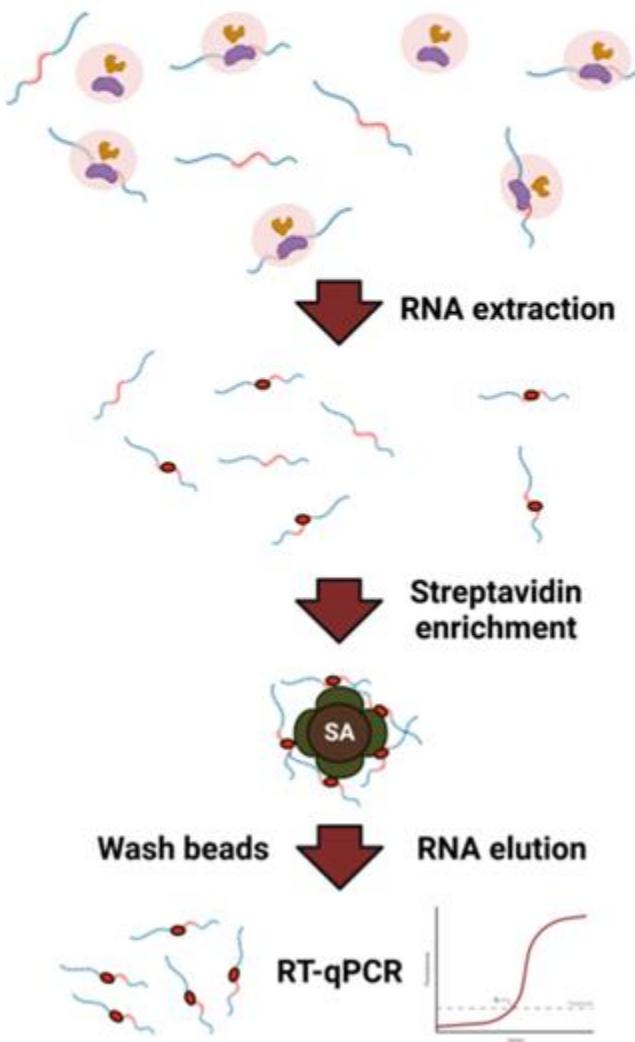
A. CWR22Rv1(dCasRx-APEX2) were induced with doxycycline for 72 hours, before being formaldehyde-crosslinked and lysed. Protein lysates were immunoprecipitated with either an HA antibody or isotype-matched IgG control. Western blot of input and pulled-down protein fractions was performed with an HA antibody. B. CWR22Rv1(dCasRx-APEX2) were induced with doxycycline alongside transfection with either NT or AR g2 gRNA and incubated for 72 hours. RIP and RT-qPCR was performed as summarised previously (Figure 5.8), in this instance using qPCR primers specific to AR-FL and AR-V7

Due to only achieving sporadic enrichment of targets in HEK293FT and CWR22Rv1(dCasRx-APEX2), a change in experimental approach was considered. It is also pertinent to note that formaldehyde concentrations used in our approach needs careful consideration, as concentrations as low as 0.3% can trap significant amounts of material in the insoluble protein fraction that is inappropriate for enrichment by IP (Patton *et al.*, 2020; Klockenbusch and Kast, 2010). Lowering our concentration from 0.2% may have aided matters, particularly as our study of an actively spliced exon could increase the chance of crosslinking with lower-solubility chromatin components, for example histones, due to splicing being tightly linked with transcription (Khan, Ng and Cheung, 2020; Merkhofer, Hu and Johnson, 2014).

A review of the literature revealed a powerful alternative that could address these challenges whilst also providing higher confidence going into proteomics workflows. Simultaneous publications from two independent groups reported on APEX-Seq, demonstrating that APEX2

proximity biotinylation of RNA is possible in addition to protein. Biotinylated RNA, likely to be proximal to their APEX2 fusion, was subsequently enriched with magnetic streptavidin beads and sequenced to characterise spatial organisation of transcripts (Padrón, Iwasaki and Ingolia, 2019; Fazal *et al.*, 2019). Importantly, these publications performed qPCR analysis of enrichment prior to RNA-Seq to confirm enrichment with their APEX2 constructs. Further work performed by others using APEX2 examined a range of aromatic biotin conjugates, in doing so identifying biotin-aniline (BAn) to mediate substantially greater live-cell RNA labelling than BP as used in the above APEX-Seq studies, again utilising streptavidin enrichment followed by RT-qPCR to confirm target RNA enrichment (Zhou *et al.*, 2019).

Based on this literature, an approach was conceived in which BAn would be used *in lieu* of BP (Figure 5.11A) in CWR22Rv1(dCasRx-APEX2) transfected with AR g2 or NT gRNA to demonstrate successful dCasRxAPEX2 binding at CE3-containing transcripts. Here, rather than immunoprecipitating protein, total RNA would be extracted and enriched using streptavidin, relying on APEX2/BAn-mediated covalent biotinylation of target RNA(s) (Figure 5.11B). Analogous to our prior efforts using formaldehyde-crosslinked immunoprecipitation, RT-qPCR should demonstrate an increased abundance of CE3 in cells transfected with AR g2 over NT gRNA. Should this approach yield success, it presents a significant advantage in that CE3 RNA enrichment will have been demonstrated using APEX2 proximity biotinylation chemistries, offering greater confidence that it will reflect the cellular context in which proteomics is performed.

**A****Biotin-phenol (BP)****Biotin-aniline (BAn)****B****22Rv1(dCasRxAPEX2-TLCV2) + BAn/H<sub>2</sub>O<sub>2</sub>**

**Figure 5.11 - Biotin-aniline mediated APEX2 proximity biotinylation provides an effective means to validate dCasRx-APEX2 binding at CE3 mRNA**

**A.** Chemical structures of biotin-phenol (BP) and biotin-aniline (BAn). Differing only by an amine (-NH<sub>2</sub>) in place of a hydroxyl (-OH) group bound to its aromatic ring, BAn mediates far greater proximity biotinylation efficiency of RNA than BP. **B.** Incubation of CWR22Rv1(dCasRx-APEX2) with BAn/H<sub>2</sub>O<sub>2</sub> would enable live-cell proximity biotinylation of RNA (pink circles). Total RNA can then be extracted and biotinylated transcripts (with red circles) enriched using streptavidin. Stringent washes are enabled by use of streptavidin, and elution of RNA followed by RT-qPCR will ascertain target RNA abundance. Transfection with AR g2 gRNA using this approach would in theory yield greater enrichment of CE3 mRNA than NT gRNA (Figure created using biorender.com)

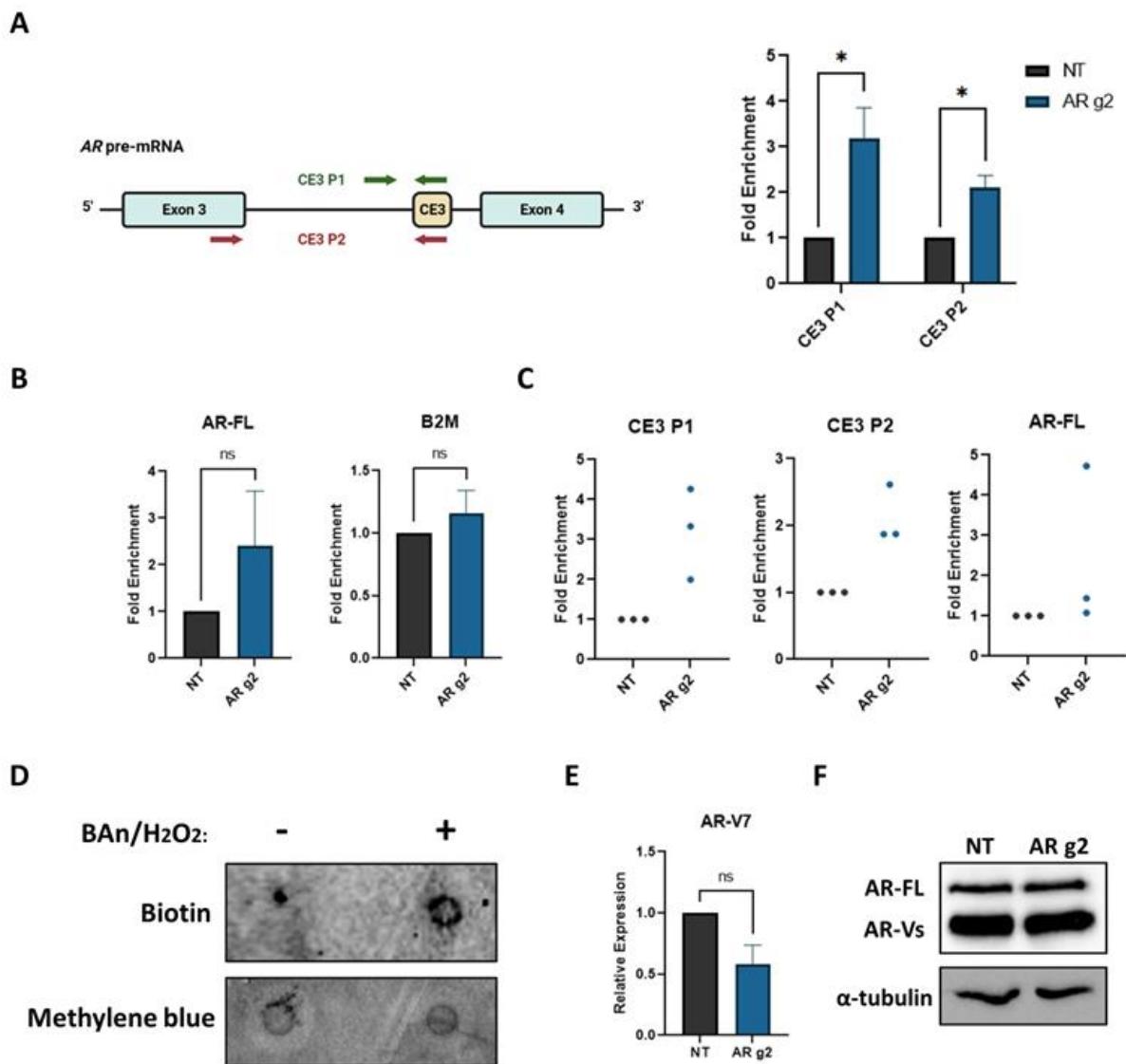
We tested this novel approach, again by inducing CWR22Rv1(dCasRx-APEX2) in concert with NT or AR g2 gRNA transfection. After 72 hours, a respective 2-hour/2-minute incubation was performed as previously, this time using BAn/H<sub>2</sub>O<sub>2</sub> in place of BP. Total RNA was extracted using TRIzol, a portion of which was retained for use as an input sample, before equal masses of the remaining sample RNAs were enriched for biotinylated transcripts using streptavidin beads. Stringent bead washing and RNA elution followed by RT-qPCR of both input and streptavidin-enriched samples thus enabled comparison of enrichment levels between NT and AR g2-transfected samples, both of which are normalised to their respective input.

Our results demonstrated a specific enrichment of CE3 using two independent primer pairs representing both *AR* pre-mRNA and mature AR-V7 transcripts, as would be predicted if dCasRx-APEX2 was active in the nuclear compartment (Figure 5.12A). This also offers encouragement that our workflow will be capable of characterising the CE3 interactome during active splicing events, rather than only targeting post-spliced AR-V7. Furthermore as predicted, our experiment failed to consistently enrich either AR-FL mRNA or negative control *B2M* transcripts (Figure 5.12B). It bears noting that although AR-FL appears to exhibit some level of enrichment (failing to reach significance at  $\alpha$  0.05), examination of individual values highlights a single replicate skewing results whereas CE3 P1 and P2 are consistently enriched across three biological replicates (Figure 5.12C). Exons encoding AR-FL are part of the same *AR* pre-mRNA transcript as CE3, so it may be that for this single replicate a level of dCasRx-APEX2 mislocalisation occurred, potentially due to factors such as variability in gRNA transfection efficiency. The precise reasons why this single AR-FL enrichment may have taken place are unclear, although crucially the wholly unrelated, yet highly expressed, *B2M* transcript never exhibited enrichment in AR g2-transfected samples (Figure 5.12B).

Additionally, we set up a parallel experiment in which dCasRx-APEX2 and BAn/H<sub>2</sub>O<sub>2</sub> RNA biotinylation was performed in the absence of gRNA. This was enriched as above and eluted RNA used to visualise biotinylation by RNA dot blot, demonstrating a detectable signal in RNA derived from BAn/H<sub>2</sub>O<sub>2</sub>-treated cells (Figure 5.12D). Interestingly a degree of AR-V7 knockdown was observed in AR g2 input samples (Figure 5.12E), though this was not to the same extent as in CWR22Rv1(CasRx) (Figure 5.2). However, no such reduction in AR species is seen at the protein level in analogous samples (Figure 5.12F). A possible explanation for this discrepancy is that RNA biotinylation may interfere to some degree with reverse transcription

reactions. Literature on APEX-Seq used an RT stop assay to evaluate the effect of biotinylation, noting that although it still enables production of full-length products an increased incidence of truncated cDNA occurs (Fazal *et al.*, 2019). Were this the case in our system, it provides an additional layer of evidence that dCasRx-APEX2 is binding at *AR* transcript CE3.

In summary dCasRx-APEX2 appears, through use of gRNA AR g2, to be binding at CE3 of *AR* mRNA as expected. Attention will now turn to application of our system for proteomics experiments and ultimately definition of the CE3 interactome in CWR22Rv1(dCasRx-APEX2).



**Figure 5.12 - RNA biotinylation with biotin-aniline enables a CE3 pulldown assay in CWR22Rv1(dCasRx-APEX2)**

A. CWR22Rv1(dCasRx-APEX2) were transfected with either NT or AR g2 gRNA and induced with doxycycline for 72 hours before being biotinylated by incubation with BAn and H<sub>2</sub>O<sub>2</sub> for 2 hours and 2 minutes, respectively. Total RNA was extracted and an RNA pulldown assay was performed with streptavidin as summarised in Figure 5.11. RT-qPCR was used with the indicated primer sets to quantify levels of pre-spliced (CE3 P1) and post-spliced (CE3 P2) CE3 mRNA enrichment between AR g2 and NT-transfected samples. Data comprises n = 3 independent

biological replicates, plotted as mean  $\pm$  SEM. Unpaired t-test was used for determination of statistical significance (\* =  $p < 0.05$ ) **B**. Enrichment data for experiment summarised in (A), using primers specific to AR-FL and B2M mRNA. Data comprises  $n = 3$  independent biological replicates, plotted as mean  $\pm$  SEM. Unpaired t-test was used for determination of statistical significance (ns =  $p > 0.05$ ) **C**. Plot of individual biological replicate enrichment values for qPCR data from (A) and (B), for the primer sets indicated **D**. Live-cell BAn/H<sub>2</sub>O<sub>2</sub> proximity biotinylation was performed in 22Rv1(dCasRx-APEX2) as in (A), followed by streptavidin enrichment of total RNA extracts. A negative control induced with doxycycline but untreated with BAn/H<sub>2</sub>O<sub>2</sub> was also included. RNA dot blot was performed using streptavidin-HRP. Methylene blue stain was used to confirm presence of RNA in each sample **E**. Input samples from pulldown assays were analysed for AR-V7 levels by RT-qPCR. qPCR data comprises  $n = 3$  independent biological replicates, plotted as mean  $\pm$  SEM. Unpaired t-test was used for determination of statistical significance (ns =  $p > 0.05$ ). **F**. Western blot depicting levels of AR-FL and AR-V protein in 22Rv1(dCasRx-APEX2) induced and transfected with the same gRNAs as in RNA-pulldown experiments.  $\alpha$ -tubulin was used as a loading control. Protein samples are sourced from proteomics experiment inputs

## 5.7 Proteomics workflows successfully enrich splicing proteins and expected CE3 mRNA interactors

With successful demonstration of dCasRx-APEX2 binding at CE3 through use of AR g2 completed, application of CWR22Rv1(dCasRx-APEX2) to proteomics workflows could begin. Our experimental setup will encompass three arms including a non-biotinylated control, as well as NT and AR g2 gRNA-transfected plates analogous to the above RNA pulldown assay (Figure 5.13). Efforts so far to optimise dCasRx-APEX2 expression, gRNA delivery, proximity biotinylation and streptavidin enrichment should enable successful transition to proteomics experiments. In order to increase protein quantities for mass spectrometry, material will be significantly upscaled to 1 x 150 mm dish per condition.

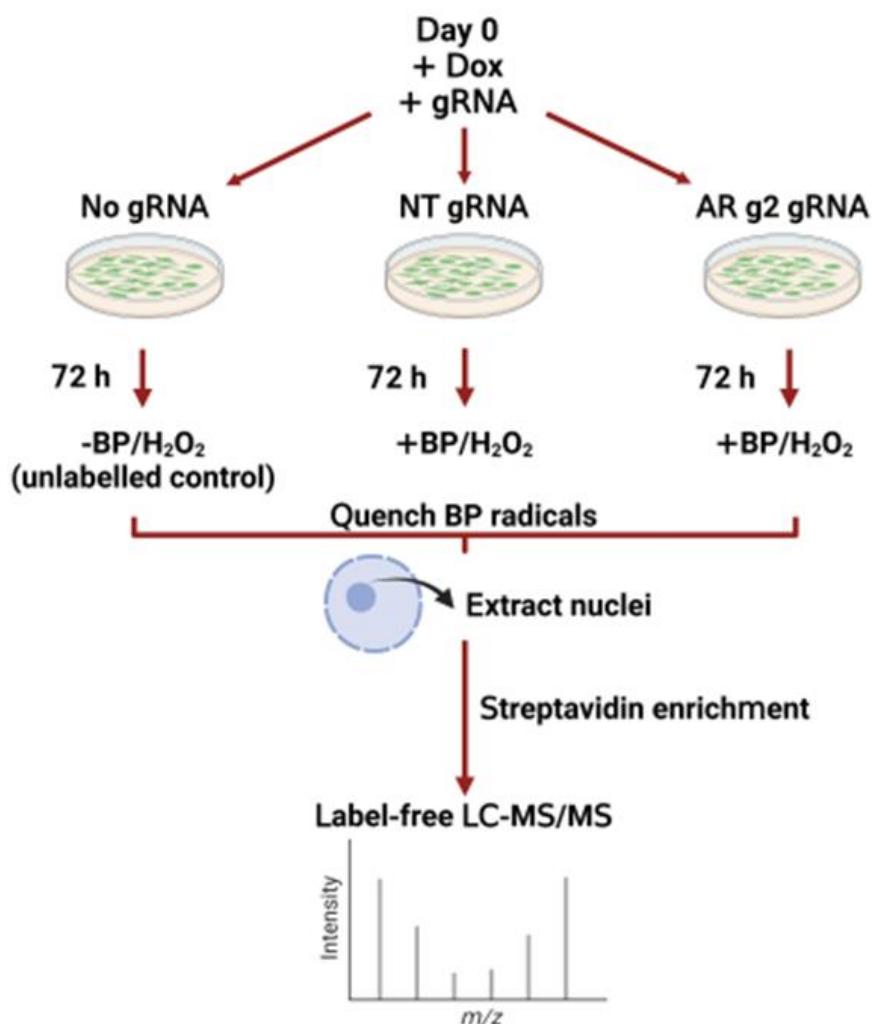


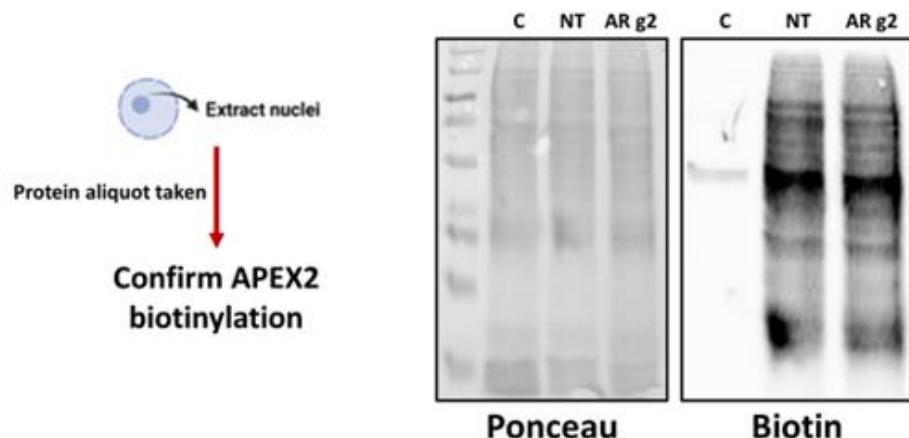
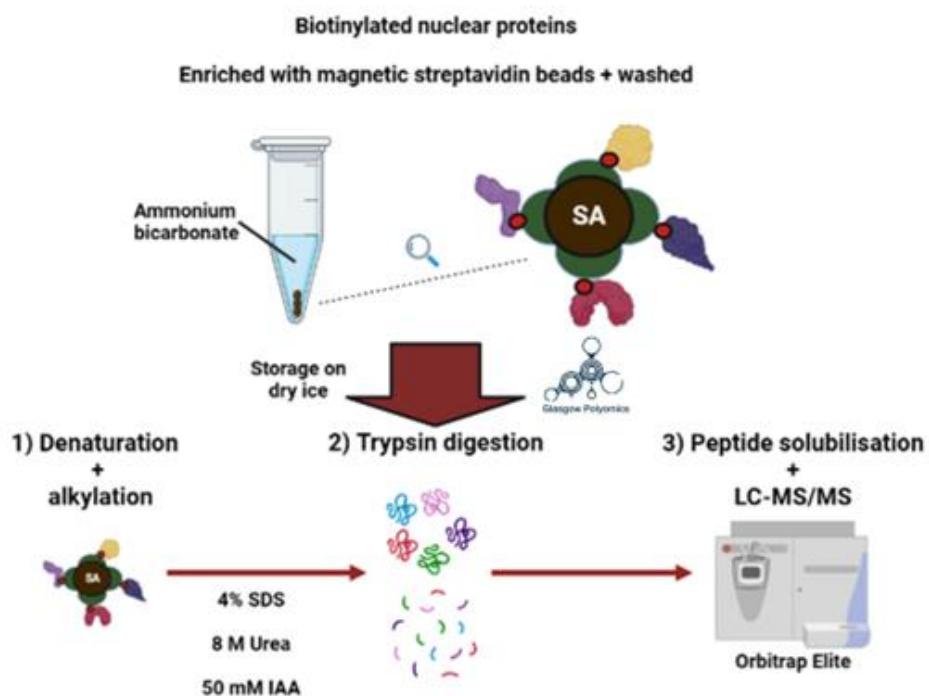
Figure 5.13 - CWR22Rv1(dCasRx-APEX2) CE3-targeted proteomics experimental setup

Three experimental arms are to be utilised for proteomics workflows. CWR22Rv1(dCasRx-APEX2) will be simultaneously transfected with relevant gRNAs and dCasRx-APEX2 expression induced with doxycycline. 72 hours later, APEX2 proximity biotinylation will be performed as before using BP/H<sub>2</sub>O<sub>2</sub>. An unlabelled control, in

which dCasRx-APEX2 expression is induced without subsequent APEX2 labelling, will also be included. Biotin-phenoxyl radical quenching washes, nuclear extraction and streptavidin enrichment will all be performed according to our previous optimisation, before samples are submitted for tryptic digest and LC-MS/MS (Figure created using biorender.com)

We opted to use label-free mass spectrometry. Label-free approaches have been shown to provide deeper proteome coverage and a higher number of protein identifications compared to isobaric tagging methods or isotope-labelled amino acids, though they do sacrifice a degree of quantification accuracy due to higher variability (Li *et al.*, 2012; Rozanova *et al.*, 2021; Stepath *et al.*, 2020). Numerous studies utilising proximity biotinylation, whether with BiOLD, APEX2 or the more recent TurboLD, have chosen this method with demonstrable success (Barnes *et al.*, 2022; Lamb *et al.*, 2016; Peterson *et al.*, 2021; Hobson *et al.*, 2022; Santos-Barriopedro, van Mierlo and Vermeulen, 2021). Furthermore, this approach has been effectively leveraged in other dCas13 and dCas9 publications with aims similar to this project (Lin *et al.*, 2021; Zhang *et al.*, 2020; Gao *et al.*, 2018).

Our approach outlined in Figure 5.13 was adopted for  $n = 4$  biological replicates to help alleviate potential variability. A nuclear lysate input sample was retained to confirm successful dCasRx-APEX2 biotinylation (Figure 5.14A), before equal amounts of 100 - 250  $\mu$ g nuclear protein (nuclear lysate concentrations varied across replicates) were used per streptavidin enrichment, performed as previously. Beads were washed in PBS and ammonium bicarbonate after post-enrichment RIPA washes in order to remove residual detergent that may interfere with mass spectrometry (Yeung *et al.*, 2008). Pelleted streptavidin beads bound to biotinylated proteins in 50 mM ammonium bicarbonate were subsequently stored on dry ice and sent to our proteomics service provider, Glasgow Polyomics. Once received, a filter-aided sample preparation (FASP) protocol was followed to denature, alkylate, and digest streptavidin-bound proteins into peptides, which were solubilised and used for nanoflow high-performance liquid chromatography (HPLC), electrospray ionisation and tandem mass spectrometry (MS/MS) (Figure 5.14B). Our primary goal with this  $n = 4$  was to give preliminary indications regarding workflow effectiveness and protein enrichment profiles using AR g2 in CWR22Rv1(dCasRx-APEX2).

**A****B**

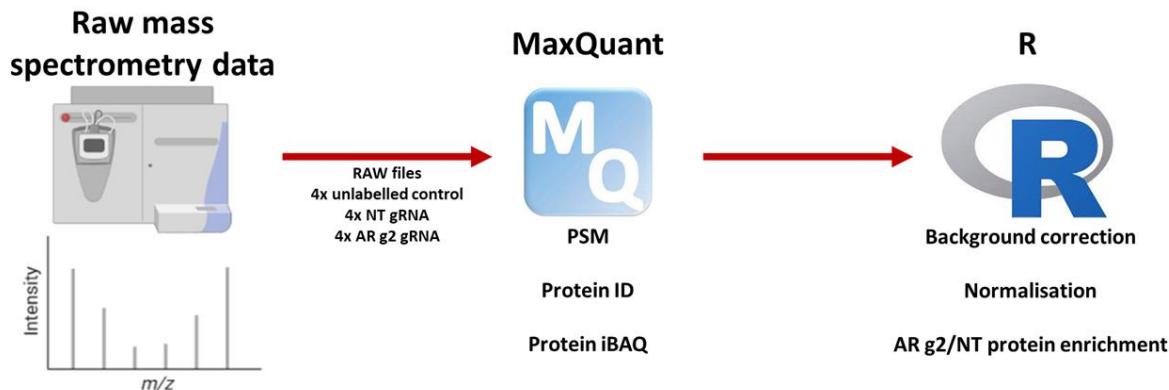
**Figure 5.14 - Nuclear lysates from proteomics experimental arms, detailed in 5.14, are utilised for LC-MS/MS**

**A.** After applying the experimental setup detailed in Figure 5.13, a portion of extracted nuclear protein is retained for SDS-PAGE and Western blot for biotinylated proteins to confirm APEX2 labelling (C = unlabelled control). Western blot is representative of  $n = 4$  proteomics experimental inputs. Ponceau was used to confirm equal protein loading **B.** Biotinylated nuclear proteins, enriched with magnetic streptavidin and washed, are retained in ammonium bicarbonate and frozen on dry ice before being sent to our proteomics service provider, Glasgow Polyomics, for filter-aided sample preparation (FASP). In doing so, samples were denatured and alkylated, digested into peptides with trypsin and solubilised for use in nanoflow high-performance liquid chromatography (HPLC), electrospray ionisation and tandem mass spectrometry (MS/MS) on an Orbitrap Elite instrument (SDS = sodium dodecyl sulphate, IAA = iodoacetamide) (Figure created using biorender.com)

For each experimental arm Glasgow Polyomics returned RAW files, a proprietary data type that contains all spectral and fragmentation data obtained from Thermo Orbitrap instruments. RAW files are amenable to analysis using MaxQuant, a popular software suite for mass spectrometry data analysis (Cox and Mann, 2008). MaxQuant employs an inbuilt peptide search engine, Andromeda, which applies a probabilistic scoring model to evaluate peptide-spectrum matches (PSM), or the likelihood of a given input spectrum to be derived from a particular peptide based on comparison with that peptide's theoretical spectrum. (Cox *et al.*, 2011). Theoretical spectra are derived from *in silico* simulated trypsin digest of a user-supplied proteome amino acid FASTA (in our case the *Homo sapiens* Uniprot database). Furthermore, peptides are assigned to proteins based on this same database.

A target-decoy approach is also utilised by MaxQuant to generate false-discovery rates (FDR) for peptide and protein assignment, whereby the proteome sequences are reversed to create a decoy database in which no 'true' matches would be expected. A statistical assumption is made that the number of matches for a peptide or protein arising from this decoy database will match the number of false positives from the original database. The proportion of false to true positives can thus be calculated as a percentage and any identifications above a specified FDR cut-off, typically 1% which we used here, are discounted (Cox *et al.*, 2011; Nesvizhskii *et al.*, 2010).

MaxQuant can also be applied to protein quantitation. An analysis pipeline was devised whereby this will be used to generate intensity based absolute quantification (iBAQ) values. Protein iBAQ is defined as the summed intensity, or abundance, of all peptides assigned to a protein divided by the number of theoretically observable tryptic peptides for that protein (Schwanhäusser *et al.*, 2011). The fold enrichment of proteins in AR g2 over NT samples can then be calculated using these values. Furthermore, unlabelled control iBAQ values will be subtracted from the latter two arms to correct for nonspecific contaminant binding. The R language for statistical software will be used for all quantitative processing of protein iBAQ values (v4.1.2, R Core Team, 2021) via the RStudio development environment (v2022.07.0, RStudio Team, 2022) (Figure 5.15).

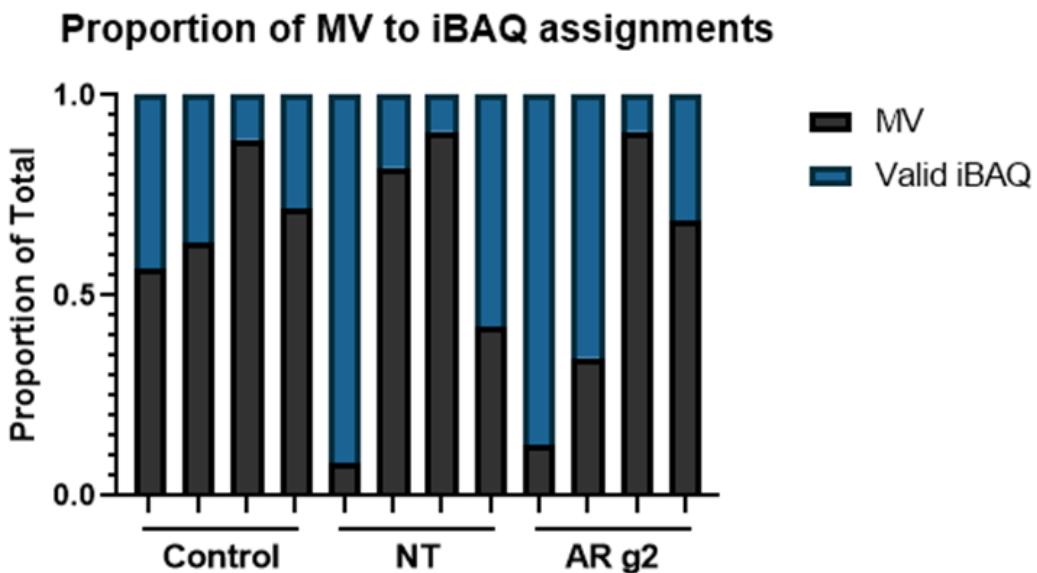


**Figure 5.15 - Simplified summary of computational analysis pipeline for mass spectrometry data**

Our experimental setup completed for  $n = 4$  biological replicates will result in 12 RAW mass spectrometry files, usable as inputs for MaxQuant software. MaxQuant will be used to determine peptide-spectrum matches (PSM), assignment of peptides to protein IDs, and calculation of intensity based absolute quantification (iBAQ) values for all identified proteins. These will subsequently be read into R, where a wide range of statistical software packages can be applied for background correction using unlabelled control values, normalisation and ultimately calculation of protein enrichment in AR g2 over NT samples

A prominent issue faced when using label-free proteomics workflows is the presence of missing values (MVs). These occur when there is insufficient information to assign detected ions to a peptide or protein resulting in an intensity score of 0, which may be due to variability in technical workflows such as trypsin mis-cleavage or ion suppression which have an element of randomness (Lazar *et al.*, 2016). More relevant to our analysis are non-random MVs arising from the inherent detection limits of mass spectrometry and associated computational analyses. Because peptide/protein assignment with MaxQuant is an FDR-based approach, it is possible that peptide ions are present but lack the intensity and statistical confidence to be assigned a PSM (Li and Smyth, 2023).

MaxQuant output tables were filtered to remove common laboratory contaminants such as keratin, identification of which is built into MaxQuant. We also adopted a conservative approach for controlling potential nonspecific contaminants by filtering for proteins identified by at least two unique peptides (Cassidy *et al.*, 2021; Higdon and Kolker, 2007). Preliminary assessment of our filtered dataset revealed a high prevalence of MVs, which will impede quantitative analysis of protein enrichment between experimental arms (Figure 5.16).



**Figure 5.16 - Proportion of MVs and valid assigned iBAQ values from initial n = 4 MaxQuant processing**

Relative proportion of missing values (MV) and successfully assigned iBAQ (valid iBAQ) scores in MaxQuant protein quantification outputs for each experimental arm. Samples are presented by experimental arm, as n1 - n4 from left to right

Despite this apparent limitation of our preliminary dataset, to garner some level of insight into workflow effectiveness we proceeded with quantitative enrichment analysis between AR g2 and NT experimental arms. First, control iBAQ values were subtracted from NT and AR g2 arms, and any resulting negative iBAQ values were set to 0. Next, due to MV prevalence we used a lenient threshold of retaining proteins that were detected in at least two out of four replicates of either gRNA sample.

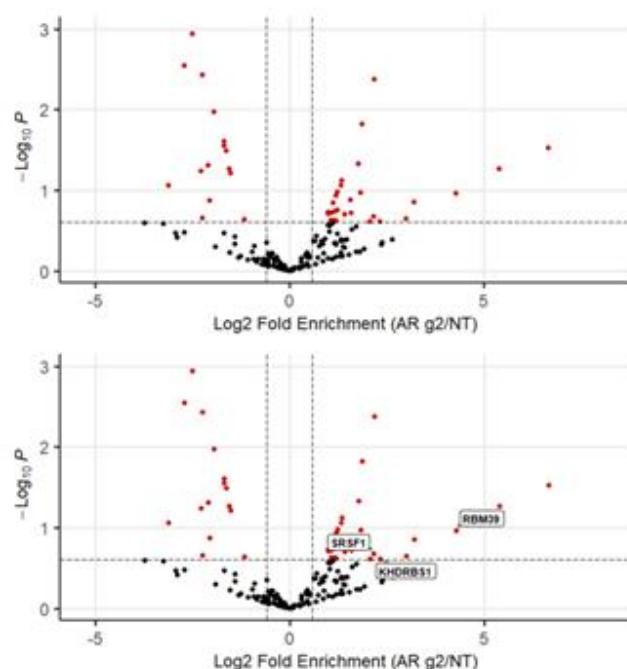
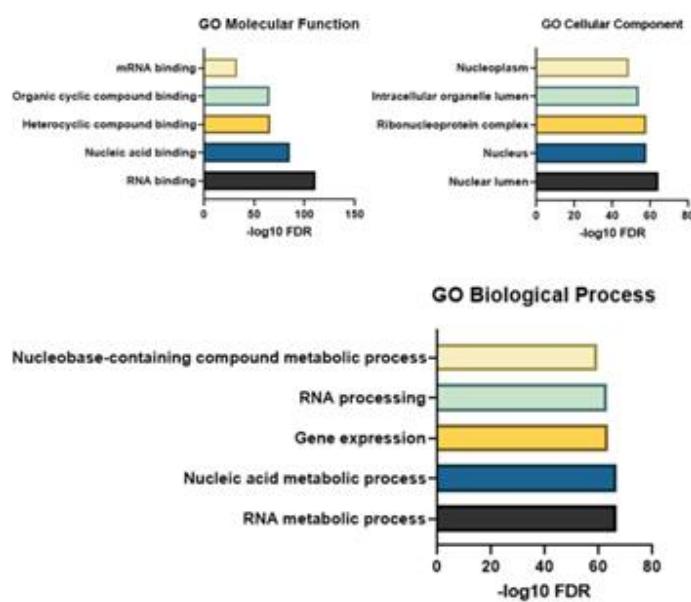
Remaining MVs were then imputed. The most appropriate strategy for MV imputation is subject to debate (Webb-Robertson *et al.*, 2015). We presumed that our MVs are non-random as a result of missing the LC-MS/MS detection limit, thus we opted for a simple imputation approach whereby MVs are replaced with half the minimum iBAQ value in its AR g2/NT replicate pair. This has proven effective in left-censored datasets (*ie* when MVs are skewed towards the lower end of the dataset, as is likely the case here) (Wei *et al.*, 2018).

In this initial n = 4 dataset, the amount of digested peptide input injected for LC-MS/MS was not quantified which may also account for the significant inter-replicate variability in number of MVs. For these samples, equal concentrations of initial nuclear lysates were first added to streptavidin beads for enrichment, washing and digestion before a fixed volume of tryptic

digest was loaded. Due to the unknown concentrations of eluted protein, as an attempt to normalise for differing peptide injection amounts we calculated relative iBAQ (riBAQ) for each protein, in which iBAQ protein values are divided by the sum of all iBAQ values in that sample (Krey *et al.*, 2018). It must be noted however that MVs will reduce riBAQ accuracy.

Log2 transformed riBAQ values were subsequently compared between AR g2 and NT arms for calculation of fold enrichment using a paired test design with R Bioconductor package *limma* (Ritchie *et al.*, 2015). Originally developed for microarrays, the linear statistical models applied by *limma* are appropriate for quantitative mass spectrometry and have been used effectively in label-free CRISPR-APEX2 workflows analogous to that applied here (Gao *et al.*, 2018; Lin *et al.*, 2021). Encouragingly, previously identified regulators of AR-V7 splicing RBM39, KHDRBS1 (Sam68) and SRSF1 (SF2) (Melnyk *et al.*, 2020; Stockley *et al.*, 2015; Liu *et al.*, 2014) were enriched in AR g2-transfected samples relative to NT, albeit relatively weakly ( $p < 0.25$ ). (Figure 5.17A).

Although our preliminary dataset lacked sufficient power to effectively analyse AR g2 enrichment trends, by taking the complete filtered list comprising 187 identified proteins used for *limma* as inputs for gene ontology (GO) analysis using STRING (Szklarczyk *et al.*, 2022), we demonstrated that RNA binding and processing terms were overrepresented (Figure 5.17B). Additionally we saw significant cellular component enrichments related to the nucleus, further garnering confidence that our workflow has strong selectivity for proteins in the nuclear fraction (Figure 5.17B). Therefore, early indications were that dCasRx-APEX2 is at a minimum enriching for RNA-associated proteins in the nucleus.

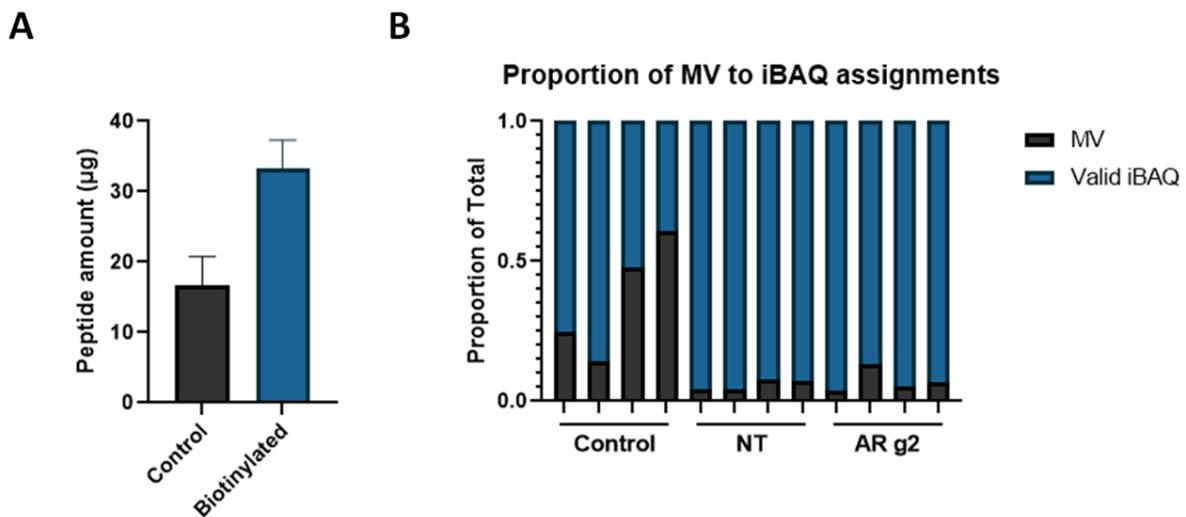
**A****B**

**Figure 5.17 - CE3-targeted dCasRx-APEX2 proteomics enriches previously published AR-V7 splicing regulators and selects for nuclear/RNA-binding proteins**

**A.** Volcano plot displaying log2 fold enrichment (x-axis) and  $-\log_{10} p$ -values (y-axis) for protein enrichment between AR g2 and NT arms in our preliminary  $n = 4$  proteomics experiment, calculated by *limma*. Previously identified AR-V7 regulators RBM39, KHDRBS1 and SRSF1 are labelled. Enrichment cutoffs for visualisation are *limma* p-value (unadjusted)  $< 0.25$  and linear fold enrichment  $> 1.5$  ( $\log_2 0.585$ ) **B.** STRING GO term analysis of all 187 proteins identified by mass spectrometry, whether statistically enriched or not, was performed. The top 5 most significant terms for each GO category are displayed, ranked by  $-\log_{10}$  STRING FDR

A low overall number of proteins compared to similar studies and a lack of statistical enrichment prompted discussion with our proteomics service provider, Glasgow Polyomics, regarding how we could enhance workflow effectiveness and improve power. As a result, two key parameters were changed. First, the amount of nuclear protein available for streptavidin enrichment was upscaled significantly by doubling all tissue culture inputs. Consequently, sufficient material was obtained to use 500 µg biotinylated nuclear lysate per pulldown, a marked increase on our initial  $n = 4$ . This subsequently provided enough downstream protein for our second adjustment: quantification of trypsin-digested peptide concentrations prior to LC-MS/MS. This approach has been successfully adopted by others to select optimal injection amounts for label-free mass spectrometry (Maia *et al.*, 2020). After consultation with Glasgow Polyomics, injection of 5 µg digested peptide per experimental arm was chosen. Prior experience from our service provider dictated that this is an optimally effective peptide concentration for their LC-MS/MS instruments, whilst injection of equal peptide amounts also provided an additional layer of normalisation between samples.

A second  $n = 4$  proteomics experiment was performed using identical experimental arms as previously and employing the above alterations. Assaying of digested peptide concentrations showed, as expected, a consistently higher amount of peptide in biotin-labelled vs. unlabelled control arms (Figure 5.18A). Furthermore, MaxQuant calculation of iBAQ values demonstrated significantly greater proportions of successful protein identifications with minimal, and substantially less variable, MVs in gRNA samples compared to initial experiments, vindicating our adapted approach (Figure 5.18B).



**Figure 5.18 - Assaying tryptic peptide concentrations enables a significantly greater proportion of MaxQuant protein identifications**

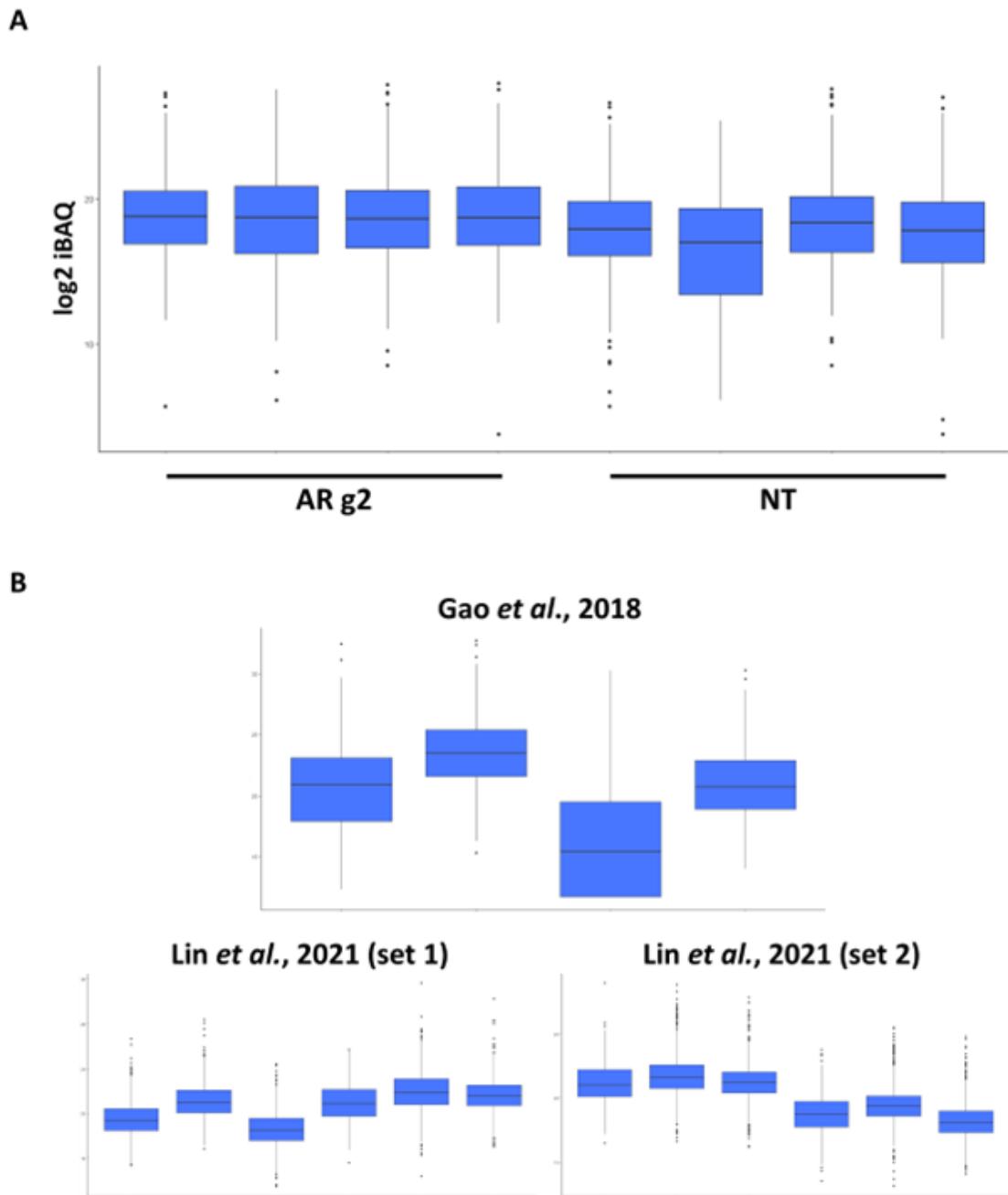
**A.** Mean total post-enrichment peptide amounts  $\pm$  SEM for  $n = 4$  proteomics input samples, assayed post-tryptic digest by Glasgow Polyomics. Biotinylated sample data points are the mean of NT and AR g2 arms **B.** The proportion of MVs to successfully assigned iBAQ values was plotted as in 5.17 for this second  $n = 4$  experiment

Generation of a more robust  $n = 4$  dataset thus enabled greater analytical power. Consequently, a more stringent protein inclusion threshold was used whereby proteins required an iBAQ value in at least 3 replicates in either gRNA arm to be candidates for imputation and enrichment analysis. Application of this threshold further demonstrated the increased power of this dataset based on number of available proteins for analysis, as after background correction and filtering for proteins as described, the resulting matrix numbered 755 proteins as opposed to 187 for our initial  $n = 4$ .

We opted not to further normalise iBAQ values beyond doing so by injecting equal digested peptide concentrations for LC-MS/MS. Label-free proteomics normalisation approaches largely rely on a background of unchanging proteins to calculate sample scaling factors (Cox *et al.*, 2014; Lee *et al.*, 2016). Given the expected differences in dCasRx-APEX2 subcellular location and hence protein interactomes between AR g2 and NT experimental arms, we lacked confidence in obtaining adequate invariant proteins for normalisation. Moreover, normalisation by riBAQ as used for our initial  $n = 4$  may also be inappropriate here. In our NT-transfected arms it would be expected that dCasRx-APEX2 would diffuse relatively freely in the nucleus, thus it may biotinylate a large range of proteins but at smaller molar amounts insufficient for PSM and iBAQ assignment. Therefore, the summation of iBAQ values may not

accurately reflect total protein, more so in NT than AR g2 samples for the aforementioned reasons. Finally, normalisation by intensity of dCasRx-APEX2 itself may also not function effectively as although a degree of APEX2 self-biotinylation would be expected, differences in protein-protein interactions and subnuclear location between samples may impact the degree to which this occurs. It is likely that other studies came to similar conclusions, as no CRISPR-APEX2 publications have normalised by the latter metric (Myers *et al.*, 2018; Han *et al.*, 2020; Lin *et al.*, 2021; Gao *et al.*, 2018).

Visualisation of protein iBAQ values subsequently used for *limma* analysis in our workflow exemplifies the benefits of our approach to normalising peptide concentrations for LC-MS/MS. Log2-transformed iBAQ intensities in AR g2 and NT samples show appreciably greater equivalence in our dataset (Figure 5.19A) than between gRNA arms in analogous label-free CRISPR-APEX2 publications that utilised the iBAQ metric (Figure 5.19B) (Gao *et al.*, 2018; Lin *et al.*, 2021). Whilst the latter publications still successfully identified and validated DNA/RNA interacting proteins by use of an appropriate gRNA, the additional normalisation utilised here by post-digest peptide quantification provides greater statistical confidence in results.

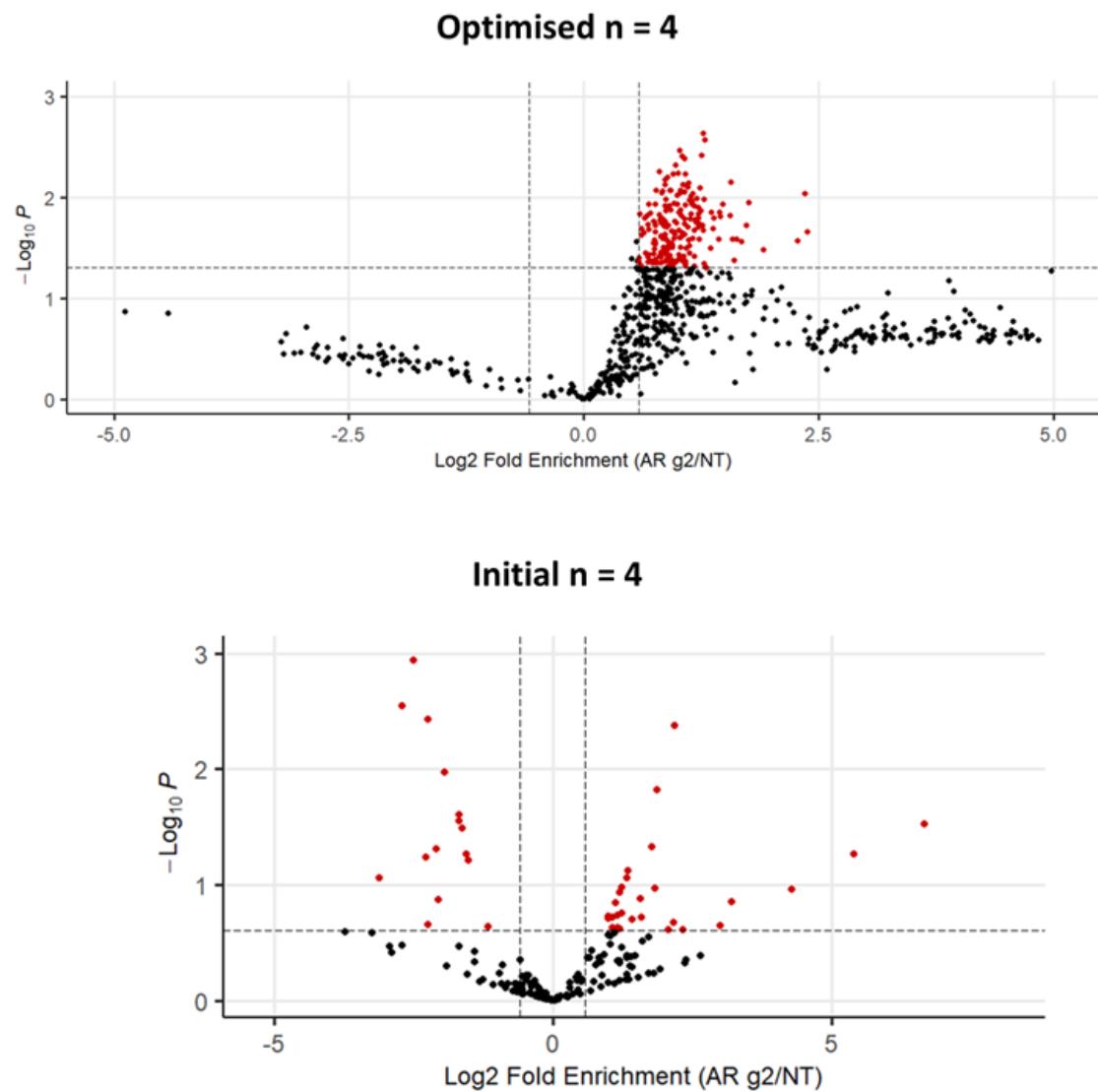


**Figure 5.19 - Injection of equal peptide concentrations for LC-MS/MS enables greater protein intensity homogeneity between samples**

**A.** Boxplot of log<sub>2</sub> iBAQ values in AR g2 and NT experimental arms, subsequently used as input for *limma* analysis. Boxplots are n1 to n4, from left to right, in each indicated gRNA arm **B.** Log<sub>2</sub> iBAQ values sourced from the indicated label-free CRISPR-APEX2 publications were plotted as in (A). gRNA labels are omitted

*Limma* analysis was applied as previously described using a paired design for calculation of protein fold enrichment between AR g2 and NT arms. Our optimised n = 4 experiment resulted in significantly greater protein enrichment than initial data (Figure 5.20).

Additionally, enrichment patterns between gRNAs were as expected. In NT-transfected samples a greater randomness of protein biotinylation would be predicted, whereas using AR g2 should result in more consistent protein intensities. Therefore the former may result in log2 fold enrichments that fail to achieve adequate repeatability for statistical significance, whereas the latter would provide such significance (Figure 5.20, top panel).

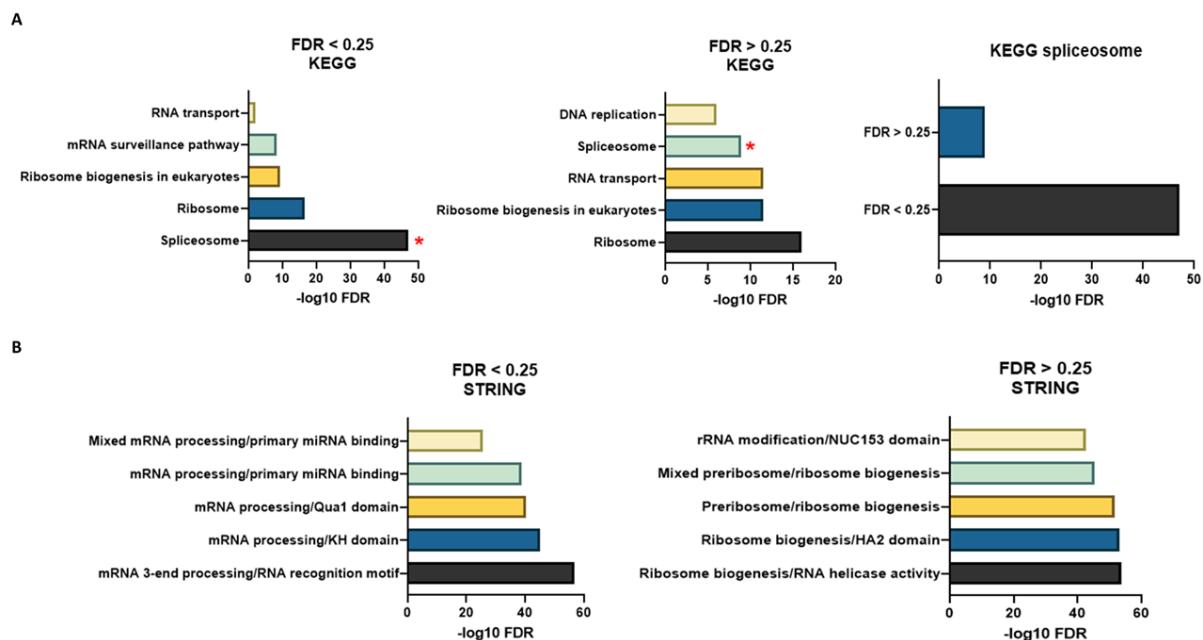


**Figure 5.20 - Experimental optimisation of proteomics workflows significantly improved dataset power**

Protein enrichment was calculated between AR g2 and NT gRNA arms using iBAQ values with the *limma* package as in Figure 5.17A. Enrichment cutoffs for visualisation in 'Optimised n = 4' are *limma* p-value (unadjusted)  $< 0.05$  and linear fold enrichment  $> 1.5$  ( $\log_2 0.585$ ). Cutoffs for initial n = 4 are as described in Figure 5.17A.

Further validation of this approach was sought via functional analysis of protein enrichment results. Protein enrichment tables were split on the basis of significance at a *limma*-adjusted p-value (FDR) of 0.25 (*limma* utilises the Benjamini-Hochberg method of FDR adjustment (Benjamini and Hochberg, 1995)). An FDR of 0.25 has historically been adopted as a threshold

for exploratory functional enrichment analysis and was applied by similar dCas13-APEX2 experiments to that presented here (Subramanian *et al.*, 2005; Lin *et al.*, 2021). Separate analysis of the two groups, comprising proteins with enrichment FDR  $</> 0.25$ , using STRING revealed a substantially greater representation of proteins related to the spliceosome in the  $< 0.25$  group (using Kyoto Encyclopaedia of Genes and Genomes (KEGG) classification) (Figure 5.21A). Additionally, STRING identified significant functional clustering related to mRNA processing within proteins enriched at an FDR  $< 0.25$ , including 3'-end processing as would be predicted given CE3 status as a 3' terminal exon (Figure 5.21B). These results reflect the expected functional enrichment of proteins seen if dCasRx-APEX2 was targeted to an actively spliced 3' terminal exon.



**Figure 5.21 - STRING analysis of proteins enriched by AR g2 demonstrates significant enrichment for the expected biological processes**

**A.** KEGG analysis, performed using STRING, was applied to proteins in two separate groups representing those enriched by AR g2 at a *limma* FDR  $< 0.25$  and  $> 0.25$ . The top 5 most significant KEGG terms are displayed for each group, ranked by  $-\log_{10}$  STRING FDR. KEGG spliceosome is highlighted by a red asterisk. KEGG spliceosome  $-\log_{10}$  STRING FDR for each group is also plotted separately **B**. The same analysis was performed as in (A), using STRING network functional classification of proteins. The top 5 most enriched STRING network category terms are displayed for each group, ranked by  $-\log_{10}$  STRING FDR

Having acquired substantial evidence that AR g2 was enriching for the expected protein classes, we leveraged previous publications reporting on CE3 interactors that promote AR-V7 splicing to further authenticate analyses. We filtered *limma* outputs for inclusion of previously

established regulators (sourced from Table 1.2, in addition to other proteins identified within those publications). In doing so a skewed trend towards positive enrichment with AR g2 was observed, although not all proteins achieved the required consistency for statistical significance (Figure 5.22A).

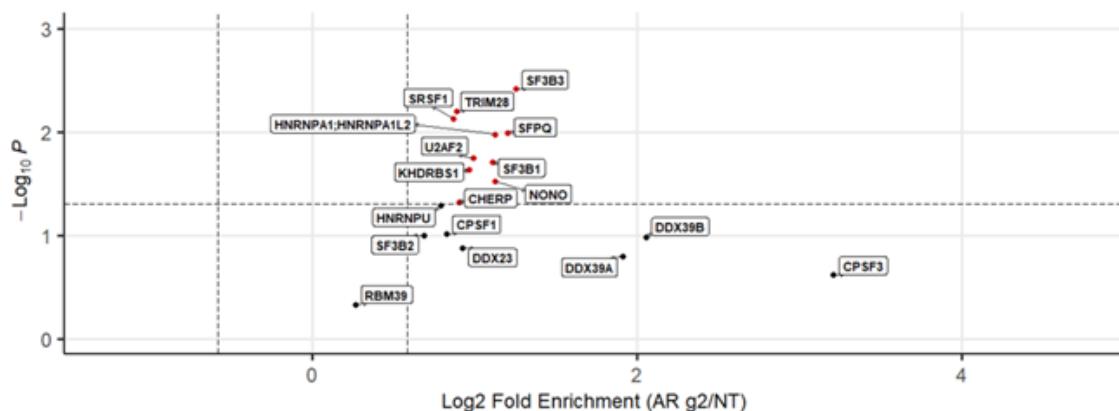
To quantitatively validate whether published AR-V7 splicing regulators trended towards enrichment in AR g2-transfected samples, rotation gene set testing (ROAST) was implemented via *limma* (Wu *et al.*, 2010; Ritchie *et al.*, 2015). The selection of gene set testing employed here was considered, as other popular statistical tests such as gene set enrichment analysis (GSEA) may have been applicable (Subramanian *et al.*, 2005). However, GSEA was deemed inappropriate for these purposes, as it is designed to accommodate large, genome-wide datasets of > 10,000 genes. Furthermore, GSEA requires a pre-ranked list as input. For our smaller protein enrichment dataset, it was preferable to perform gene set testing of enriched proteins (or the genes encoding them) based on statistical significance, which factors in consistency of enrichment. P-values are not directional thus unable to be used for GSEA ranking, as they do not indicate in which sample a protein is enriched, whilst ranking by raw fold enrichment will not select for the proteins of most interest, namely those that were repeatedly enriched across replicates and thus at a greater significance, representing the strongest candidate CE3 splicing regulators (Figure 5.20, top panel). ROAST utilises information relevant to this purpose such as correlations between replicates, therefore factoring in directionality and statistical significance, and is shown to be robust in smaller datasets (Wu *et al.*, 2010). Additionally, it is directly applicable to datasets analysed using linear statistical models, as performed here using *limma* (Ritchie *et al.*, 2015). Given that these characteristics are well suited to our data, and that the ROAST algorithm is implementable within the same *limma* software used in all other protein enrichment analyses, it represented an appropriate choice of test. ROAST substantiated our observations by proving a statistically significant directionality, showing that published AR-V7 regulators have an asymmetrical distribution towards more significant FDRs in our dataset (Figure 5.22B). This provided greater confidence that AR g2 is capturing proteins as expected given the enrichment trends seen amongst these documented proteins.

Optimisation of CE3-targeted dCasRx-APEX2 proximity biotinylation, streptavidin enrichment and LC-MS/MS, allied to application of suitable quantitative analyses, has demonstrated

significant enrichment of splicing factors. Furthermore, pre-existing knowledge of published interactors served as a vital positive control to confirm workflow effectiveness. Satisfied with validation of results, proteomics experiments provided a powerful dataset representing a rich source of hitherto unidentified CE3-proximal proteins in the CRPC cell line CWR22Rv1. Efforts now turn to closer examination of enriched proteins and ultimately application of this dataset to uncover novel regulators dictating AR-V7 generation.

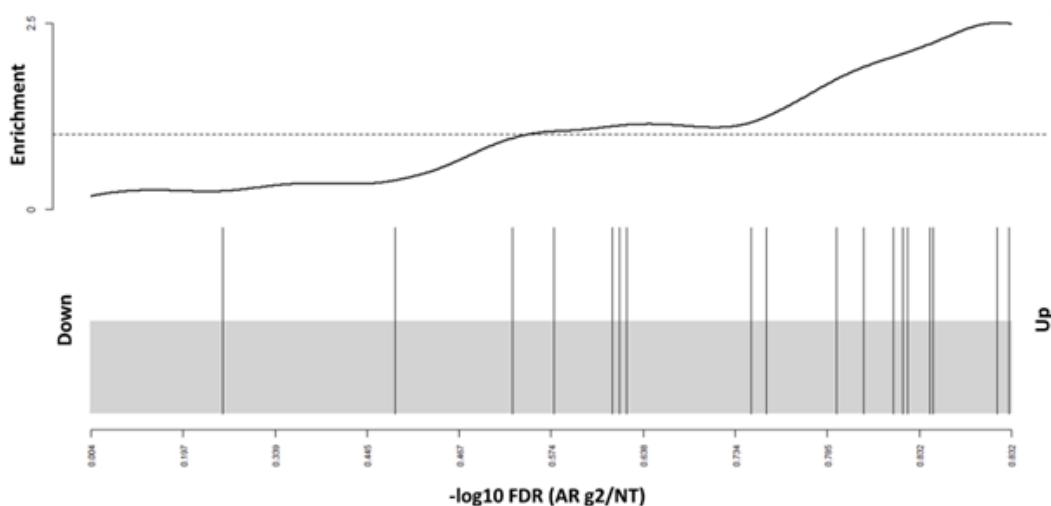
**A**

### Published AR-V7 regulators



**B**

### ROAST Enrichment FDR = 0.035



**Figure 5.22 - Previously identified regulators of AR-V7 generation are enriched by our dCasRx-APEX2 workflow**

**A.** Protein enrichment results were filtered for previously published AR-V7 regulators as described in text. Volcano plot enrichment cutoffs are as used in Figure 5.20 'optimised n = 4' **B.** ROAST was performed using the list of proteins in (A) as input. FDR represents ROAST-calculated significance of statistical enrichment for this list. Barcode plot represents ranking of  $-\log_{10}$  FDR values for this list amongst all proteins from *limma* analysis

## 5.8 Discussion

The primary objective of this project, identification of novel AR-V7 splicing regulators, hinged on generation of a high-quality proteomics dataset. The processes undertaken in this chapter have achieved this through meticulous development of CE3-targeted proximity biotinylation in CWR22Rv1(dCasRx-APEX2), requiring: i) creation of transgenic doxycycline-inducible CWR22Rv1; ii) synthetic CasRx gRNA design and validation for targeting of CE3 mRNA; iii) optimisation of dCasRx-APEX2 proximity biotinylation and streptavidin enrichment of proteins; and iv) application of RNA biotin-labelling with BAn to confirm dCasRx-APEX2 presence at CE3. Collectively, these enabled powerful enrichment of proteins using CE3-targeting gRNA AR g2, statistical significance of which was quantified using the linear models available in the R Bioconductor package *limma*.

Functional interrogation of the most effectively enriched proteins using STRING demonstrated an overrepresentation of proteins involved in mRNA splicing, processing and 3'-end maturation, all processes that would be predicted at an actively spliced 3' terminal exon. In addition, we were able to utilise the current knowledgebase of published AR-V7 splicing regulators as a positive control and secondary layer of validation. By doing so, we confirmed an association using ROAST whereby previously identified factors were skewed towards statistical enrichment in our dataset.

Additional proximity biotinylation technologies have emerged since the advent of BioID and APEX2. One such example, TurboID, was developed by directed evolution yeast display screening of biotin ligase BirA, using the same approach as for development of APEX2 to select for mutants with the greatest biotinylation capability (Branon *et al.*, 2018; Lam *et al.*, 2015). TurboID, like BioID, presents an advantage over APEX2 methods in that it uses soluble exogenous biotin as opposed to the more toxic combination of BP/H<sub>2</sub>O<sub>2</sub> which precludes use of the latter in *in vivo* or organoid systems (Che and Khavari, 2017). Moreover, biotin exhibits far greater membrane permeability than BP, further enabling its potential applications in more complex systems (Mair and Bergmann, 2022; Murata *et al.*, 2021). By progressively reducing biotin incubation times with BirA mutants, TurboID was developed as a ligase capable of biotinylation after 10-minute biotin treatments (Branon *et al.*, 2018; Cho *et al.*, 2020). To date, only one recent publication has exploited CRISPR-TurboID approaches in the form of a dCas9-TurboID fusion (Yheskel, Sidoli and Secombe, 2023). However, in this instance

it was used without a gRNA as a background control for KDM5-TurboID protein enrichment rather than being targeted to a specific genomic locus.

More novel still are ultraID and AirID. As the smallest developed enzyme to date for proximity biotinylation at just 19.7 kDa, ultraID is even more favourable for fusion constructs yet still retains biotinylation catalysis comparable to TurboID, biotinyling proteins within a 10-minute exogenous biotin incubation (Kubitz, Bitsch and Zhao *et al.*, 2022). AirID, though it performs better than first and second-generation BioID enzymes, does not have the same biotinyling potency as TurboID. However, it has been shown to mediate more specific proximity biotinylation with fewer off-targets. Additionally, unlike TurboID it has no detectable toxicity as a fusion protein, even when expressed constitutively for several days (Kido *et al.*, 2020; Branion *et al.*, 2018).

Theoretically, the various BioID-derived methods discussed fulfil criteria appropriate for our study of alternative splicing regulators. APEX2 still holds the kinetic advantage of a brief 1 - 2 minute labelling reaction upon H<sub>2</sub>O<sub>2</sub> addition, providing greater temporal resolution. Another significant advantage is its ability to catalyse biotin-radical formation with a range of aromatic substrates, thereby expanding the range of biomolecules amenable to biotinylation, whereas methods that have evolved from BioID solely label lysine residues of proteins (Sears, May and Roux, 2019). Most notably for our experiments, BAn is appropriate for RNA labelling, whilst biotin-naphthylamine also enables effective DNA biotinylation (Zhou *et al.*, 2019). This capability proved crucial to project success by enabling us to confirm dCasRx-APEX2 localisation to our mRNA of interest by BAn/H<sub>2</sub>O<sub>2</sub> RNA biotinylation.

The repertoire of available enzymes for proximity biotinylation has thus expanded in recent years, each having their own merits. Prudence is advised when selecting an enzyme, as choices are best judged on the specific experimental goals. It remains to be seen whether more novel enzymes will be used as CRISPR-based fusions, as has been successfully achieved by numerous groups with APEX2. Although APEX2 functioned effectively for this study, other available enzymes do hold advantages as discussed. They open the possibility of proximity biotinylation assays in a wider range of models such as CRPC organoids or *in vivo* systems, which may identify an additional cohort of AR-V splicing regulators. Performing an analogous experiment using the same gRNA AR g2 in CWR22Rv1 with, for example, a dCasRx-TurboID fusion to compare enriched protein populations would certainly be edifying.

We opted for label-free mass spectrometry for its ability to provide deeper proteome coverage, whilst also performing a high number of biological replicates to help alleviate the variability seen in such workflows (Li *et al.*, 2012; Rozanova *et al.*, 2021; Stepath *et al.*, 2020). Significant issues regarding high MV incidence were solved through quantification of trypsin-digested peptides prior to LC-MS/MS, ensuring an optimal and equal amount of peptide was loaded for each run. Label-free proteomics necessitates separate LC-MS/MS runs for every sample, which is a major source of variability in using this method. An alternative is to perform multiplexed proteomics, either by metabolic labelling using stable-isotope labelling by amino acids in cell culture (SILAC), or isobaric labelling, commonly performed with tandem mass tags (TMT) or isobaric tags for relative and absolute quantification (iTRAQ).

SILAC is performed by culturing cells through multiple passages with addition of amino acids that contain heavy isotopes such as  $^{13}\text{C}$ , which behave identically to conventional amino acids when incorporated into proteins. Therefore, by incubating different experimental arms with either regular or heavy amino acids, LC-MS/MS runs can be multiplexed and the sample from which peptides are derived determined by isotope mass shifts (Chen *et al.*, 2015).

Isobaric labelling introduces a chemical tag to tryptic peptides which undergoes fragmentation at a predictable point during MS2. By introducing different mass tags to separate samples, they can be multiplexed as for SILAC, although when using TMT/iTRAQ tags peptides are of the same mass at the MS1 level and a mass differential is only introduced after collision-induced fragmentation (Pappireddi, Martin and Wühr, 2019). Though these approaches undoubtedly improve quantification precision between independent LC-MS/MS runs (Rozanova *et al.*, 2021), it was decided that label-free methods would be utilised in order to maximise protein identifications, with 4 independent biological replicates being performed to help mitigate variability.

The current range of published CRISPR-APEX2 studies show that any of label-free, metabolic and isobaric labelling methods can be successfully implemented for gRNA-directed proximity proteomics (Gao *et al.*, 2018; Yi *et al.*, 2020; Lin *et al.*, 2021; Qiu *et al.*, 2019; Han *et al.*, 2020; Myers *et al.*, 2018). It has been shown that label-free mass spectrometry worked effectively for this study, although any potential further improvements by multiplexing are hard to ascertain without direct testing. Furthermore, the approach utilised here negated

requirements to purchase and optimise additional labelling reagents or metabolic incorporation protocols, thus simplifying the overall workflow.

Undertaking a similar experimental design using APEX2 fused with the DNA-targeting ability of dCas9, as performed by others (Gao *et al.*, 2018, Myers *et al.*, 2018), would be of great interest. Given that alternative splicing can be influenced by chromatin architecture, histone marks and histone modifying enzymes (Naftelberg *et al.*, 2015; Agirre *et al.*, 2021), this could identify a distinct class of splicing regulator. For example histone demethylase KDM4B, which has been shown to bind chromatin at the genomic CE3 locus (Duan *et al.*, 2019), was not enriched by this dataset whereas KDM4B-associated factors TRIM28, SF3B3 and U2AF2 were. Our method of nuclear extraction has likely isolated soluble proteins, neglecting to isolate the insoluble fraction that often constitutes chromatin-associated elements (Takat *et al.*, 2009). Therefore, matched workflows that label proteins proximal to DNA/chromatin or RNA at CE3 in a gRNA-directed manner using dCas9 and dCasRx, respectively, may be combined with optimisation of protein isolation to provide complementary datasets. Indeed this is an area of active research in the host group, and colleagues are currently undertaking development of a dCas9-based APEX2 methodology.

In retrospect, significant advantages may have been gained by fusing catalytically active CasRx with APEX2 in the same manner as dCasRx. In order to validate dCasRx-APEX2 RNA-targeting functionality and verify that fusion with APEX2 was not deleterious, it was necessary to develop a BAn/H<sub>2</sub>O<sub>2</sub> RNA-labelling and streptavidin pulldown assay. Though this methodology achieved the desired goals, it is costly, time-consuming, and technically laborious compared to a simpler qPCR or western blot knockdown readout.

Had CasRx been fused with APEX2 from the outset with an identical linker to dCasRx-APEX2, the latter approaches could have been used by making the assumption that if a CasRx-APEX2 fusion could effectively utilise a gRNA for target mRNA degradation, it must be capable of gRNA-mRNA binding and recognition. Therefore, it follows that dCasRx-APEX2, which only differs by four catalytic residues, would similarly bind mRNA in a gRNA-directed manner. This would have created the possibility to screen multiple gRNAs against a target more efficiently, possibly using a higher throughput technique such as FACS. Nevertheless, we designed and validated the necessary gRNA for this project although this idea would warrant testing for future use of this method at other RNA interactomes of interest.

Our attention now turned to harnessing this proteomics dataset for selection and validation of novel AR-V7 splicing regulators, which forms the focus of the subsequent and final results chapter.

# Chapter 6 Selection and validation of novel AR-variant splicing factors

## 6.1 Introduction and rationale

With proximity biotinylation proteomics targeted at CE3 mRNA in CWR22Rv1 successfully validated, protein enrichment data now required application towards achieving the project's primary objective: identification of novel AR-V7 splicing regulators. All work performed up to this point including reagent generation, technical optimisation and proteomic analyses have ultimately had this goal in mind, therefore this next stage represents a crucial proof of concept that this approach is more than a demonstration of novel techniques and can help answer biologically relevant questions. Using a dCasRx-APEX2 approach with gRNA AR g2, we enriched 203 CE3-proximal proteins at  $p < 0.05$  and FDR (adjusted  $p$ -value)  $< 0.25$  (Figure 5.20). Therefore, the next step of analysis will be formulation of a selection strategy from this list of proteins to identify candidates for functional validation.

Careful assessment of the dataset ruled out simply selecting a panel of the most enriched CE3-proximal proteins. Functional analyses of experimentally identified proteins demonstrated significant overrepresentation of splicing processes (Figure 5.21), providing confidence of our targeting to an actively spliced mRNA exonic region. Therefore, we were cognizant of the fact that selecting the most highly enriched factors could merely reflect proteins that are abundantly present at all splicing junctions. The purpose of this project was to select factors that are reasonably specific for CE3 inclusion in AR-V7, in the hope that their inhibition may provide superior selectivity and minimise global alternative splicing perturbations.

Numerous publications have examined AR-V splicing factor inhibition in CRPC as summarised in Table 1.2. However, these have largely explored members of the core spliceosomal machinery or polyadenylation complexes. Given the essential cellular functions of alternative splicing, targeting these core components would be likely to cause prohibitive toxicity as seen in clinical trials of pharmacological splicing inhibition to date, which have failed in solid tumours (Eskens *et al.*, 2013; Hong *et al.*, 2014; NCT00499499 and NCT00459823). Although

the remit of this project will not move beyond preclinical work, we hope to demonstrate an element of specificity for AR-V7 splicing among our chosen factors.

It was therefore decided to leverage a range of publicly available PCa and CRPC clinical datasets to assist target selection for two primary reasons. Firstly, by correlating expression of proteins (or the genes encoding them) with a range of parameters including, for example, AR-V7 levels, activation of AR-V7 target genes, or patient survival, it was hypothesised that this would maximise the chances of selected protein(s) being *bona fide* AR-V7 splicing regulators. Such metrics should complement this interactome dataset well to collectively provide information on splicing factors that are: i) interacting with CE3 mRNA in a CRPC model expressing high levels of AR-Vs; and ii) associated with AR-V7 splicing and activity in patient cohorts. This combined information is significantly more powerful than the latter alone, as correlation with AR-V7 expression signatures does not necessarily reflect causation. However, linking correlative data with our experimental validation of CE3 interaction would strongly indicate a direct role in AR-V7 generation.

The second reason to utilise such data is in an effort to raise translational relevance. The approach used here has been designed, optimised, and applied in a single model of CRPC. Although future work could undoubtedly include expansion of techniques to other PCa lines that express AR-Vs, at present only CWR22Rv1 have been used. Whilst this has provided an excellent proof of concept for this biomolecular technique, there will naturally be concerns regarding whether identified factors are specific to this cellular and genetic background only. Therefore, use of patient-derived datasets will be invaluable for any future utility of validated proteins as predictive biomarkers of AR-Vs or therapeutic targets to block AR-V7 splicing.

Splicing factor aberrations in solid tumours are largely gene amplification or overexpression events as opposed to the high frequency of splicing factor mutations seen in haematological malignancies (Anczuków and Krainer, 2016). Hence, use of clinical cohort gene expression data represents an appropriate means of selection from our AR g2-enriched protein list. Clinical proteomics data would have been of additional value, particularly phosphoproteomics as splicing regulators, including the SR protein family, are regulated by phosphorylation status (Jeong, 2017). However, there is a marked paucity of patient-derived proteomics data. Although public repositories including the Clinical Proteomic Tumor Analysis Consortium (CPTAC) are available, their depth is limited, and currently the wealth of available

data using genomics and transcriptomics far exceeds that of proteomics (Kwon *et al.*, 2021; Seydel, 2022). The Human Proteome Project (HPP), launched in 2010, is attempting to bridge the knowledge gap (Adhikari *et al.*, 2020). However considering the range of possible post-translational modifications and lack of nucleic acid-based signal amplification techniques such as PCR, the technical complexity of large-scale proteomics is readily apparent. Therefore gene expression analyses will form the mainstay of our work.

Discussion of specific datasets and how they have been used will follow in subsequent results sections. Ultimately, the aim is to select protein candidates for subsequent manipulation *in vitro*, thereby experimentally confirming their involvement in AR-V splicing. It is likely that these experiments will result in a lead protein showing the most encouraging data, for which more extensive functional investigation will be performed. This validation will form the final stages of this PhD project as proof that dCasRx-APEX2 technology can successfully unveil novel splicing regulators, in this instance being applied to AR-V7 in CRPC.

## 6.2 Specific materials and methods

All catalogue numbers are supplied at first reference to use of reagent in the document, after which they are omitted.

### 6.2.1 Analysis of publicly available RNA-Seq clinical cohorts

For interrogation of the SU2C/PCF mCRPC cohort (Abida *et al.*, 2019), the R Bioconductor package *cBioPortalData* (v2.6.1, Ramos *et al.*, 2020) was used to extract relevant study data for analysis as detailed in results sections. Unless specified otherwise, Pearson's correlation coefficient was used for determining correlations between parameters. Specific gene sets used for AR and AR-V7 activity correlations are specified within analysis results. For determining mean correlation across signatures, Fisher z-transformation was applied to correlation coefficients. Custom R scripts were used to analyse parameters for each of the 63 gene panel.

For the TCGA-PRAD dataset, *TCGAbiolinks* (v2.15.3) was used to extract raw gene expression counts. Information on patient V7 status was obtained from cBioPortal (Gao *et al.*, 2013) entries, with presence or absence of AR-V7 defined as in the TCGA-PRAD 2015 analysis (TCGA, 2015). Differential gene expression analysis between AR-V7 positive and negative patients was performed using *edgeR* (v3.32.0) (Robinson, McCarthy and Smyth, 2010). log2 counts-per-million were generated using the *edgeR* trimmed-mean of M values normalisation.

The R software environment was used via RStudio, with versions as described in Section 5.2.12.

### 6.2.2 Analysis of publicly available microarray datasets

Microarray gene expression data was obtained via gene expression omnibus series GSE35988, microarray platform GPL6480 (Grasso *et al.*, 2012). Differential gene expression analysis between tumour types was implemented using *limma* via the GEO2R online tool (Barrett *et al.*, 2012). log2 normalised microarray expression values generated by GEO2R were extracted for visualisation and plotting.

### 6.2.3 Determination of gene essentiality

Data on pan-cellular gene essentiality was obtained from the cancer dependency map (DepMap) portal (Tsherniak *et al.*, 2017), available at <https://depmap.org/portal/>.

#### **6.2.4 Patient survival analysis**

Patient survival in the SU2C/PCF mCRPC cohort was determined by Pearson's correlation coefficient between log2 FPKM expression for the genes specified and overall patient survival, measured in months, which was obtained from cBioPortal (Gao *et al.*, 2013). Patient survival between top and bottom quartiles of gene expression in the TCGA-PRAD dataset was determined by logrank test. All TCGA-PRAD survival analysis and plots were performed using the Tumor online Prognostic analysis Platform (Ouyang *et al.*, 2022), available at <http://www.biostatistics.online/topp/>.

#### **6.2.5 Amino acid multiple alignment between TRA2A and TRA2B**

UniProt amino acid FASTA sequences for TRA2A (UniProt Q13595) and TRA2B (UniProt P62995) were aligned using the constraint-based alignment tool for multiple protein sequences (COBALT) (Papadopoulos and Agarwala, 2007), available at <https://www.ncbi.nlm.nih.gov/tools/cobalt/>

#### **6.2.6 siRNA transfection**

CWR22Rv1 cells were reverse transfected in 6-well plates with siRNA at a 25 nM concentration using 0.2% lipofectamine™ RNAiMAX according to manufacturer instructions. Opti-MEM™ I was used as serum-free media for creation of transfection complexes. 72 hours later, protein or RNA was harvested and RT-qPCR or western blot was performed as described Sections in 3.2 and 3.3.

VCaP cells were seeded in 6-well plates in either 10% FBS or 10% dextran-coated charcoal stripped FBS as specified and grown until 80% confluent. At 80% confluence, siRNA was transfected into cells and enzalutamide was added. Briefly, media was refreshed with the addition of 10 µM enzalutamide where applicable and siRNA was transfected at a 25 nM concentration using 0.2% lipofectamine™ RNAiMAX according to manufacturer instructions. Opti-MEM™ I was used as serum-free media for creation of transfection complexes. 72 hours after enzalutamide treatment/siRNA transfection, protein or RNA was harvested and RT-qPCR or western blot was performed as described in Sections 3.2 and 3.3.

All siRNA sequences used are listed in Table 6.1. SMARTpool siRNA was used at 25 nM final concentration. For combined TRA2A/B depletion, 25 nM of each TRA2 siRNA was used.

siRNA	Sense sequence (5' - 3')
non-targeting (NT) (Sigma)	UUCUCCGAACGUGUCACGU
THRAP3 SMARTpool (Horizon L-019907-00-0005)	THRAP3 SP 1: GGUUAAGCUCCGAGAUGA THRAP3 SP 2: CCGAGUGACUGCUUAUAAA THRAP3 SP 3: CAACAUUAUGUGACCAUUG THRAP3 SP 4: CAAAUGGGAGGGCCUGGUA
SART1 SMARTpool (Horizon L-017283-00-0005)	SART1 SP 1: GCUACAAACCCGACGUUAA SART1 SP 2: CCGAAUACCUACGCCUGA SART1 SP 3: GAACCGAUCGUGAAUAGGG SART1 SP 4: UAAAGACCCUAGGAGAGGA
TRA2A (Thermo s26664)	GGAUCUUCGUGAAGUAUUU
TRA2B (Thermo s12749)	GGAGGAUACAGAUCACGUU

Table 6.1 - siRNA sequences used in knockdown experiments

### 6.2.7 CWR22Rv1 TRA2 knockdown RNA-Seq analysis

CWR22Rv1 were reverse transfected as in Section 6.2.6, using NT, TRA2A, TRA2B or combined TRA2A/B siRNA. 72 hours post-transfection, RNA was extracted using the GenElute™ mammalian total RNA miniprep kit. n = 3 biological replicates were prepared and sent on dry ice to GENEWIZ™ for paired-end, strand-specific RNA-Seq. RNA-Seq libraries were prepared using poly(A) selection and ~25 - 40 million 2 x 150 bp reads were sequenced per sample on an Illumina NovaSeq™ 6000. FASTQ files were quality checked using *FastQC* (v0.11.9, Andrews, 2010) and a quality summary report was compiled with *MultiQC* (v1.9, Ewels *et al.*, 2016). Illumina universal adapter sequences were trimmed from reads using *Cutadapt* (v3.5 Martin, 2011), and successful adapter removal was confirmed by running trimmed FASTQ files through *FastQC* again (v0.11.9).

A genome index was generated using *STAR* (v2.7.0e, Dobin *et al.*, 2013) *genomeGenerate* mode with the GENCODE GRCh38 primary assembly and associated GTF annotation (release v43). SAM files were generated against this index with *STAR* (v2.7.0e), using default settings. *SAMtools* (v1.16.1, Li, Handsaker *et al.*, 2019) was subsequently used for sorting files by genomic coordinate. *SAMtools* was also used for conversion of files to BAM format. A gene counts matrix was created with *featureCounts* (Subread v1.4.0) (Liao *et al.*, 2014), using GENCODE primary assembly GTF annotation (release v43).

For differential gene expression analysis, gene-level counts generated by *featureCounts* were exported. *DESeq2* (v1.34.0, Love, Huber and Anders, 2014) was used with default settings for calculation of differential gene expression between samples, and the *ashr* algorithm for log2 fold change shrinkage (Stephens, 2017) was implemented within *DESeq2*. Genes were sorted by log2 fold change for gene set enrichment analysis using GSEA software (v4.0.3) (Subramanian *et al.*, 2005). Gene expression counts, normalised by *DESeq2*, were extracted for plotting.

For differential exon usage analysis, exon-level counts were generated from sorted BAM files using the *QoRTs* package (v1.3.6, Hartley and Mullikin, 2015) with GENCODE primary assembly GTF annotation (release v43). *JunctionSeq* (v1.2.4, Hartley and Mullikin, 2016) was used with default settings to calculate differential exon usage of genes between siRNA conditions. Exon coverage counts, normalised by *JunctionSeq*, were extracted for plotting.

All Linux software was run on the Rocket high performance computing server at Newcastle University. All R packages were run using R and RStudio as described in Section 6.2.1.

### **6.2.8 Sulforhodamine B (SRB) assay**

In 96-well plates, 2,500 and 5,000 cells per well of CWR22Rv1 and VCaP cells, respectively, were reverse transfected as in Section 6.2.6. Transfections were performed in 90 µl total media volume, with a 25 nM concentration of the indicated siRNAs. After 24 hours, 10 µl DMSO or enzalutamide was added to a concentration of 0.1% DMSO, or 10 µM enzalutamide in 0.1% DMSO, in a final culture volume of 100 µl. At the indicated post-treatment timepoints, cells were fixed using 10% (w/v) ice-cold TCA (Sigma T4885) and stored at 4°C until processing. Once all timepoints were fixed, plates were washed in diH<sub>2</sub>O and air dried. Cells were then stained using 0.4% (w/v) SRB (Sigma S9012) dissolved in 1% acetic acid (v/v) for 30 minutes at room temperature. Plates subsequently underwent two rounds of 5 washes in 1% acetic acid (v/v) and were air dried. Cell-bound dye was solubilised in 100 µl 10 mM Tris-HCl (pH 10.8) by shaking at room temperature for 15 minutes, before 570 nm absorbance readings were taken using a microplate reader (Bio-Rad). All SRB data was normalised to respective day 0 samples, before being scaled within each biological replicate to the NT siRNA/DMSO-treated experimental arm. SRB datapoints were calculated as an average of technical triplicates within each biological replicate.

### **6.2.9 Phosphorodiamidate morpholino and decoy RNA oligomer transfection**

For phosphorodiamidate morpholino oligomers (PMO) experiments, CWR22Rv1 and VCaP cells were reverse transfected in 12-well plates with the indicated PMO at a 10  $\mu$ M concentration using 0.2% lipofectamine<sup>TM</sup> RNAiMAX according to manufacturer instructions. Opti-MEM<sup>TM</sup> I was used as serum-free media for creation of transfection complexes. 48 hours later, transfection was confirmed by live-cell imaging of fluorescein-conjugated oligomers with a Nikon<sup>TM</sup> TE2000 fluorescence microscope, before RNA was extracted for RT-qPCR as detailed in Section 3.3. PMO sequences used are: CE3 block, 5' - AGTCAGCCTTCTTCAGGGTCTGGT - 3'; control, 5' - CCTCTTACCTCAGTTACAATTATA - 3'.

For decoy RNA experiments, CWR22Rv1 cells were reverse transfected in 12-well plates with the indicated decoy RNA oligomers at a 5  $\mu$ M concentration using 0.2% lipofectamine<sup>TM</sup> RNAiMAX according to manufacturer instructions. Opti-MEM<sup>TM</sup> I was used as serum-free media for creation of transfection complexes. 48 hours later, RNA was extracted for RT-qPCR as detailed in Section 3.3. Decoy RNA sequences used are: TRA2 decoy, 5' - (AGAA)5<sub>n</sub> - 3', control decoy, 5' - (GCAAUCC)3<sub>n</sub> - 3'.

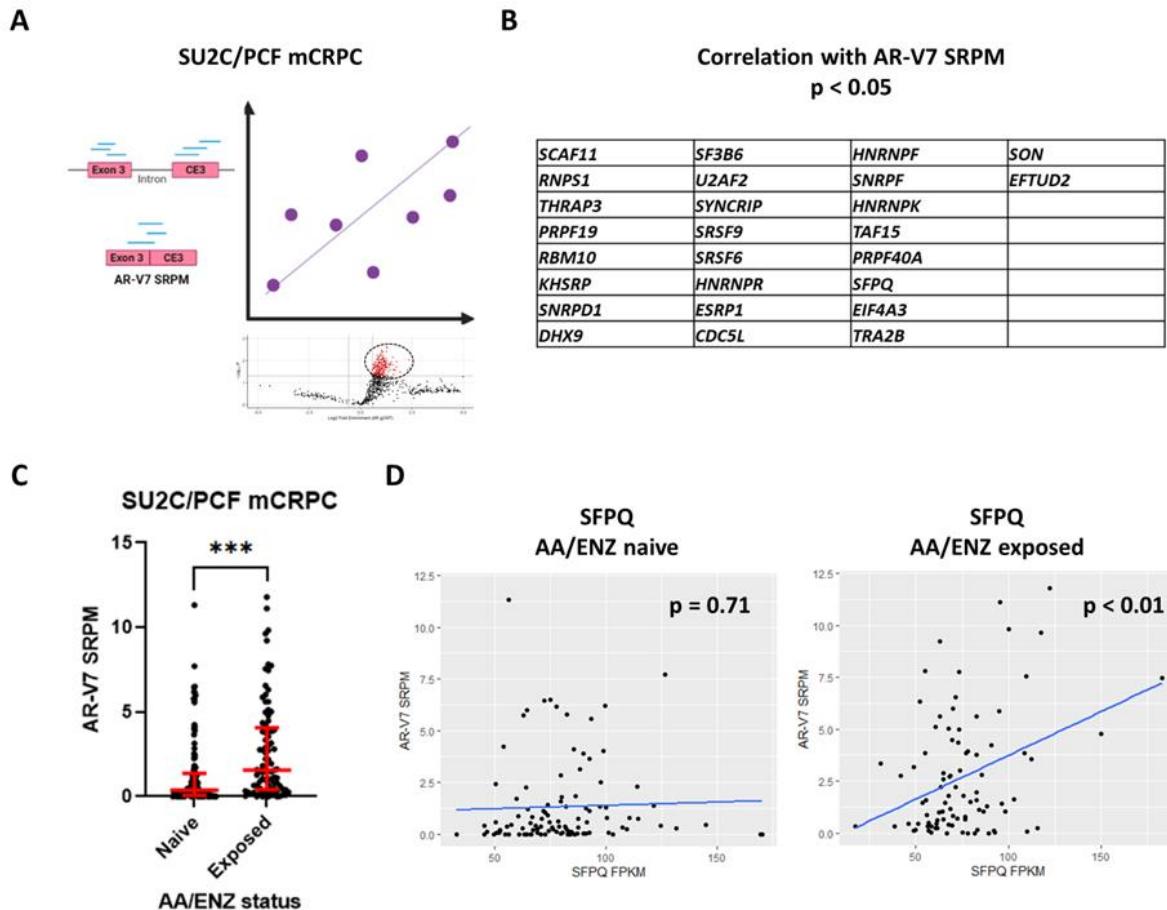
### 6.3 Publicly available prostate cancer data facilitates splicing factor selection

For selection of protein candidates from the CE3 interactome, a range of publicly available datasets were employed. Before analysing these, the *limma*-derived protein enrichments were filtered, as despite seeing significant overrepresentation of splicing processes dCasRx-APEX2 also captured ribosomal proteins (Figure 5.21A). Furthermore, protein subunits of RNA polymerase II were enriched, indicating that this dataset selected for transcriptional processes to some degree. Although splicing occurs cotranscriptionally and is influenced by transcriptional rates (Merkhofer, Hu and Johnson, 2014), this project's primary focus was on splicing factor identification. To generate a final protein list for further analysis, *limma* outputs were filtered to solely include proteins enriched at  $p < 0.05$  and FDR (adjusted p-value)  $< 0.25$  that are also functionally involved in any category of splicing activity based on annotation sourced from QuickGO (Binns *et al.*, 2009), resulting in 63 protein candidates for subsequent interrogation.

A large cohort of 429 mCRPC patients from the international Stand Up to Cancer/Prostate Cancer Foundation (SU2C/PCF) consortium was selected, from which tumour RNA-Seq gene expression data is publicly available for more than 200 samples (Abida *et al.*, 2019). All data for this study was accessed for analysis using the R Bioconductor package *cBioPortalData* (Ramos *et al.*, 2020). Additionally, this dataset contained two forms of RNA-Seq expression library, derived from exon capture technology and poly(A) selection. We chose the former ( $n = 208$  patients) due to its greater tolerance for RNA degradation, which is more likely to be a concern for patient tumour-derived RNA extractions (Cieslik *et al.*, 2015).

First, fragments per kilobase million (FPKM) RNA-Seq expression values from the final 63 gene panel were correlated with AR-V7 spliced reads per million (SRPM), whereby a spliced read is classed as an RNA-Seq read mapping across the exon 3/CE3 splice junction (Figure 6.1A) (Abida *et al.*, 2019). Correlation analysis revealed 26 genes to have an association with AR-V7 at  $p < 0.05$  (Figure 6.1B). As AR-V7 levels have been shown to further increase upon treatment with enzalutamide (ENZ) and/or abiraterone acetate (AA) both in patients and xenograft models (Sharp *et al.*, 2019; Sowalsky *et al.*, 2022), which was also observed in this dataset (Figure 6.1C), it was of interest to separate this  $n = 208$  cohort into groups of AA/ENZ naïve and treated patients. Separate analysis of naïve/treated groups demonstrated striking differences in the strength of correlation upon treatment for certain genes including

published AR-V7 regulator SFPQ (Takayama *et al.*, 2017) (Figure 6.1D). Analysis of expression changes in response to these agents will further aid target selection from CE3-interacting proteins identified by dCasRx-APEX2 workflows.



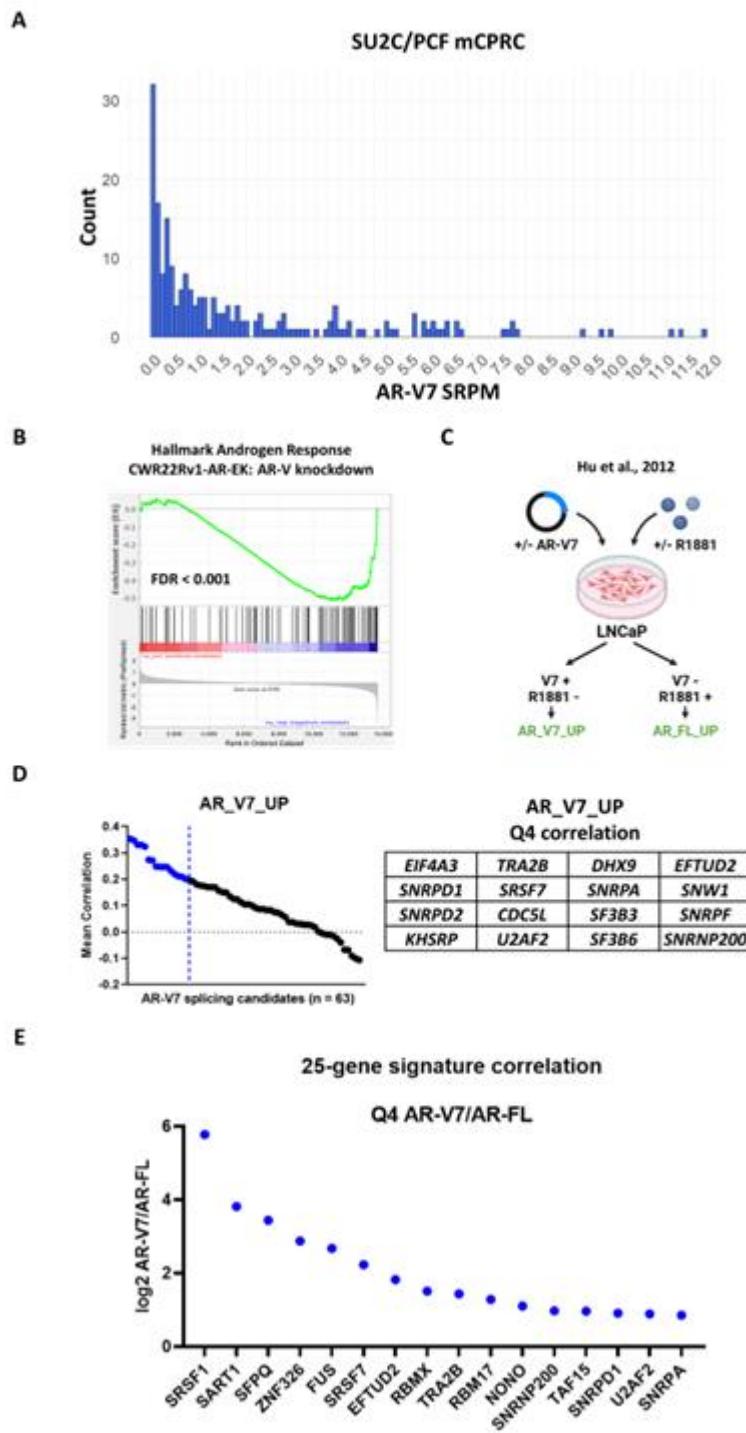
**Figure 6.1 - Correlation of enriched proteins' gene expression with AR-V7 splicing in an mCRPC patient cohort**

**A.** Proteins enriched using dCasRx-APEX2 at  $p < 0.05$  and  $FDR < 0.25$ , that were also associated with splicing gene ontology, were selected for correlation analysis using an  $n = 208$  mCRPC patient cohort (Abida *et al.*, 2019). Genes encoding this list of 63 proteins were examined for correlation between gene expression (FPKM) and AR-V7 splicing (SRPM), the latter being classed as the number of RNA-Seq mapped reads spanning the exon 3/CE3 boundary per million reads **B.** Expression of 26 genes correlated with AR-V7 SRPM at  $p < 0.05$  using Pearson's correlation coefficient **C.** Mean AR-V7 SRPM was compared between AA/ENZ naïve and exposed patients. AA/ENZ naïve comprised  $n = 106$  patients, AA/ENZ exposed comprised  $n = 89$  patients. Individual patient AR-V7 SRPM values are plotted, significance was determined by unpaired t-test ( $*** = p < 0.001$ ) **D.** Numerous genes exhibit significantly greater FPKM expression correlation with AR-V7 SRPM in AA/ENZ-treated patients than AA/ENZ-naïve. Correlation was determined using Pearson's correlation coefficient. *SFPQ* is displayed as an example

Examination of AR-V7 SRPM across the analysed cohort demonstrated numerous patients had SRPM values of either 0 ( $n = 24$ ) or  $< 0.5$  ( $n = 61$ ) (Figure 6.2A). Given that the AR-V7 SRPM measure relies on reads spanning the exon 3/CE3 junction, it was theorised that a secondary AR-V7 activity readout may act as a surrogate marker for AR-V7 generation where SRPM is at the limits of detection. To this end, a range of gene sets and signatures representing AR-V7 activity were leveraged. One of these was a 2012 study that utilised ectopic expression of AR-V7 in LNCaP PCa cells, which are negative for AR-V7, in the presence or absence of androgen to derive a pair of distinct 25-gene signatures, AR\_FL\_UP and AR\_V7\_UP, that delineate transactivation by AR-FL and AR-V7, respectively (Hu *et al.*, 2012).

Precise categorisation of AR-FL/AR-V7 specific target genes is somewhat challenging, as the cistromes of AR species overlap considerably (Cato *et al.*, 2019). Furthermore, the host lab has generated a CRISPR-engineered CRPC cell line, CWR22Rv1-AR-EK, that expresses AR-Vs in the complete absence of AR-FL (Kounatidou *et al.*, 2019). Knockdown of AR-Vs in this cell line results in a highly significant negative gene set enrichment analysis (GSEA) result for the 'Hallmark Androgen Response' (Figure 6.2B) (Liberzon *et al.*, 2015; Subramanian *et al.*, 2005). This is a well-defined gene set representing the AR transcriptional response to androgen treatment, further evidencing the intersection of AR-FL/AR-V transactivation capability.

Therefore, expression of the 25-gene AR-FL and AR-V7 signatures described (Figure 6.2C) was correlated with the final list of genes encoding 63 splicing factors enriched by dCasRx-APEX2. Two metrics were analysed. First, genes were ranked on the basis of mean correlation with AR\_V7\_UP genes in  $n = 208$  mCRPC cases (Figure 6.2D). Given the aforementioned overlap in AR-FL/AR-V7 targets, an assessment of preferential correlation with AR-V7 transactivation over AR-FL would also be insightful. Consequently, the same correlation was performed using AR\_FL\_UP genes to provide a readout of AR-FL activity, and the relative correlation with AR\_V7\_UP over AR\_FL\_UP was determined. Genes in the top quartile (Q4) of AR\_V7\_UP over AR\_FL\_UP relative correlation are plotted (Figure 6.2E).



**Figure 6.2 (previous page) - Target gene expression data may provide an auxiliary readout of AR-V7 generation**

**A.** Numerous RNA-Seq samples in the analysed SU2C/PCF mCRPC cohort had either 0 (n = 24) or < 0.5 (n = 61) AR-V7 SRPM **B.** GSEA of the Hallmark Androgen Response from an RNA-Seq experiment in which AR-Vs are depleted in CWR22Rv1-AR-EK, a CRPC cell line that expresses AR-Vs only (Kounatidou *et al.*, 2019) **C.** Experimental summary of a study that used combinations of AR-V7 plasmid transfection and treatment with synthetic androgen R1881 in LNCaP PCa cells, which are AR-V7 negative, to derive AR-FL and AR-V7 25-gene transactivation signatures (Hu *et al.*, 2012) (Figure created using biorender.com) **D.** Mean correlation analysis was performed for each of the 63-gene list with the AR\_V7\_UP 25-gene signature (Q4 - top quartile marked by blue dotted line and points). Pearson's correlation coefficient was used. Fisher z-transformation was applied for calculation of mean correlation across the 25-gene signature. Q4 correlating genes are displayed in the table **E.**

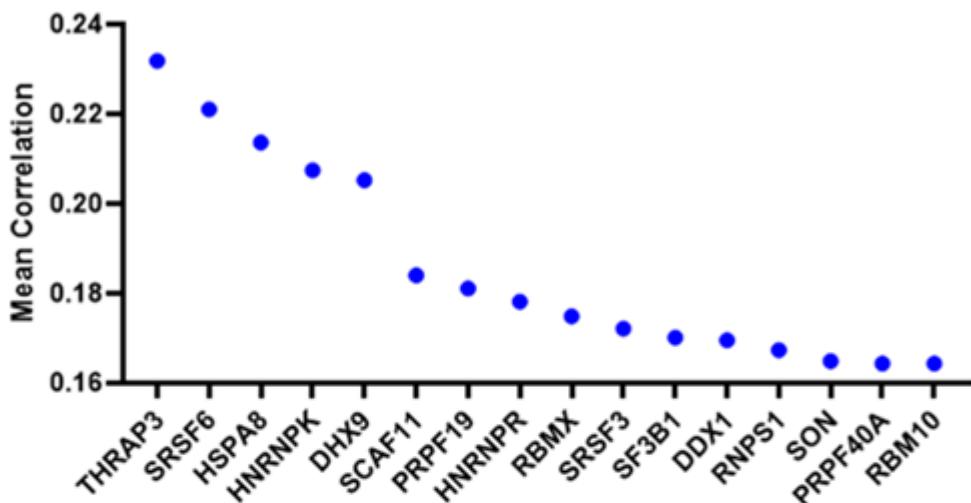
The same correlation analysis was performed as in (D), using the AR\_FL\_UP signature. From this, the relative correlation with AR-V7 over AR-FL target genes was derived. Log2 AR-V7/AR-FL correlation is plotted for the top quartile (Q4) of genes.

To further probe correlations, two additional gene signatures associated with AR-V7 levels were sourced from the literature. Both studies analysed tumours for gene expression patterns, deriving 59-gene (Sharp *et al.*, 2019) and 41-gene (Cai *et al.*, 2018) signatures. The former validated an association between this 59-gene signature and nuclear AR-V7 protein expression in CRPC (Sharp *et al.*, 2019), whilst the latter study performed specific knockdown of AR-V7 in CWR22Rv1 before going on to demonstrate that of the transcriptome alterations by knockdown, a signature of 41 genes correlated with AR-V7 positivity in patient samples (Cai *et al.*, 2018).

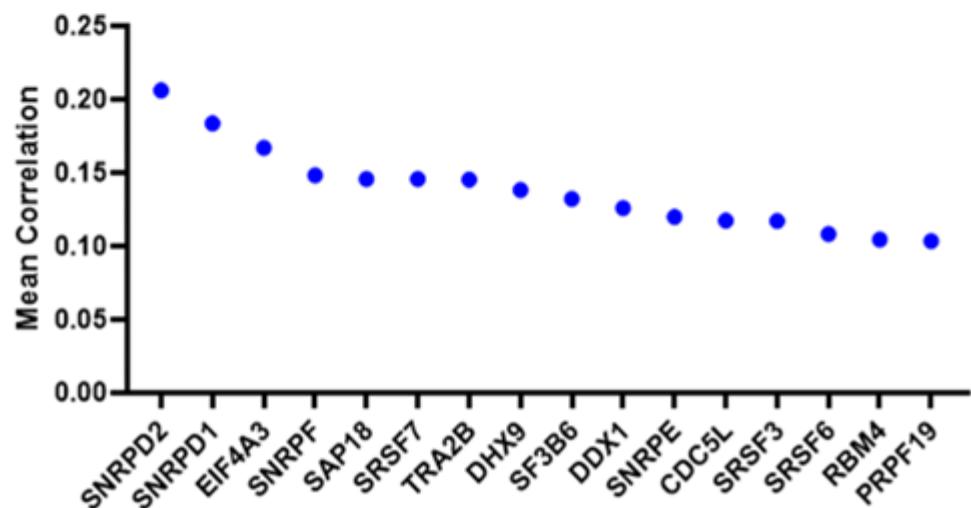
Correlation analysis was performed again, this time against each of the above signatures (Figure 6.3A, B). Mean Pearson's coefficients were not as great as those seen for AR\_V7\_UP (Figure 6.2D), although this would largely be expected given that AR\_V7\_UP was defined in a tightly controlled *in vitro* model of PCa as opposed to large CRPC tumour cohorts that are likely to exhibit significantly greater heterogeneity. Nevertheless, an awareness of how genes that encode the 63-protein list of enriched splicing regulators correlate with these clinically confirmed signatures will aid target selection.

**A**

Sharp et al. 2019  
59-gene AR-V7 signature

**B**

Cai et al. 2018  
41-gene AR-V7 signature



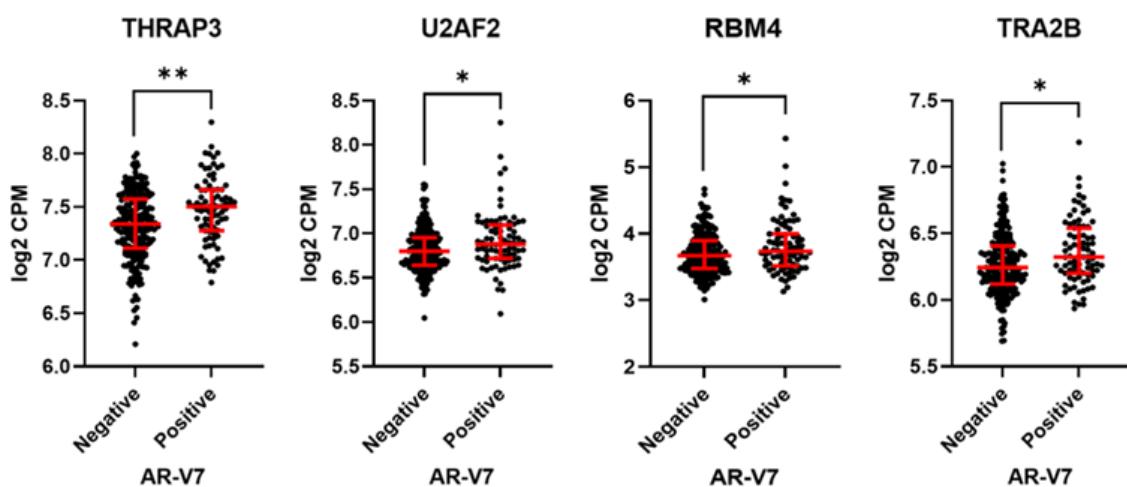
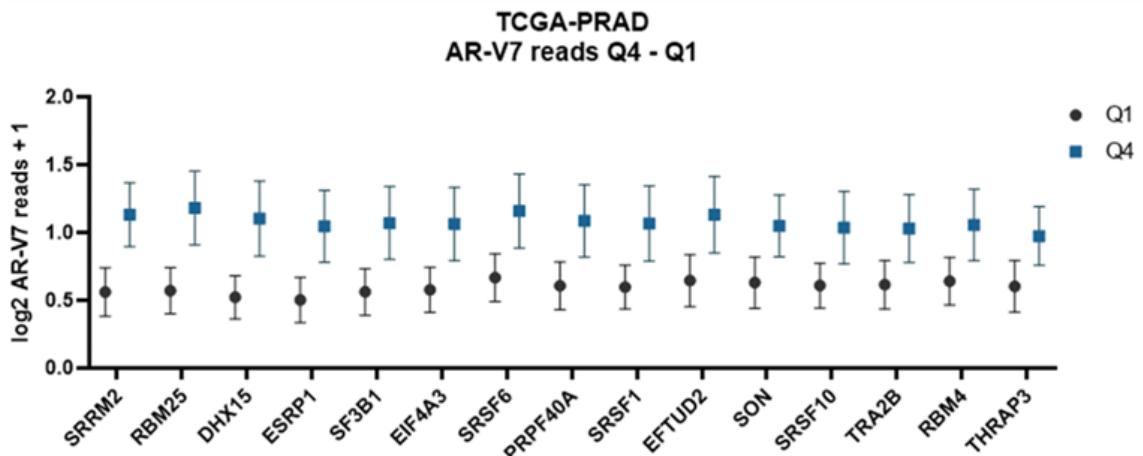
**Figure 6.3 - Correlation of enriched protein genes with clinically identified AR-V7 signatures provides additional evidence for protein selection**

Correlation analysis of genes encoding the 63-protein list of splicing regulators enriched by dCasRx-APEX2 was performed as in Section 6.2, in this instance against two AR-V7 target gene signatures (A - Sharp et al., 2019, B - Cai et al., 2018) confirmed to associate with AR-V7 expression in CRPC tumour cohorts. Mean Pearson's coefficient is plotted for the top quartile of correlated genes

In addition to this  $n = 208$  mCRPC cohort, *TCGAbiolinks* was used (Colaprico *et al.*, 2016) to access RNA-Seq count data derived from The Cancer Genome Atlas (TCGA) prostate adenocarcinoma cohort (TCGA-PRAD); specifically a 2015 analysis of 320 patients that included sample information on AR-V7 mapped reads (TCGA, 2015). It could be argued that analysis of AR-V7 in a primary prostate cancer cohort such as this is less relevant to splicing factor selection, given the lower AR-V7 frequency observed compared to the SU2C/PCF mCRPC dataset and also observed by others (Sharp *et al.*, 2019). However, the fundamental mechanisms of CE3 sequence recognition and binding by splicing factors and, ultimately, AR-V7 generation, are likely to be similar between primary PCa and CRPC as there are no observed changes to the CE3 sequence in *AR* pre-mRNA between these disease states. Rather, it is probable that changes in AR-V7 frequency between primary PCa and CRPC are due to alterations in splicing factor expression levels. Therefore, it was decided that expression analysis within TCGA-PRAD data in addition to the SU2C/PCF cohort would still be informative.

The Data from the TCGA-PRAD cohort was used to split patient samples into two groups based on a binary classification of AR-V7 presence ( $n = 83$ ) or absence ( $n = 237$ ), in which the authors defined 'presence' as  $\geq 2$  total reads spanning the junction. Transcriptome-wide differential gene expression analysis (DGEA) between groups was subsequently performed using the *edgeR* Bioconductor library (Robinson, McCarthy and Smyth, 2010), and results were filtered for genes in the panel of 63 that also exhibited upregulation in AR-V7 positive patients at an FDR  $< 0.05$ . This returned a group of just four genes: *THRAP3*, *U2AF2*, *RBM4* and *TRA2B* (Figure 6.4A).

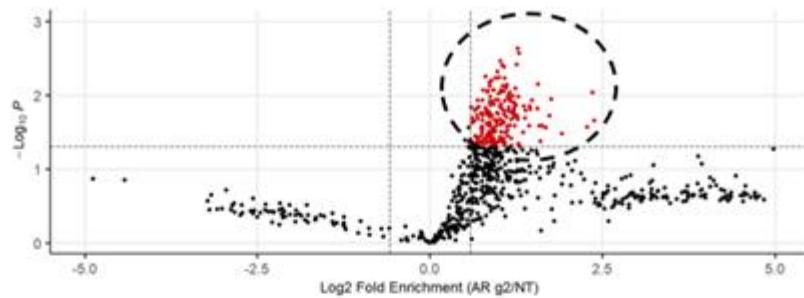
As an alternative measure of association, patients were split into the top and bottom quartiles of expression for each of the 63-gene list before assessing differences in AR-V7 read numbers between bottom (Q1) and top (Q4) quartile groups. This analysis demonstrated that when split by expression quartiles, 15 genes on our panel exhibit significantly greater ( $p < 0.05$ )  $\log_2$  AR-V7 reads in Q4 vs Q1 samples (Figure 6.4B). Therefore, analysis of this TCGA-PRAD cohort provides additional information that supplements prior examination of an  $n = 208$  mCRPC patient dataset (Abida *et al.*, 2019). Collectively, analysis of these datasets provided comprehensive information helping inform selection of splicing factors from the panel of 63 for functional validation (Figure 6.5).

**A****B**

**Figure 6.4 - A panel of genes encoding dCasRx-APEX2-enriched splicing factors are associated with AR-V7 expression in TCGA-PRAD samples**

A. Patients from the TCGA-PRAD cohort for which AR-V7 spliced read information was available ( $n = 320$ ), were split into two groups denoting absence (negative,  $n = 237$ ) or presence (positive,  $n = 83$ ) of AR-exon 3/CE3-spanning RNA-Seq alignments. Transcriptome-wide differential gene expression analysis was performed using *edgeR* and results were filtered for genes differentially expressed at an FDR  $< 0.05$  that were also in the 63-gene panel encoding AR g2-enriched splicing factors. Just 4 genes: *THRAP3*, *U2AF2*, *RBM4* and *TRA2B* met these criteria. Log<sub>2</sub> *edgeR*-normalised counts-per-million (CPM) are plotted. Asterisks denote *edgeR* analysis FDR significance (\* = FDR  $< 0.05$ , \*\* = FDR  $< 0.01$ ). Individual data points for each sample are plotted in black, red denotes the median value and 25<sup>th</sup>/75<sup>th</sup> percentiles. B. In the same TCGA-PRAD dataset, patient samples were split into two groups comprising the upper (Q4) and lower (Q1) quartiles of *edgeR* normalised log<sub>2</sub> CPM values. This was performed for each gene in the panel of 63. For each of these 63 divisions, log<sub>2</sub> AR-V7 reads were compared between Q4 and Q1 by unpaired t-test. 15 genes, plotted above, had significantly greater AR-V7 reads in Q4 at a  $p < 0.05$ . Mean log<sub>2</sub>(AR-V7 reads + 1) are plotted as the mean  $\pm$  95% confidence intervals

## dCasRxAPEX2 CE3-targeted proteomics enrichment



$p < 0.05$   
FDR  $< 0.25$

203 proteins



Involvement in splicing

63 proteins

TCGA-PRAD  
n = 500

mCRPC  
n = 208



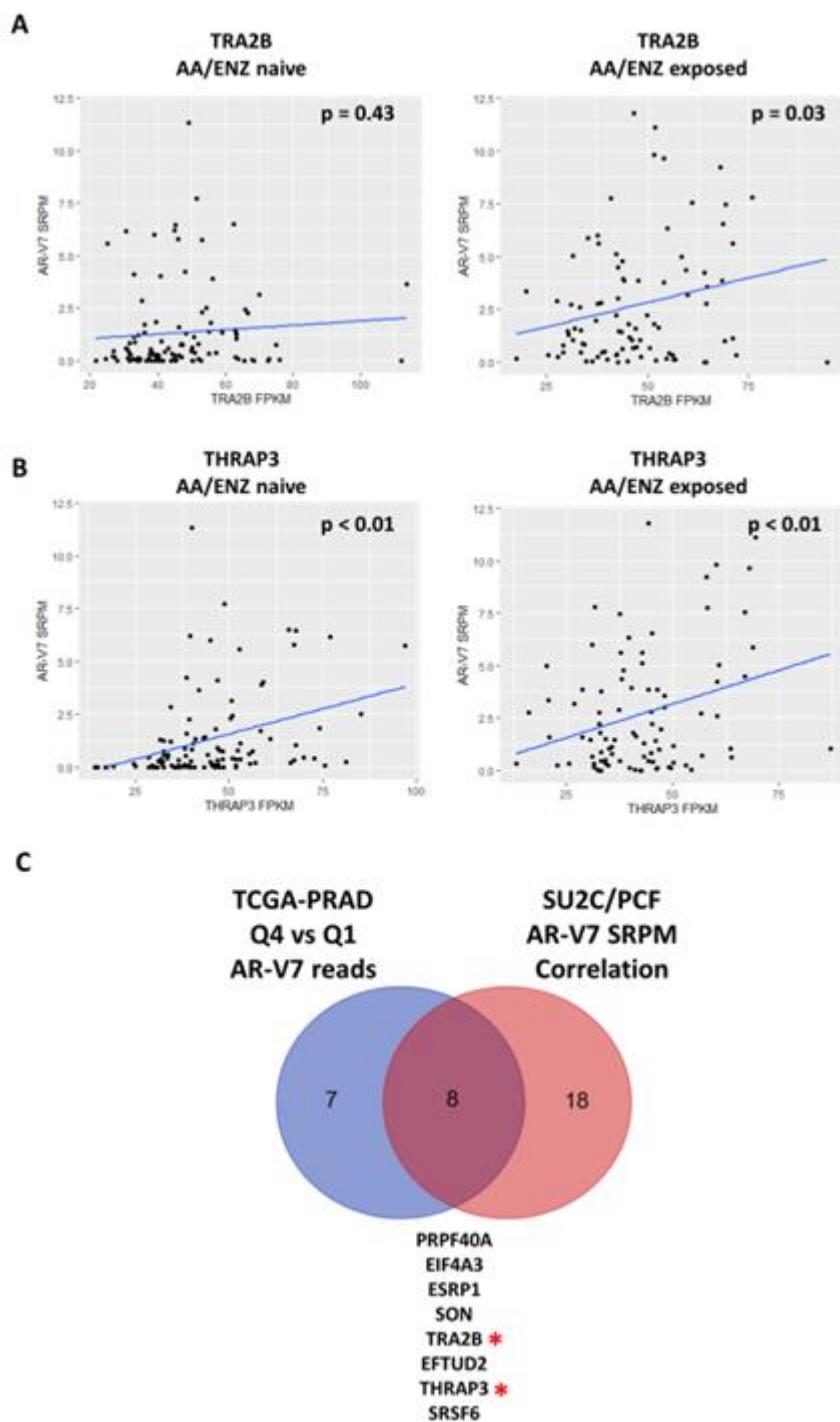
Correlation with AR-V7 splicing and expression signatures

Figure 6.5 - Summary of CWR22Rv1 CE3 interactome triangulation with publicly available datasets

dCasRx-APEX2 workflows in CWR22Rv1 enriched 203 CE3-interacting proteins at  $p < 0.05$  and FDR  $< 0.25$ . This panel was further refined for proteins with established splicing activity, yielding a list of 63. Two large RNA-Seq datasets obtained from PCa and mCRPC patients enrolled into TCGA and SU2C/PCF studies were analysed for associations between AR-V7 splicing, activity signatures and expression of genes encoding the panel of 63 proteins

As a result of these analyses, two protein candidates were taken forward for experimental investigation: TRA2B and THRAP3. Both genes encoding these were in our panel of 26 that significantly correlate with AR-V7 SRPM in the SU2C/PCF mCRPC cohort (Figure 6.1B), and *TRA2B* expression correlates to a far greater degree in AA/ENZ-exposed than AA/ENZ-naïve patients (Figure 6.6A). *THRAP3* exhibited strong correlation with AR-V7 SRPM in both subgroups of patients, at greater statistical significance than *TRA2B* (Figure 6.6B). In addition, *TRA2B* was the second highest-correlating gene with the AR\_V7\_UP activity signature and was amongst the top quartile of genes exhibiting differentially increased correlation between AR\_V7\_UP and AR\_FL\_UP (Figure 6.2D, E). Furthermore, *THRAP3* and *TRA2B* were amongst the genes correlating highest with the 59-gene (Sharp *et al.*, 2019) or 41-gene (Cai *et al.*, 2018) clinically observed AR-V7 expression signatures (Figure 6.3).

Both genes were in a group of just 4 out of the 63-gene list to have greater expression in AR-V7 positive TCGA-PRAD patients at an FDR < 0.05 (Figure 6.4A), whilst also being one of 15 to display significantly higher AR-V7 spliced reads in Q4 over Q1 gene expression groups (Figure 6.4B). *TRA2B* and *THRAP3* also belonged to an 8-gene set that both correlated with AR-V7 SRPM in the SU2C/PCF cohort and demonstrated higher AR-V7 between quartile groups in the TCGA-PRAD dataset at  $p < 0.05$  (Figure 6.6C).



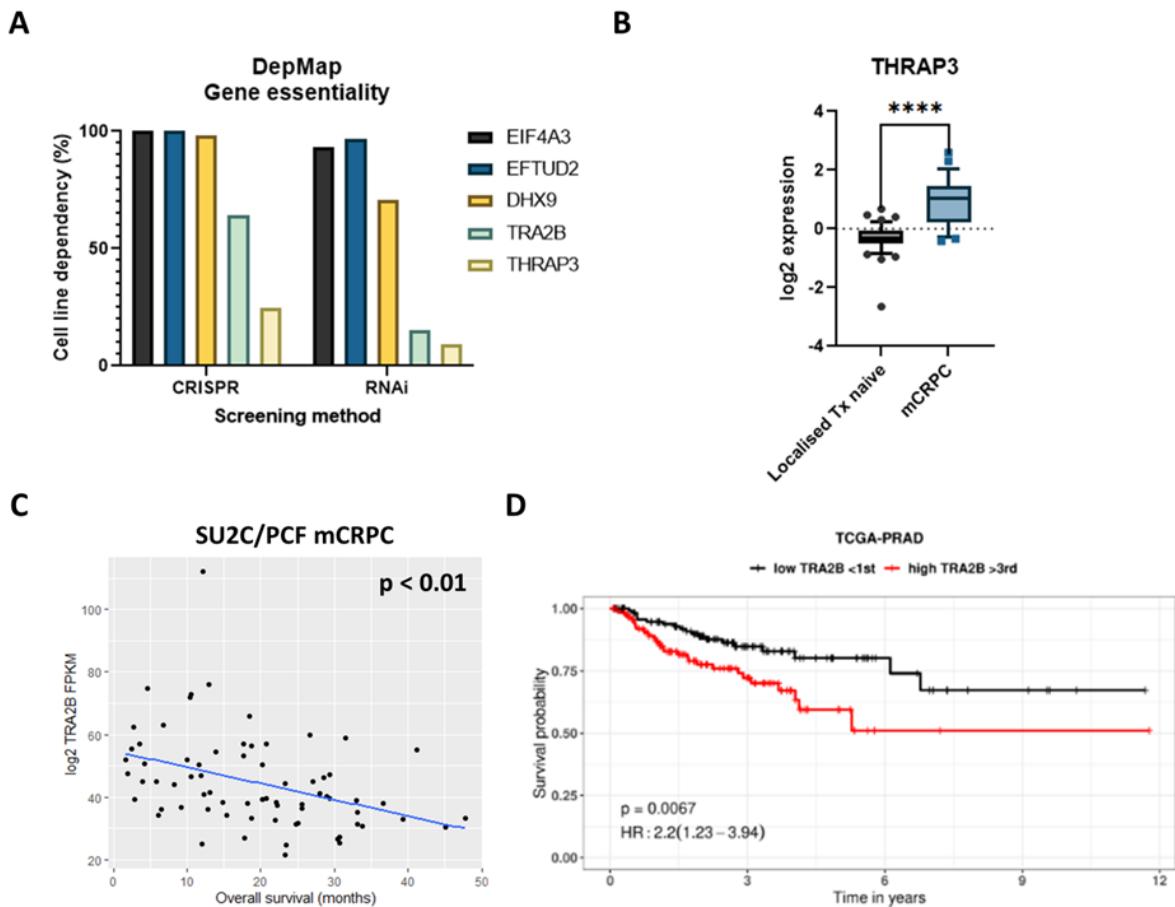
**Figure 6.6 - TRA2B and THRAP3 emerge as lead splicing candidates for functional analysis**

**A.** *TRA2B* FPKM expression was correlated with AR-V7 SRPM in AA/ENZ naïve and exposed patients of the SU2C/PCF mCRPC cohort (correlation data sourced from analysis summarised in 6.1) **B.** *THRAP3* FPKM expression was correlated with AR-V7 SRPM in AA/ENZ naïve and exposed patients of the SU2C/PCF mCRPC cohort (correlation data sourced from analysis summarised in 6.1) **C.** Eight genes were statistically significant in both the analysis performed in (A) and (B), and in the Q4 vs Q1 AR-V7 analysis summarised in 6.4B. Two of these genes were *TRA2B* and *THRAP3* (highlighted with red asterisk)

Other genes also presented as strong potential candidates from our analyses, including *EIF4A3*, *EFTUD2*, and *DHX9*. All genes of interest were examined on the cancer dependency map (DepMap) portal, an analysis led by the Broad Institute whereby genome-wide CRISPR and RNAi screens were used to determine pan-cancer cellular gene essentiality (Tsherniak *et al.*, 2017). *EIF4A3*, *EFTUD2* and *DHX9* are all classified as ‘common essential’ genes, depletion of which is deleterious in the vast majority of tested cell lines indicating likely involvement in core spliceosomal machinery. Encouragingly, *TRA2B* and *THRAP3* displayed more favourable profiles, suggesting both proteins may be involved in more selective splicing regulatory processes that are not absolutely essential for cell viability (Figure 6.7A).

Additional supplementary lines of evidence further supported selection of *TRA2B* and *THRAP3*. Analysis of microarray gene expression data from a cohort of localised, treatment-naïve PCa and mCRPC tumours (Grasso *et al.*, 2012) highlights *THRAP3* as one of the most highly overexpressed of all assayed genes in mCRPC (Figure 6.7B). Though this dataset does not include information on AR-V7, thus was omitted from original analyses, it provided auxiliary justification for *THRAP3* selection. Moreover, survival analysis revealed *TRA2B* expression to negatively correlate with overall survival in the SU2C/PCF mCRPC dataset (Figure 6.7C) and to significantly associate with lower progression-free survival in the TCGA-PRAD cohort, further bolstering *TRA2B* selection as a clinically relevant splicing factor in CRPC that merits study (Figure 6.7D).

Lastly, both *TRA2B* and *THRAP3* have been noted in the literature previously, though not in relation to AR-V splicing. A 2015 study of an independent PCa cohort in China noted *TRA2B* expression to associate with poor survival outcomes, biochemical recurrence, and lymph node metastases (Diao *et al.*, 2015). *THRAP3* was examined in another publication that performed phosphoproteomic comparisons between LNCaP and an androgen-independent derivative, LNCaP-AI. They found multiple differences in *THRAP3* phosphorylation sites between LNCaP-AI and LNCaP, which was shown to alter its association with a range of splicing factors (Ino *et al.*, 2016). As LNCaP-AI are known to express a degree of AR-V7 compared to their androgen-sensitive parental cell line (Liu *et al.*, 2014), this study potentially implicates *THRAP3* in this process. Neither of these publications examined AR-V expression or splicing thus conferring a sufficiently novel angle for our study whilst also providing encouragement for our selected factors.



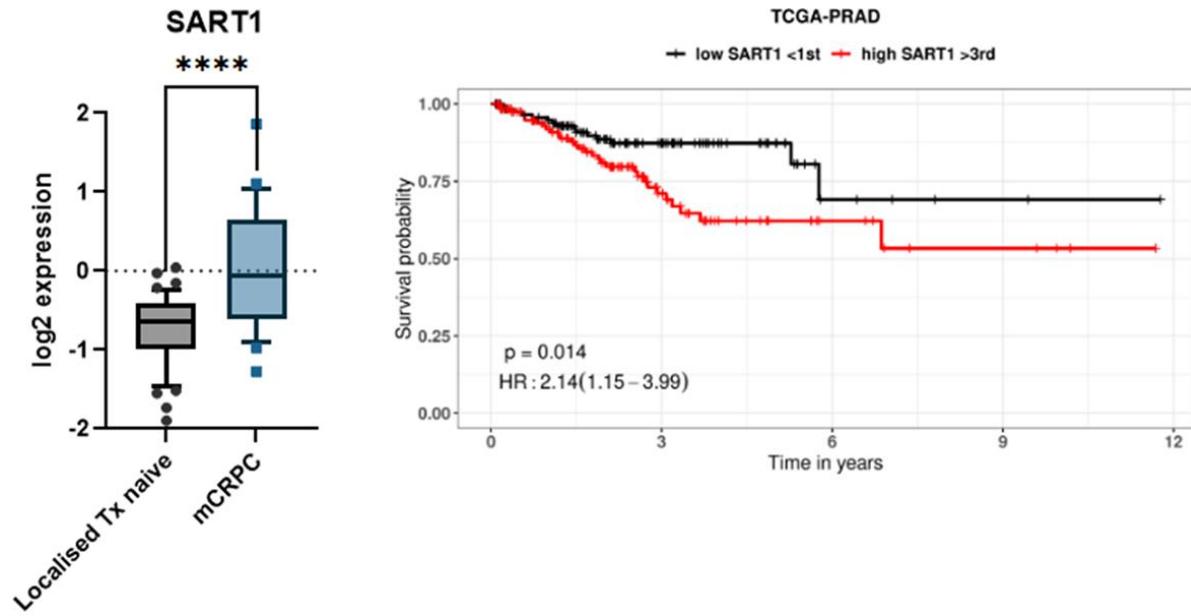
**Figure 6.7 - Supplementary analyses further vindicate selection of TRA2B and THRAP3 as candidates**

**A.** Pan-cellular essentiality of *TRA2B* and *THRAP3*, along with other genes that scored well in our analyses, was examined using the cancer dependency map (DepMap). DepMap assesses deleterious effects of gene depletion across several hundred cell types by both CRISPR knockout (CRISPR) and RNA interference (RNAi). Cell line dependency indicates the percentage of cell types screened by DepMap for which a given gene is essential to survival **B.** Analysis of expression differences was performed between localised, treatment (Tx) naïve PCa vs mCRPC tumours from a previous microarray study (Grasso *et al.*, 2012). Differential expression was analysed between groups using *limma* via the GEO2R online tool (Barrett *et al.*, 2013). Log 2 normalised expression is plotted for *THRAP3*. Significance is determined by *limma* analysis FDR (\*\*\*\* = p < 0.0001) **C.** Pearson's correlation coefficient between log2 *TRA2B* FPKM and overall survival (months) in the SU2C/PCF mCRPC cohort was determined **D.** Difference in progression-free interval between the top (red) and bottom (black) patient quartiles of *TRA2B* expression in the TCGA-PRAD cohort was determined by logrank test. Survival analysis in (D) was performed using the Tumor online Prognostic analysis Platform (ToPP) (Ouyang *et al.*, 2022)

An additional factor to be included in functional experiments was SART1. Given the aforementioned possibility of workflows selecting for cotranscriptional as well as splicing processes, it would be of interest to select a protein that was significantly enriched by dCasRx-APEX2 proteomics that may not be an AR-V7 specific regulator based on analyses presented here. A previous CRISPR screen noted that SART1 depletion appears to mediate a degree of depletion in both AR-FL and AR-V7 (Tang *et al.*, 2022). Furthermore by using some of the same

analyses as in Section 6.7, *SART1* expression is significantly increased in mCRPC patients and negatively associates with survival in the TCGA-PRAD cohort (Figure 6.8). Therefore its enrichment by dCasRx-APEX2 and association with these metrics, whilst lacking AR-V7 specificity, may indicate that it functions as a more general oncogenic regulator of the *AR* gene.

TRA2B, THRAP3 and SART1 proteins will now be functionally investigated in the forthcoming chapter by analysing any effects their depletion has on AR-FL/AR-V7 levels.



**Figure 6.8 - SART1 was selected as an additional factor upregulated in mCRPC and associated with poor PCa survival**

*SART1* expression differences between mCRPC and therapy-naïve PCa were analysed by the same methods as used in Figure 6.7B. Survival analysis between the top (red) and bottom (black) *SART1* patient expression quartiles in the TCGA-PRAD cohort was performed as in Figure 6.7D

## 6.4 Experimental validation confirms TRA2B as an AR-V7 splicing factor

To investigate functional involvement of the selected factors in AR-V7 splicing, each gene was depleted using siRNA. All experiments used a non-targeting (NT) siRNA as a control for comparison with splicing factor-targeting siRNA. CWR22Rv1 were used initially in order to validate observations from proteomics experiments. This was first tested by qPCR, whereby levels of AR-FL and AR-V7 were assayed using a common *AR* exon 3 primer in combination with a reverse primer selective for either exon 4 (AR-FL) or CE3 (AR-V7) (Figure 6.9A). Neither TRA2B or THRAP3 depletion mediated reductions in AR-FL or AR-V7 transcripts, instead the opposite effect was observed with both mRNAs increasing in abundance (Figure 6.9B). Conversely, SART1 knockdown caused a significant decrease in both AR-FL and AR-V7, potentially alluding to its function as a general driver of *AR* transcription (Figure 6.9B). Levels of target gene depletion were also checked, and it is notable that siRNA only modestly downregulated TRA2B in this instance (Figure 6.9C).

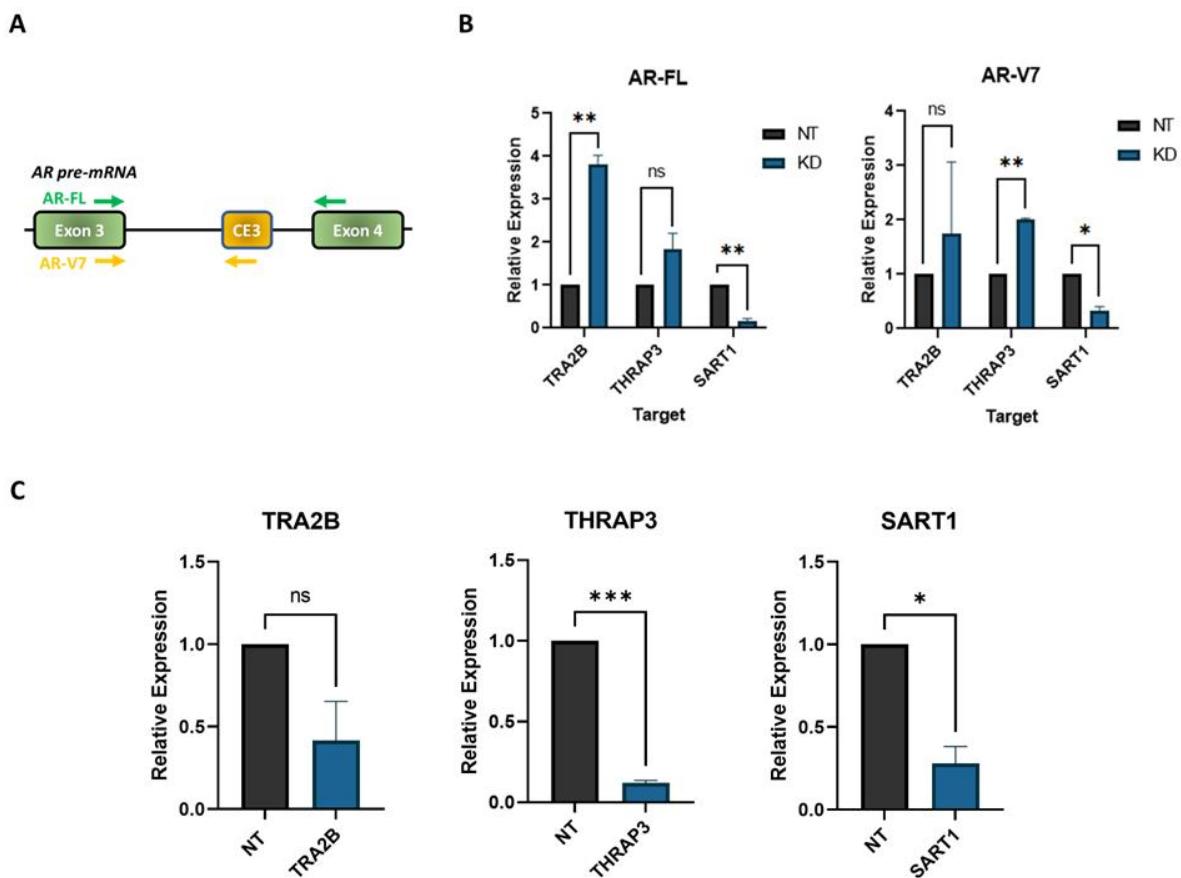


Figure 6.9 - Initial assessment of AR-FL and AR-V7 mRNA levels upon target depletion in CWR22Rv1

**A.** Summary of qPCR assay used to determine AR-FL and AR-V7 mRNA levels. A forward primer which binds within exon 3 can distinguish between AR-FL and AR-V7 by use of differing reverse primers selective for either

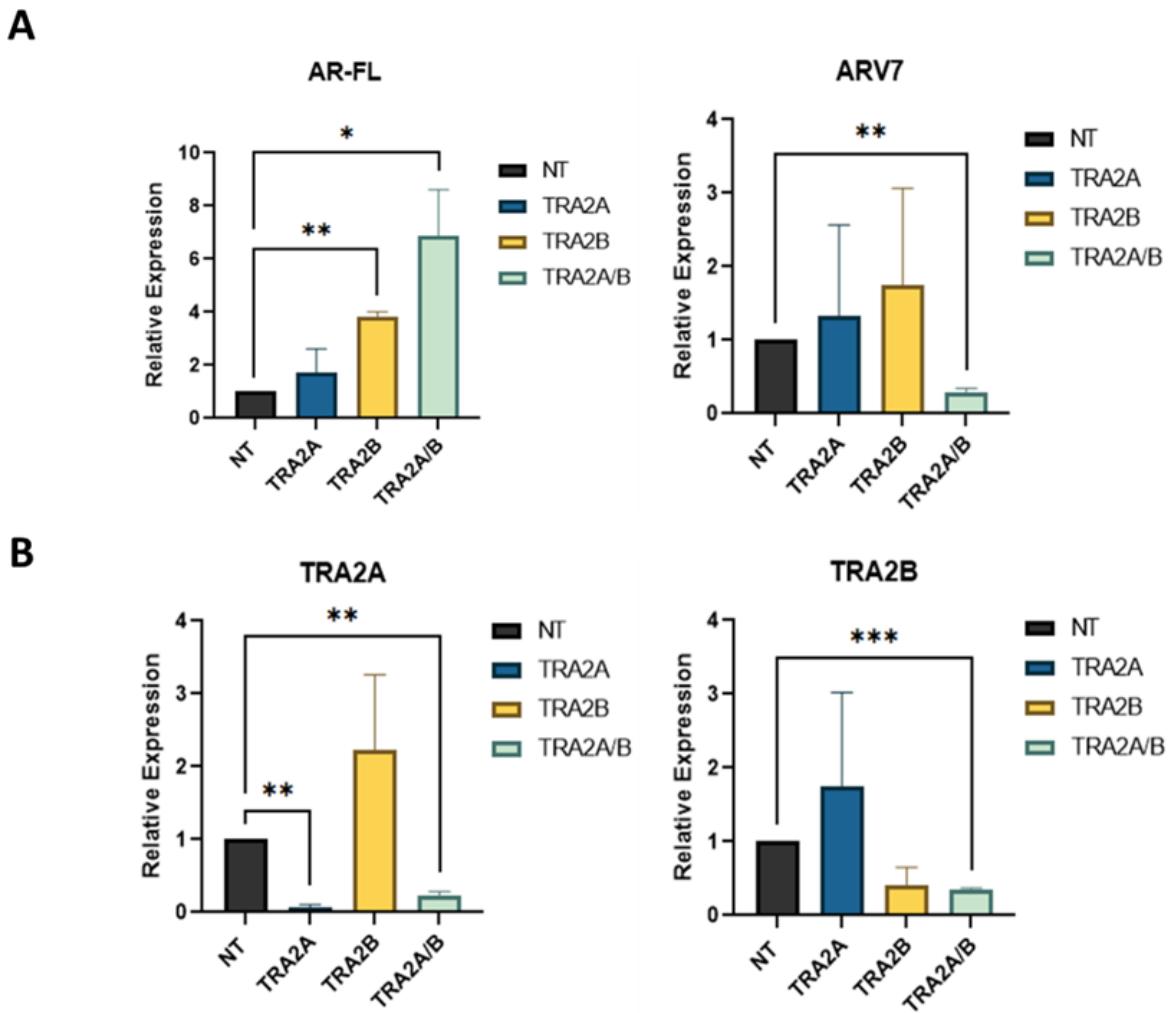
*AR* exon 4 or CE3 **B**. CWR22Rv1 were transfected with 25 nM siRNA that targets TRA2B, THRAP3 or SART1 and incubated for 72 hours before RNA was extracted. RT-qPCR was used to analyse levels of AR-FL and AR-V7 transcript. Data comprises n = 2 independent biological replicates, plotted as mean  $\pm$  SEM. Unpaired t-test was used for determination of statistical significance (\* = p < 0.05, \*\* = p < 0.01, ns = p > 0.05) **C**. Samples described in (B) were also checked for depletion of their respective target genes. Data comprises n = 2 independent biological replicates, plotted as mean  $\pm$  SEM. Unpaired t-test was used for determination of statistical significance (\* = p < 0.05, \*\*\* = p < 0.001, ns = p > 0.05)

Despite enrichment by dCasRx-APEX2 proteomics workflows and strong evidence from clinical cohorts, depletion of TRA2B or THRAP3 actually increased levels of both AR transcripts assayed. A subsequent review of the literature and consultation with the Elliott lab in Newcastle University, a group that works extensively on TRA2 proteins, highlighted that TRA2B has a closely related paralog, TRA2A. Their 2014 study in the breast cancer cell line MDA-MB-231 revealed that dual depletion of both TRA2B and TRA2A was required for alterations in splicing of *CHEK1* exon 3, with lone TRA2B or TRA2A knockdown failing to elicit this effect (Best *et al.*, 2014).

Encouraged by this, TRA2B was re-examined by using siRNA targeted to TRA2A, TRA2B, or a combination of the two. This revealed a striking difference in effect on AR-FL and AR-V7 mRNA in contrast with single siRNA treatment. Dual TRA2A/TRA2B depletion mediated an even greater increase in AR-FL than TRA2B knockdown alone (Figure 6.10A, lower significance for TRA2A/B arm is due to replicate variability). Crucially, this dual depletion arm resulted in a significant decrease in AR-V7 mRNA (Figure 6.10A). This reciprocal effect on AR-FL and AR-V7 levels is highly relevant to this study. Were TRA2 proteins truly controlling *AR* mRNA splicing in the absence of an effect on gene transcription, it naturally follows that if the exon 3/CE3 (ARV-V7) splice site is not being selected as efficiently there will be a concomitant increase in exon 3/exon 4 (AR-FL) splice junctions, precisely as we observed.

An element of reciprocal control exists between TRA2 proteins, as knockdown of one paralog has been shown to mediate increases in the other in a model of breast cancer (Best *et al.*, 2014). This is due to *TRA2* autoregulation in which TRA2B mediates inclusion of a 'poison exon', encoding a premature stop codon, into its own transcript as an autoregulatory mechanism (Stoilov, 2004). Furthermore, poison exon regulation occurs between *TRA2* genes, with knockdown of each paralog reducing premature stop codon inclusion in the other (Best *et al.*, 2014). We observed a similar effect in CRPC cell line CWR22Rv1, with gene expression

of TRA2B increasing upon TRA2A depletion, and vice versa (Figure 6.10B). We also confirmed that dual TRA2A/B depletion robustly reduces expression of both genes (Figure 6.10B).

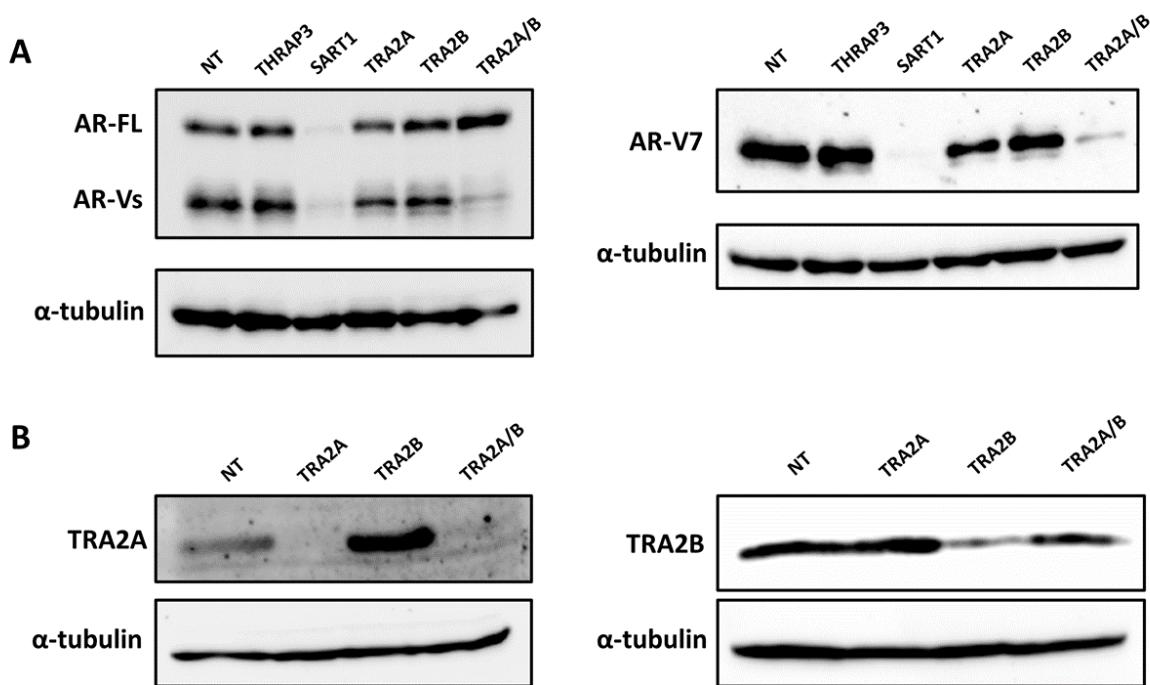


**Figure 6.10 - Dual TRA2 depletion in CWR22Rv1 alters AR mRNA splicing decisions to significantly deplete AR-V7 levels**

**A.** CWR22Rv1 were transfected with 25nM siRNA targeting TRA2A, TRA2B or with a combination of TRA2A and TRA2B siRNA (TRA2A/B). Cells were incubated for 72 hours before RNA was extracted and RT-qPCR was used to analyse levels of AR-FL and AR-V7 transcript. Data comprises  $n = 2$  independent biological replicates, plotted as mean  $\pm$  SEM. Unpaired t-test was used for determination of statistical significance. Only results significant at  $\alpha = 0.05$  or lower have significance denoted (\* =  $p < 0.05$ , \*\* =  $p < 0.01$ ) **B.** Samples described in (A) were also checked for levels of TRA2A and TRA2B. Data comprises  $n = 2$  independent biological replicates, plotted as mean  $\pm$  SEM. Unpaired t-test was used for determination of statistical significance. Only results significant at  $\alpha = 0.05$  or lower have significance denoted (\*\* =  $p < 0.01$ , \*\*\* =  $p < 0.001$ )

Next, confirmation of these events at the protein level was sought. Western blot results recapitulated qPCR data, with THRAP3 knockdown having no effect, SART1 abrogating AR-FL and AR-V7 levels, and a combined TRA2A/B depletion increasing and decreasing levels of AR-

FL and AR-V7, respectively (Figure 6.11A). Reciprocal upregulation of TRA2A upon TRA2B knockdown was even more apparent at the protein level, whereas transfection with TRA2A siRNA had little discernible effect on TRA2B protein (Figure 6.11B). Additionally, TRA2B knockdown was incomplete at the protein level. Interestingly, both the latter two observations were observed in breast cancer cell line MDA-MB-231, where no apparent increase in TRA2B protein was seen after TRA2A RNAi (Best *et al.*, 2014). The authors also failed to elicit as great a reduction in TRA2B protein as for TRA2A, which we also observed in CWR22Rv1 (Best *et al.*, 2014). Whether this is a consequence of siRNA potency or TRA2B regulatory mechanisms is unclear.



**Figure 6.11 - Effects of dual TRA2A/B depletion in CWR22Rv1 are also observed at the protein level**

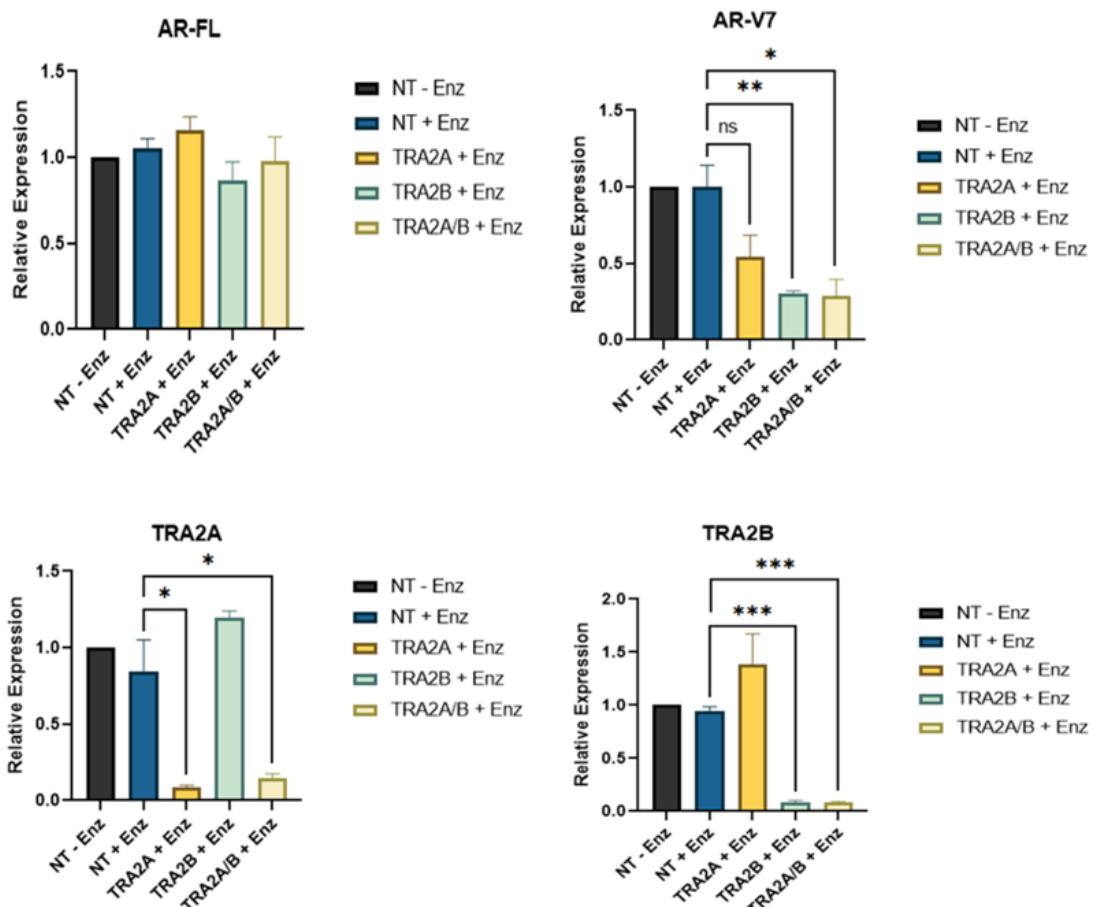
**A.** CWR22Rv1 were transfected with the indicated siRNAs at 25 nM and incubated for 72 hours. Then, cell lysates were harvested for analysis of AR-FL, AR-Vs and AR-V7 protein levels by western blot **B.** Sample described in (A) were also analysed for levels of TRA2A and TRA2B. Western blots are representative of  $n = 3$  biological replicates.  $\alpha$ -tubulin was used as a loading control throughout

Given this apparent specificity for AR-V7 production, TRA2A/B were taken forward as proteins of most interest. Based on analysis of publicly available data, it was likely that TRA2 splicing control is not solely restricted to the CWR22Rv1 genetic background. Therefore analysis was expanded to a second AR-V expressing PCa cell line, VCaP. CWR22Rv1 express constitutively high levels of AR-V7, whereas VCaP exhibit lower steady-state levels of AR-V7 which can be

augmented by treatment with NSAA enzalutamide or androgen deprivation (Paschalis, Welti *et al.*, 2021; Liu *et al.*, 2014). VCaP were transfected with the same siRNAs as for CWR22Rv1 in steroid-depleted media (SDM) and treated with 10  $\mu$ M enzalutamide to maximise AR-V7 levels. An NT siRNA arm without enzalutamide treatment was also included to check for NSAA-induced increases in AR-V7 levels.

qPCR analysis shows that this experimental setup failed to trigger an enzalutamide-mediated spike in AR-V7 splicing between NT -/+ enzalutamide arms at the 72 hour timepoint. Nevertheless, by comparing results between enzalutamide-treated arms the effects of TRA2 depletion in reducing AR-V7 mRNA is clear (Figure 6.12A). Reductions in AR-V7 are, however, not accompanied by the significant induction of AR-FL seen in CWR22Rv1 (Figure 6.12A). The *AR* gene is amplified to upwards of 20 copies in VCaP (Makkonen *et al.*, 2011). Given that an *AR* auto-regulatory loop has been noted in models of CRPC (Alimirah *et al.*, 2006; Isaacs *et al.*, 2012), it is possible that this high-level *AR* amplification may restrain increases in AR-FL relative to CWR22Rv1.

Interestingly, in the VCaP cell line, it appears single knockdown of TRA2B, and to a lesser extent TRA2A, depletes AR-V7 levels. Importantly AR-FL expression is largely unaffected, indicating that *AR* mRNA TRA2 splicing control in VCaP is specific to AR-V7 (Figure 6.12A). The same extent of TRA2 paralog compensation seen in CWR22Rv1 is not readily apparent in qPCR gene expression readouts (Figure 6.12B). This requires confirmation at the protein level, however. Precisely why this difference would occur between cell lines and why poison exon-mediated reciprocal control between TRA2A/B is apparently absent in VCaP is not immediately obvious.

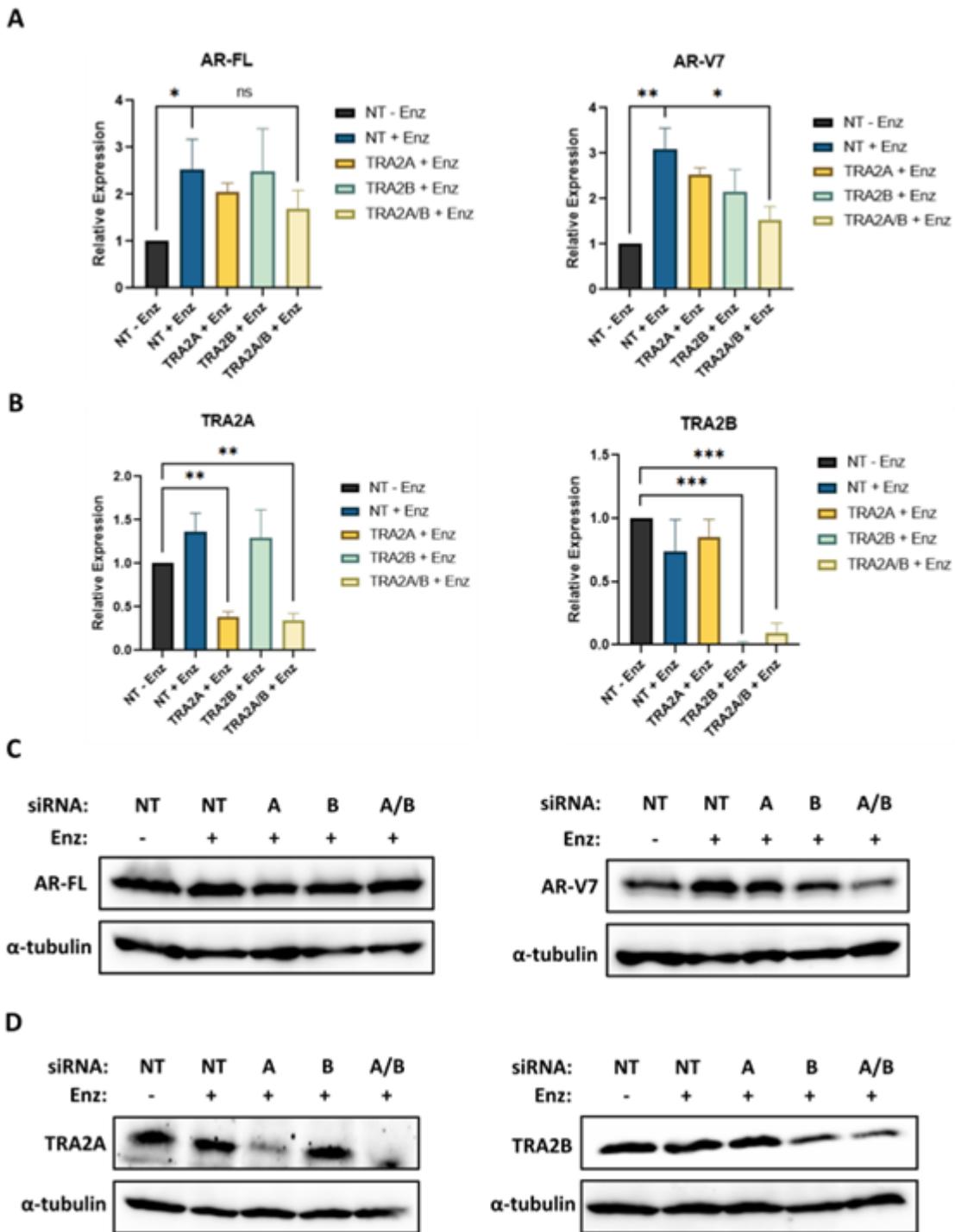
**A****Figure 6.12 - AR-V7 splicing is controlled by TRA2 in VCaP**

**A.** VCaP were transfected with siRNA in SDM and treated with 10  $\mu$ M enzalutamide (Enz) and incubated for 72 hours, before RNA was harvested for RT-qPCR analysis of AR-FL and AR-V transcript levels. An enzalutamide-untreated arm was also included to confirm NSAA-mediated induction of AR-V7 levels. **B.** Samples described in (A) were analysed for levels of TRA2A and TRA2B mRNA. All qPCR data in (A) and (B) comprises  $n = 3$  independent biological replicates, plotted as mean  $\pm$  SEM. Unpaired t-test was used for determination of statistical significance. Not all statistical test results are plotted for ease of visualisation (\* =  $p < 0.05$ , \*\* =  $p < 0.01$ , \*\*\* =  $p < 0.001$ , ns =  $p > 0.05$ )

Initial results in VCaP were encouraging, although successful induction of AR-V7 with enzalutamide, and potential abrogation of this with TRA2 knockdown, would be a substantial illustration of AR-V7 splicing control by TRA2. It was hypothesised that performing experiments in SDM may have negated any effect of enzalutamide on AR-V7 production, essentially raising the baseline above which further NSAA-induced increases would be inconsequential to the rate of *AR* transcription. Alternatively, combining the complete androgen absence of SDM with enzalutamide and siRNA transfection may have been too great a cellular stressor.

Cell culture media supplemented with 10% FBS contains a castrate level of androgen at < 20 ng/dl testosterone, and despite this not being as low as androgen-null SDM, use of standard 10% FBS with enzalutamide should reasonably reflect a typical clinical scenario (Sedelaar and Isaacs, 2009; Song and Khera, 2014). The above experimental setup in VCaP cells was therefore repeated; this time using standard 10% FBS as opposed to 10% dextran-coated charcoal stripped FBS used in SDM.

With this approach, 10  $\mu$ M enzalutamide significantly increased levels of AR-V7 and, to a lesser extent, AR-FL mRNA as measured by qPCR (Figure 6.13A). TRA2 depletion in enzalutamide-treated VCaP lowered AR-FL levels to an extent, however there is a significantly greater differential between enzalutamide-exposed TRA2A/B and NT siRNA arms for AR-V7 (Figure 6.13A). Furthermore as previously observed, no reciprocal TRA2 paralog upregulation was detected in VCaP (Figure 6.13B). These observations were recapitulated at the protein level with no apparent change in AR-FL seen between arms, whereas AR-V7 is augmented with enzalutamide, an effect that is quashed by TRA2 knockdown (Figure 6.13C). Furthermore, consistent with qPCR results, no compensatory TRA2 upregulation is seen in western blot data (Figure 6.13D).



**Figure 6.13 - Enzalutamide mediates an elevation in AR-V7 in VCaP, which is quashed by TRA2 depletion**

**A.** VCaP were transfected with siRNA in full media containing 10% FBS and treated with 10  $\mu$ M enzalutamide (Enz) and incubated for 72 hours, before RNA was harvested for RT-qPCR. An enzalutamide-untreated arm was also included to confirm NSAA-mediated induction of AR-V7 levels **B.** Samples described in (A) were analysed for levels of TRA2A and TRA2B mRNA. All qPCR data in (A) and (B) comprises  $n = 3$  independent biological replicates, plotted as mean  $\pm$  SEM. Unpaired t-test was used for determination of statistical significance. Not all statistical test results are plotted for ease of visualisation (\* =  $p < 0.05$ , \*\* =  $p < 0.01$ , \*\*\* =  $p < 0.001$ , ns =  $p > 0.05$ ) **C.** Lysates from VCaP cells treated as in (A) were analysed for levels of AR-FL and AR-V7 by Western blot **D.** Samples described in (C) were also analysed for levels of TRA2A and TRA2B. Western blot data is representative of  $n = 3$  biological replicates.  $\alpha$ -tubulin was used as a loading control throughout

It is notable that despite enzalutamide increasing AR-V7 levels in VCaP, and the ability to block this by depleting TRA2 proteins, expression of TRA2A or TRA2B do not themselves increase with enzalutamide exposure (Figure 6.13). The reason for this may lie with TRA2 protein functional regulation, which will be explored in Section 6.5. Given these results in CWR22Rv1 and VCaP cells suggest a specificity for AR-V7, this TRA2 splicing axis has met criteria for the original project aims and will form the focus of subsequent experiments.

## 6.5 Transcriptomic analyses further characterise TRA2 functions in CWR22Rv1

Based on results presented in Section 6.4, TRA2B and its paralog TRA2A were taken forward as lead candidates for analysis due to their specific effect on AR-V7 levels being likely to indicate a splicing effect rather than general transcriptional modulation. First, their functional involvement in splicing will be explored further, as the remainder of this thesis will focus on the TRA2 splicing axis.

Transformer 2 (TRA2) was first characterised as an alternative splicing regulator involved in sex determination in *Drosophila melanogaster*, before two mammalian homologues, TRA2 $\alpha$  (TRA2A) and TRA2 $\beta$  (TRA2B), were discovered (Förch and Valcárcel, 2003; Dauwalder, Amaya-Manzanares and Mattox, 1996; Beil, Scretton and Stamm, 1997). TRA2A and TRA2B share approximately 75% sequence homology and are members of the SR protein family of splicing factors, which typically function to recruit spliceosome components at nascent transcript splice sites (Jeong, 2017). Structurally, both homologs have an RNA-recognition motif (RRM) flanked by N and C-terminal arginine/serine rich (RS) domains responsible for protein-protein interactions (Figure 6.14A) (Xue, Ma and Zhang, 2023). Using the constraint-based alignment tool for multiple protein sequences (COBALT) (Papadopoulos and Agarwala, 2007), comparison of UniProt amino acid FASTA sequences for TRA2A (UniProt Q13595) and TRA2B (UniProt P62995) highlights the extensive similarity between TRA2 paralogs, though there are notable differences including polyglycine repeats after their respective RRMs (Figure 6.14A). TRA2 RRMs also exhibit 81% homology (Figure 6.14A), which would be expected to provide a level of overlap between nucleotide binding preferences.

The apparent functional redundancy existing between TRA2 proteins has been noted by this work on AR-V7 in CWR22Rv1, as well as work by others in breast cancer where depletion of both TRA2A and TRA2B was required for appreciable splicing changes in exons encoding a range of genes including *CHEK1*, *ATRX*, *GLYR1* and *CEP95* (Best *et al.*, 2014). Further supporting this is experimental evidence that ectopic expression of either TRA2A or TRA2B can activate splicing of the *TRA2A* poison exon (Grellscheid, Dalglish and Best *et al.*, 2011). This suggests that for at least some target RNAs, TRA2 paralogs can compensate adequately.

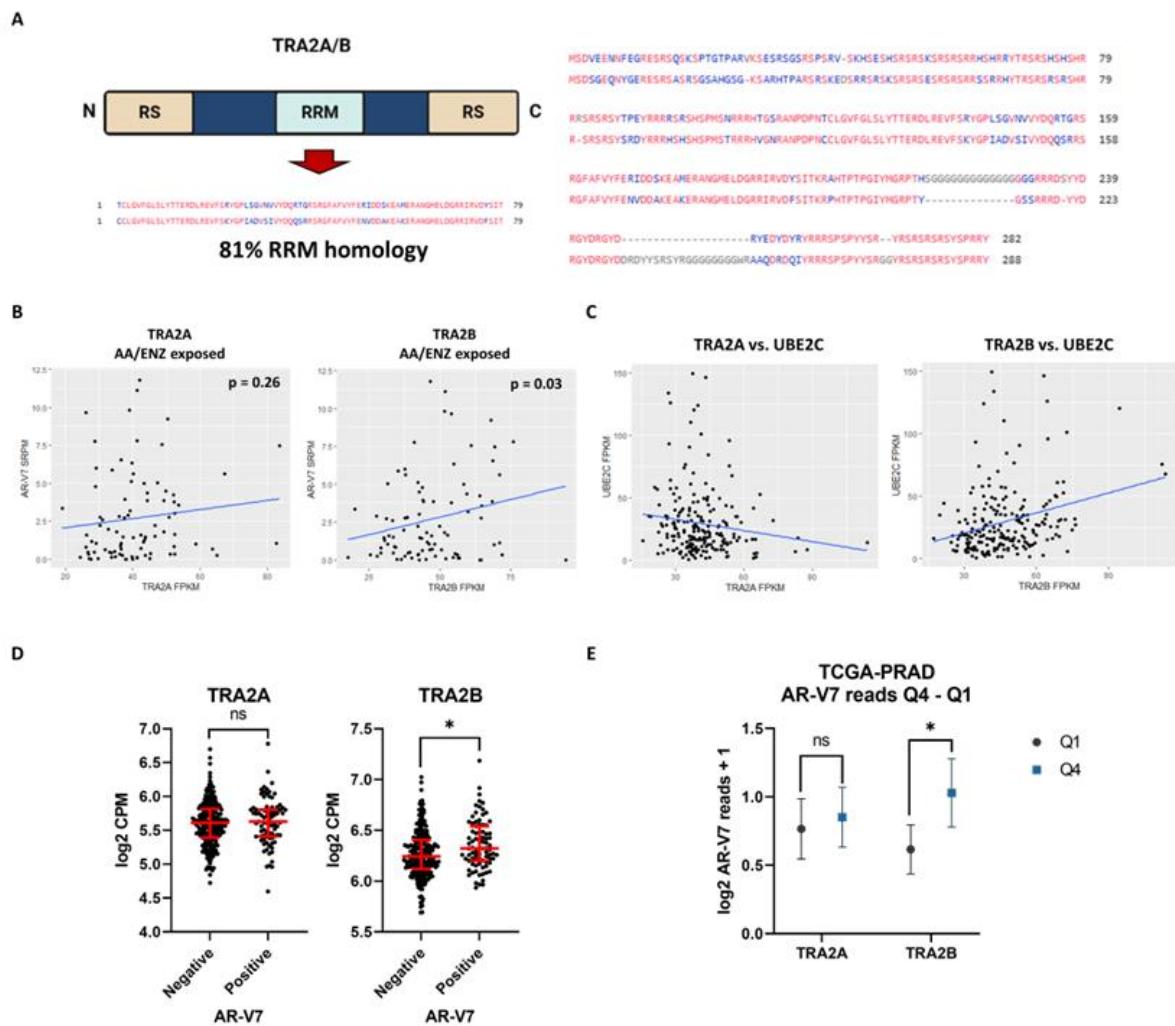
However, *TRA2B* knockouts in mice are embryonic lethal despite retention of functional TRA2A (Mende *et al.*, 2010). Additionally, *TRA2A* and *TRA2B* exhibit inverted expression

patterns between maturation-normal and maturation-deficient human oocytes (Li *et al.*, 2020); whilst our experiments using VCaP show that TRA2B knockdown alone has a greater effect on AR-V7 than TRA2A, and is almost equivalent to that of dual TRA2A/B depletion (Figure 6.12, Figure 6.13). Therefore, it is evident that TRA2 proteins do not exhibit complete compensatory ability. Furthermore, the TRA2 RRM differs in amino acid composition between paralogs by ~20% (Figure 6.14A). Given the tightly regulated nature of alternative splicing, this difference may confer variation in splicing enhancer specificity between TRA2A and TRA2B.

Interestingly, TRA2 proteins have been shown as preferential activators of alternative rather than constitutive splicing. Previous work performed using HeLa nuclear extracts has demonstrated that both TRA2 paralogs bind to purine-rich RNA splicing enhancer elements (Tacke *et al.*, 1998). Additionally, by using SR-protein deficient cytoplasmic S100 protein extracts, it was shown that unlike other SR proteins, TRA2 proteins alone are unable to render these extracts splicing-proficient, and TRA2A or TRA2B only enhanced splicing of specific purine-rich sequences when supplemented with low concentrations of nuclear extract (Tacke *et al.*, 1998). Consistent with dCasRx-APEX2 proteomics having been targeted to an exonic region using AR g2, TRA2 proteins are known to predominantly bind exonic, rather than intronic, splicing enhancer elements (Grellscheid *et al.*, 2011; Lam *et al.*, 2003).

TRA2 paralog dysregulation has been implicated in a range of solid cancers including breast, lung, and colon, where they are associated with progression and therapeutic resistance. This has been reviewed extensively by others, therefore in the interests of maintaining focus on PCa, CRPC, and AR gene splicing, readers are directed to the relevant reviews (Best *et al.*, 2013; Xue, Ma and Zhang, 2023). TRA2A correlation with some of the same patient cohort metrics used for TRA2B selection in Section 6.2 was examined. Only TRA2B exhibits a statistically significant positive correlation with AR-V7 SRPM in the SU2C/PCF cohort (Figure 6.14B). Furthermore, in this cohort TRA2B expression has a strong positive association with canonical AR-V7 target gene UBE2C, part of the AR\_V7\_UP signature and used extensively by the host lab as an AR-V7 activity readout, whereas TRA2A correlates negatively (Figure 6.14C). Finally TRA2B, but not TRA2A, expression is elevated in AR-V7 positive patients from the TCGA-PRAD cohort (Figure 6.14D), whilst separation of TCGA-PRAD samples into the top and

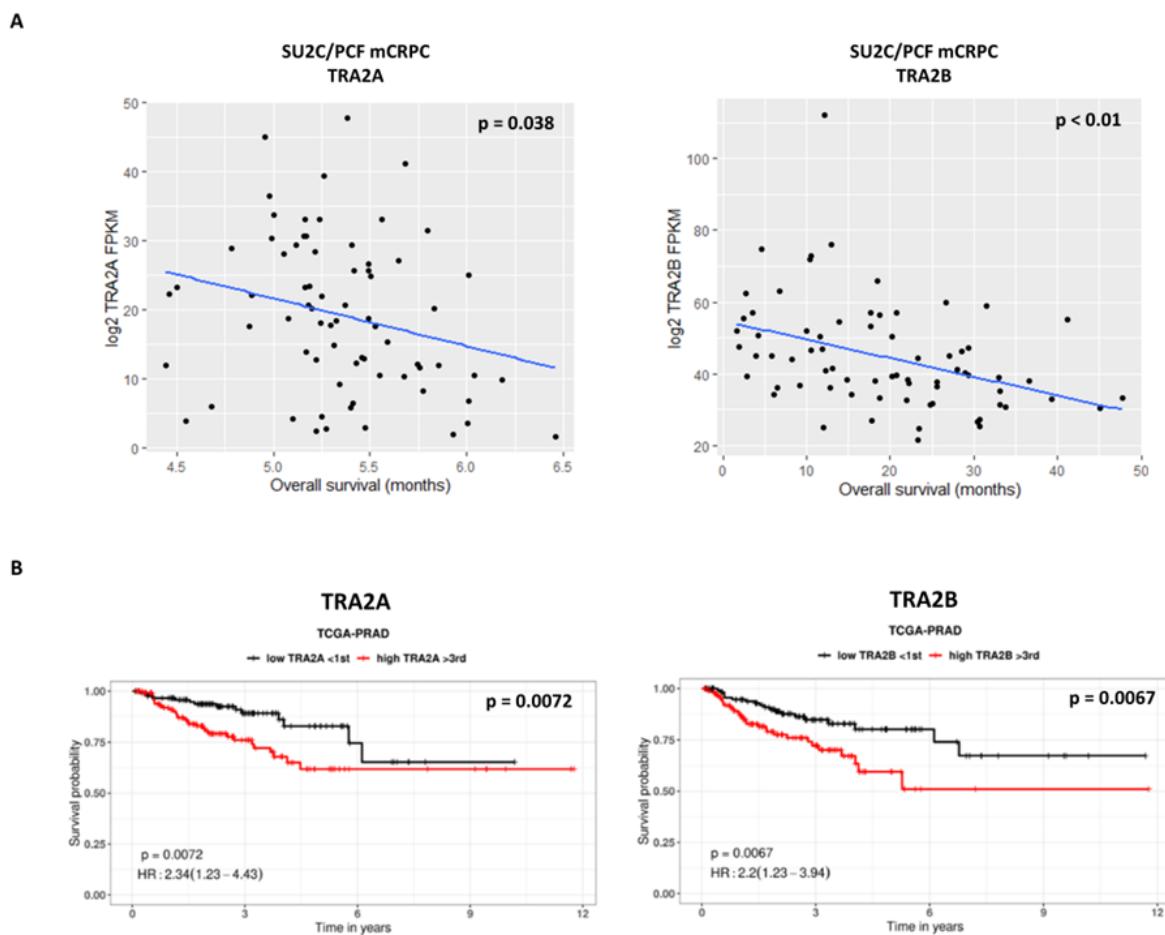
bottom quartiles of expression only reveals an appreciable difference in AR-V7 reads between quartiles for TRA2B (Figure 6.14E).



**Figure 6.14 - TRA2 protein TRA2B specifically correlates with AR-V7 in clinical patient cohorts**

**A.** Both TRA2A and TRA2B, which share approximately 75% amino acid homology, are part of the SR family of splicing proteins. They differ from other SR proteins in having both N and C-terminal RS domains, with a single RRM (Figure created using biorender.com). Alignment analysis between TRA2A and TRA2B amino acid sequences using COBALT highlights homologous amino acids (pink), as well as those that differ (blue). Polyglycine repeats regions are in grey. The RRMs of TRA2A and TRA2B exhibit ~81% homology **B.** Pearson's correlation coefficients were compared between AR-V7 SRPM and FPKM expression values of *TRA2A* and *TRA2B* in the SU2C/PCF mCRPC patient cohort **C.** Pearson's correlation coefficients were compared between *UBE2C* FPKM expression values and those of *TRA2A* and *TRA2B* in the SU2C/PCF mCRPC patient cohort **D.** Differential gene expression analysis was performed between AR-V7 positive and negative patients in the TCGA-PRAD cohort using *edgeR*. Normalised log2 counts per million (CPM) for *TRA2A* and *TRA2B* were extracted and plotted. Individual data points for each sample are plotted in black, red denotes the median value and 25th/75th percentiles (\* = FDR < 0.05, ns = FDR > 0.05) **E.** In the same TCGA-PRAD dataset as in (D), patient samples were split into groups comprising the upper (Q4) and lower (Q1) quartiles of normalised log2 CPM values for *TRA2A* and *TRA2B*. log2 AR-V7 reads were compared between Q4 and Q1 by unpaired t-test (\* = p < 0.05, ns = p > 0.05). Mean log2(AR-V7 reads +1) are plotted as the mean  $\pm$  95% confidence intervals

Only TRA2B was enriched by dCasRx-APEX2 workflows using AR g2, whereas TRA2A was not detected in any experimental arms. The above analysis further exemplifies that of the TRA2 proteins, it is TRA2B specifically that appears to associate with AR-V7 splicing events. However, the apparent functional redundancies and compensatory mechanisms between TRA2 proteins seen in CWR22Rv1 are of interest. Particularly as despite this lack of clinical AR-V7 association TRA2A still exhibits correlation with poorer survival outcomes in both SU2C/PCF (Figure 6.15A) and TCGA-PRAD (Figure 6.15B) cohorts, though to a lesser extent than TRA2B.



**Figure 6.15 - Both TRA2 paralogs are associated with poorer clinical outcomes in prostate cancer**

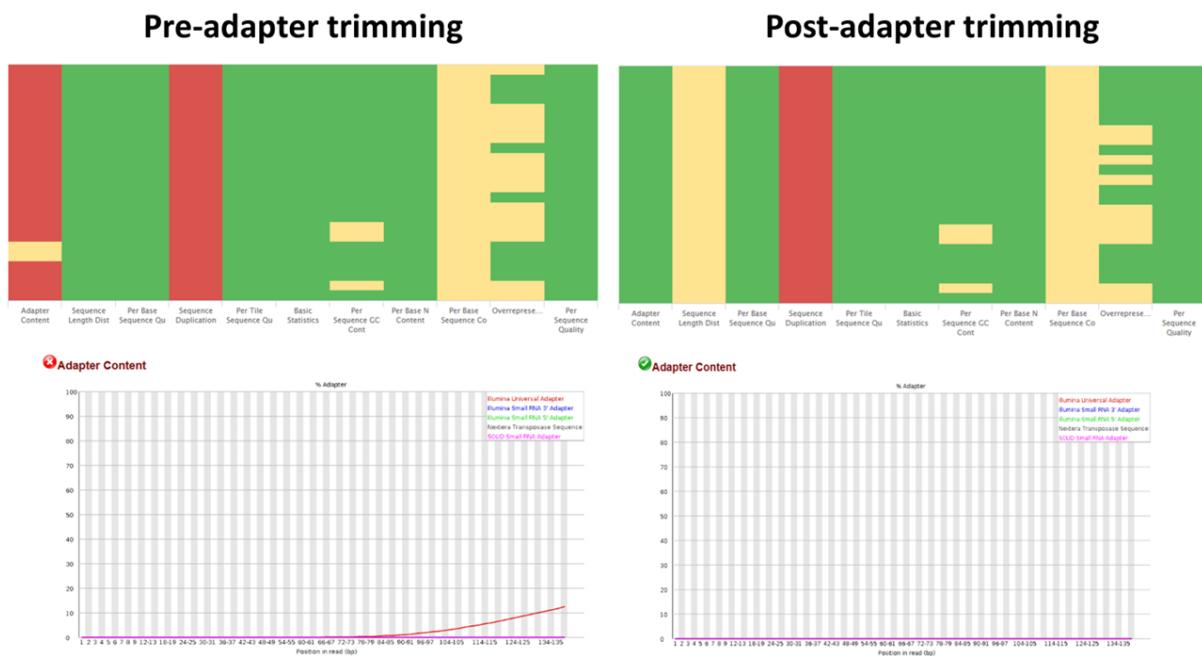
**A.** Pearson's correlation analysis was performed between log2 FPKM expression of both *TRA2A* and *TRA2B* and overall survival (months) in the SU2C/PCF mCRPC cohort. **B.** Differences in progression-free interval times between the top (red) and bottom (black) quartiles of *TRA2A* and *TRA2B* expressing patients in the TCGA-PRAD cohort were determined by logrank test. Survival analysis in (B) was performed using the Tumor online Prognostic analysis Platform (ToPP) (Ouyang *et al.*, 2022)

In order to acquire a deeper understanding of functional overlap and transcriptomic alterations resulting from TRA2 knockdown in CRPC, RNA-Seq of TRA2-depleted CWR22Rv1 was performed using the same experimental arms as previously: NT, TRA2A, TRA2B and TRA2A/B siRNA. Previous work examining transcriptomic changes in MDA-MB-231 utilised two arms: a negative control and combined TRA2A/B knockdown (Best *et al.*, 2014), whereas our setup will provide insight into how TRA2A, TRA2B and combined TRA2A/B depletion differ to provide a fuller picture of TRA2 redundancy in the CWR22Rv1 model of CRPC.

Each of the above experimental arms was performed for three independent biological replicates for a total of 12 samples. RNA was extracted and as for proteomics, material was outsourced for paired-end RNA sequencing with GENEWIZ™. This returned ~30-40 million 2 x 150 bp RNA-Seq reads per sample in the form of raw FASTQ sequencing files for analysis.

First, RNA-Seq library quality was assessed using *FastQC* (Andrews, 2010) and a quality report for all samples was compiled with *MultiQC* (Ewels *et al.*, 2016). *FastQC* assesses a multitude of quality metrics including confidence of a correctly assigned base across the read, GC content and the presence of any significantly overrepresented sequences that could indicate contamination (Andrews, 2010). This revealed overall good quality sequencing based on these metrics, with just one QC failure (Figure 6.16, left panel) indicating adapter contamination. Note that the ‘sequence duplication’ QC score (Figure 6.16, fourth column) is expected to fail for RNA-Seq, due to the presence of highly expressed genes and deduplication is not recommended (Parekh *et al.*, 2016).

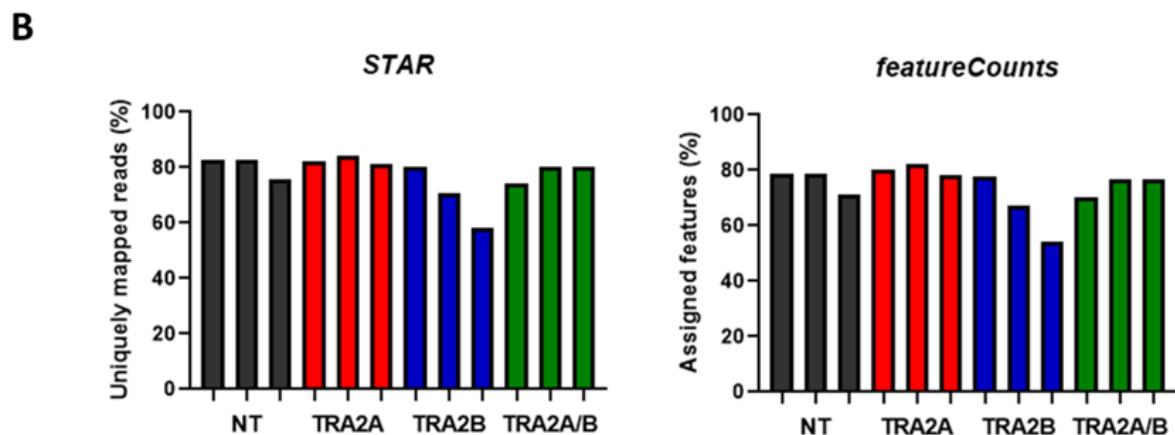
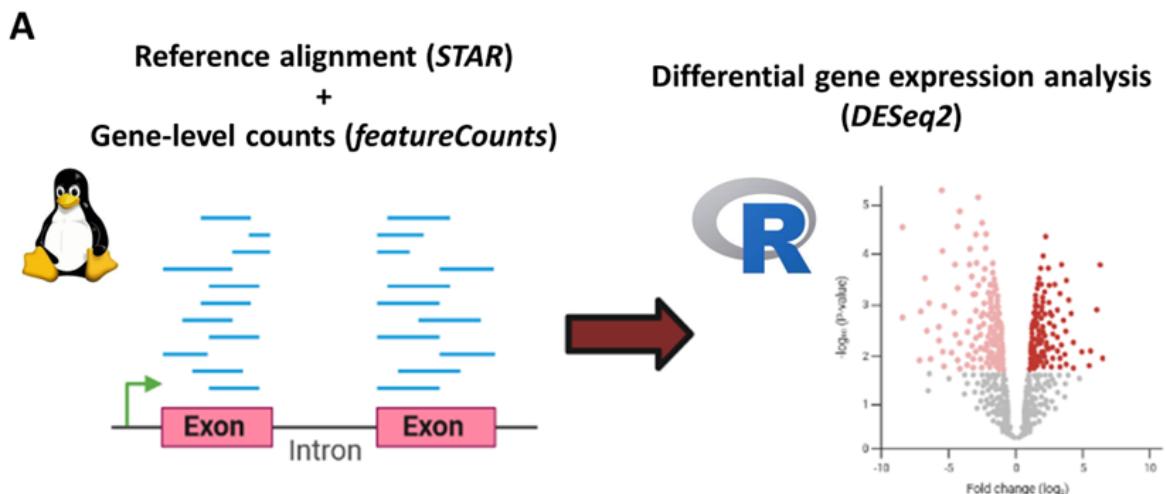
Adapter contamination transpired to be due to the Illumina Universal sequencing adapter: 5' - AGATCGGAAGAG - 3'. These sequences were removed using *Cutadapt* (Martin, 2011), as adapter removal can enhance the mapping efficiency during RNA-Seq alignment (Zhou and Rokas, 2014). *Cutadapt* successfully removed adapter sequences (Figure 6.16, right panel), enabling analysis to proceed to the next stages.



**Figure 6.16 - Adapter trimming with *Cutadapt* successfully eliminates Illumina adapter contamination**

RNA-Seq library QC was compared before (left) and after (right) adapter trimming with *Cutadapt*. Note that sequence duplication (fourth column) is expected to fail QC for RNA-Seq due to the presence of highly expressed genes. Top panel is *MultiQC* summary for all samples, bottom panel is example adapter content plot from *FastQC* showing the presence (left) and absence (right) of adapter sequences (red line) at the 3' end of RNA-Seq reads. Green = passed QC, amber = passed with warnings, red = QC failure

In order to ascertain global gene expression changes resulting from TRA2 depletion, differential gene expression analysis (DGEA) was performed. RNA-Seq reads were aligned using *STAR* (Dobin *et al.*, 2013) to an hg38 human reference genome, sourced from the GENCODE project (Harrow *et al.*, 2012). Aligned reads were subsequently assigned to exons to derive gene-level expression counts using *featureCounts* (Liao, Smyth and Shi, 2014), before the resulting counts matrix was used for DGEA between samples with the R Bioconductor package *DESeq2*. *DESeq2* uses sophisticated approaches based on a background of unchanging genes to normalise for sequencing library depth and composition, enabling between-samples comparisons (Love, Huber and Anders, 2014) (Figure 6.17A). *STAR* and *featureCounts* successfully aligned reads and assigned genes for the majority of reads (Figure 6.17B), giving at least 16 million reads per sample for DGEA, a more than adequate amount for well-annotated genomes (Liu, Zhou and White, 2014).



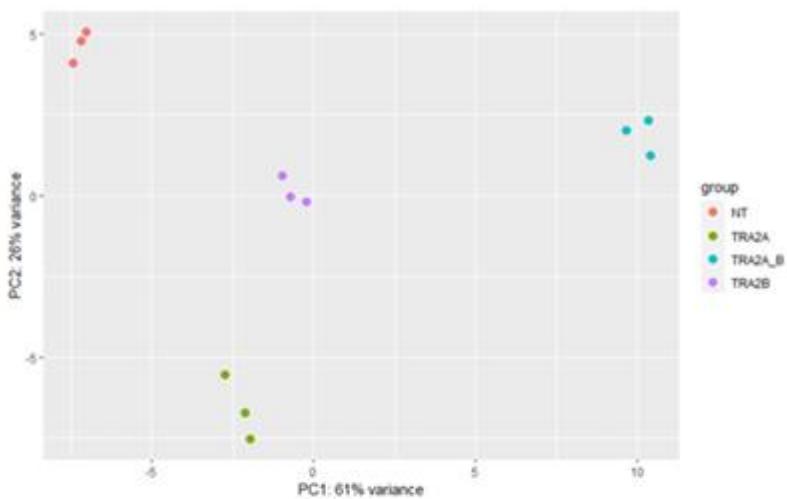
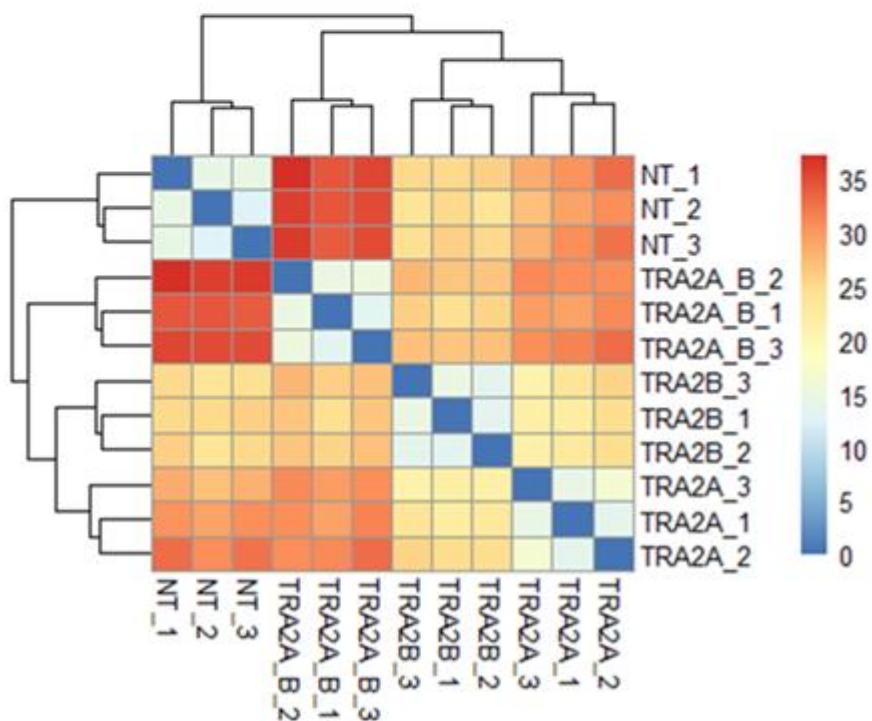
**Figure 6.17 - A computational pipeline enables differential gene expression analysis of RNA-Seq data**

**A.** After passing QC, FASTQ read files were aligned to a reference genome with *STAR*, before gene-level counts were generated with *featureCounts*. Both these steps were performed using Linux software packages. The resultant matrix of gene counts was subsequently used as input for R Bioconductor package *DESeq2* for DGEA. Figure created using biorender.com **B.** *STAR* and *featureCounts* successfully aligned reads and assigned genes for the majority of reads, providing > 16 million reads per sample for *DESeq2* DGEA

After library normalisation, samples were visualised by principal component analysis (PCA) of normalised gene counts to assess intra-arm similarity and acquire insight into inter-arm differences (Groth *et al.*, 2013). The resulting PCA plot highlighted that samples clustered well between replicates of the same arm, whilst clear differences were apparent between target-specific siRNA knockdowns (Figure 6.18A). Interestingly, TRA2A and TRA2B knockdown samples had a greater separation based on the 2<sup>nd</sup> principal component (PC2) than the 1<sup>st</sup> principal component (PC1) (Figure 6.18A). As PC1 represents a greater percentage of variability between samples than PC2 (Groth *et al.*, 2013), this indicates that TRA2A and

TRA2B knockdowns are more closely related to each other than to NT or TRA2A/B, from which they were predominantly separated on PC1 (Figure 6.18A). This suggests a potential degree of TRA2 overlap as alluded to previously. Crucially no batch effects were present, providing confidence in DGEA.

As an alternative means of assessing sample similarity, hierarchical clustering was performed. The Euclidean distance, or divergence of expression profiles (Glazko and Mushegian, 2010), between normalised gene counts for each sample was determined and plotted as a sample distance heatmap (Figure 6.18B). This algorithm recapitulated results of PCA analysis in showing that: i) samples cluster more closely within than between experimental arms, and ii) TRA2A and TRA2B depleted samples have more similar expression profiles than any other experimental arm, as reflected by their separation further down the hierarchical clustering tree than any other pairs of samples (Figure 6.18B).

**A****B**

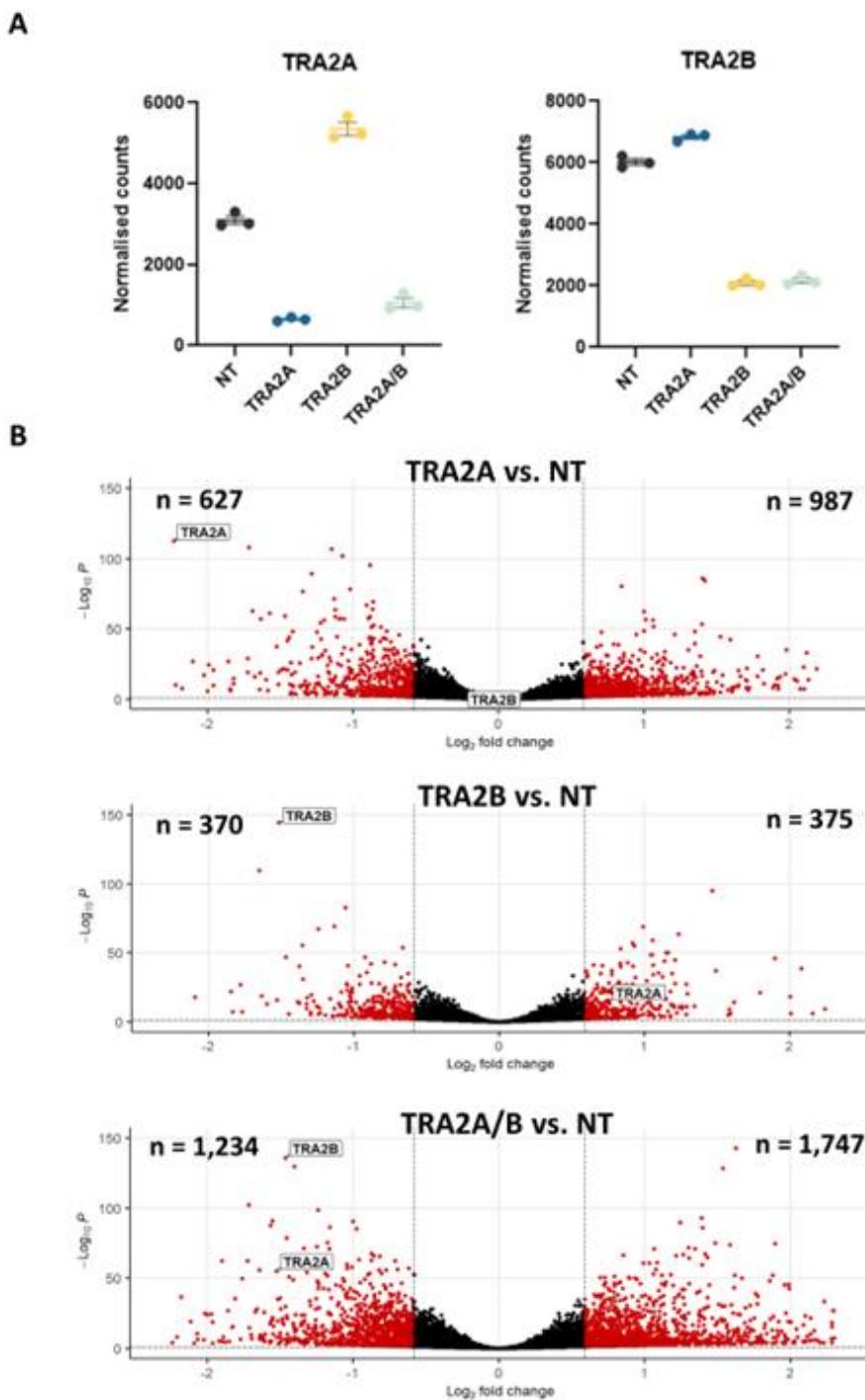
**Figure 6.18 - Principal component analysis and hierarchical clustering demonstrate sample expression similarities and differences between experimental arms**

**A.** Principal component analysis (PCA) of *DESeq2*-normalised gene counts **B.** Heatmap of hierarchical clustering determined by Euclidean distance between normalised gene counts for each sample. Colour scale denotes count profile divergence between sample pairings from perfect overlap (blue) to maximum divergence (red)

DGEA was subsequently performed using *DESeq2*. Although RT-qPCR validation of successful siRNA-mediated target depletion in the RNA-Seq samples prior to sample submission had

already been performed (data not shown), confirmation was sought by RNA-Seq. RNA-Seq expression profiling of TRA2 revealed a strikingly similar pattern to that seen by qPCR and western blot experiments, with TRA2B depletion causing a nearly twofold increase in TRA2A counts (Figure 6.19A). Conversely, TRA2A knockdown does not enact the same effect on TRA2B levels (Figure 6.19A).

Volcano plots were created to summarise all fold changes and FDRs (*DESeq2* Benjamini-Hochberg adjusted p-values) of differentially expressed genes (DEGs) across each experimental contrast. These reiterate TRA2A or TRA2B as amongst the most downregulated genes in their respective siRNA treatments (Figure 6.19B). Moreover, summation of total positively or negatively DEGs at a cut-off of FDR  $< 0.05$  and linear fold change  $+/- 1.5$  shows that combined TRA2A/B knockdown generates an appreciably greater number of DEGs than the total of lone TRA2A and TRA2B depletions (Figure 6.19B). It is particularly interesting that TRA2B knockdown results in significantly fewer DEGs than TRA2A, as although reductions in TRA2B mediate significant TRA2A elevation, the inverse only occurs to a minimal extent. This potentially implies that TRA2B depletion is more effectively buffered by augmented TRA2A, resulting in fewer total DEGs.

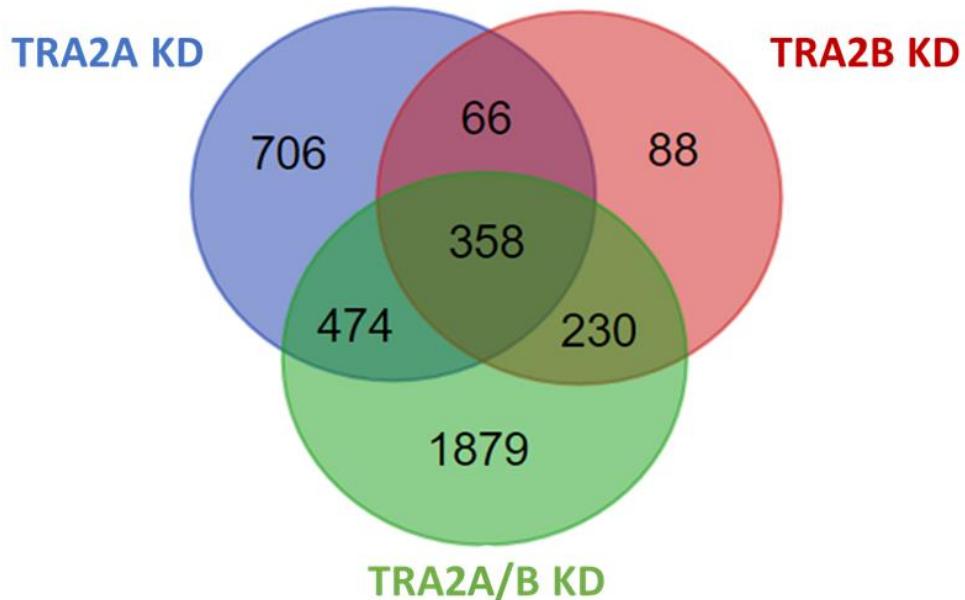


**Figure 6.19 - TRA2 paralogs are efficiently depleted by siRNAs and combined depletion results in significantly greater differential expression**

A. Summary of RNA-Seq gene expression counts, normalised by *DESeq2*, for *TRA2A* and *TRA2B* in each sample  
 B. Volcano plot of DEGs resulting from each respective *TRA2* siRNA vs NT DGEA. *TRA2A* and *TRA2B* are highlighted in boxes. Cutoffs for significantly DEGs (points in red) are FDR < 0.05 and linear fold change +/- 1.5.

In order to more closely examine this, the overlap in identity of DEGs at FDR < 0.05 and linear fold change +/- 1.5 between experimental contrasts was ascertained. More than 700 genes

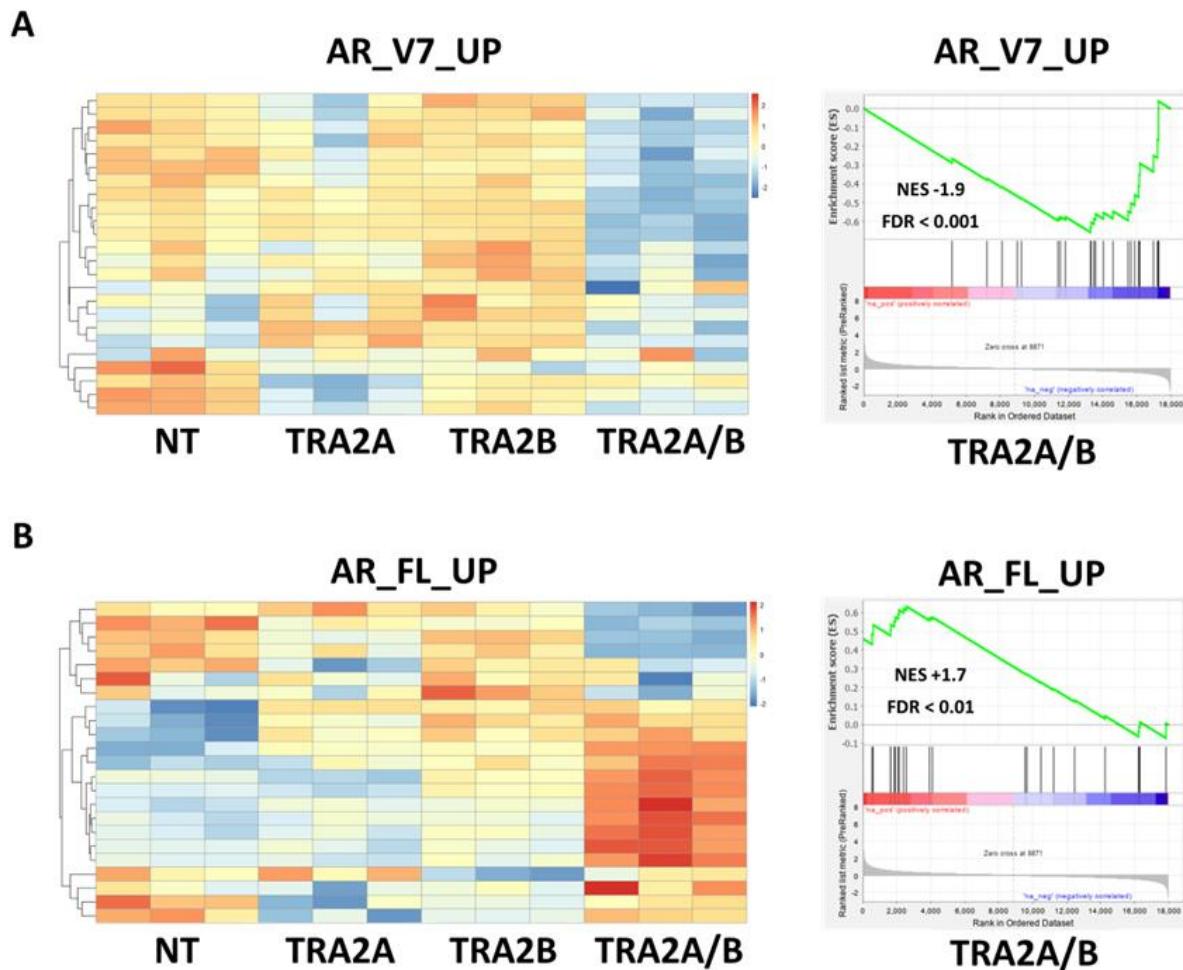
were uniquely altered by lone TRA2A knockdown, compared to just 88 for TRA2B (Figure 6.20). This may suggest the aforementioned TRA2A compensation is quashing the effects of TRA2B depletion, although it is impossible to answer this definitively as it may simply be that TRA2B has a narrower range of targets in CWR22Rv1. Additionally, the TRA2A siRNA used appears to have a greater ability to diminish its target than TRA2B siRNA (Figure 6.11, 6.13). Notably, nearly 1,900 genes were significantly differentially expressed by combined TRA2A/B depletion (Figure 6.20). This large number of unique DEGs alludes to a degree of functional overlap between TRA2A and TRA2B. In theory, combined knockdown of two entirely functionally unrelated genes would return the sum of DEGs resulting from each individual knockdown, whereas combined TRA2 depletion resulting in a number of unique DEG identities far in excess of their individual knockdowns would suggest intersecting roles. Nevertheless, many DEGs unique to single TRA2 siRNA treatments were observed. Therefore as expected given their 75% amino acid homology, TRA2 paralogs are likely to regulate both unique and shared RNA targets, of which AR is putatively the latter.



**Figure 6.20 - Overlap analysis of TRA2 knockdown significantly DEGs alludes to both redundant and unique expression changes**

The identities of genes differentially expressed at FDR < 0.05 and linear fold change +/- 1.5 for each TRA2 siRNA vs. NT contrast were extracted. Overlap between genes in each experiment contrast was determined

Regardless of global transcriptomic effects resulting from TRA2 paralog knockdown, RNA-Seq data will provide greater insight into the effects of TRA2 depletion on this project's primary goal; validation of a novel AR-V splicing factor. Given preliminary indications from qPCR and western blot analysis that TRA2A/B depletion in CWR22Rv1 is mediating a splicing switch from AR-V7 to AR-FL, readouts of respective AR protein activity were analysed. Here, 25-gene expression signatures AR\_V7\_UP and AR\_FL\_UP (Hu *et al.*, 2012), previously used in proteomics target selection (Figure 6.2), were profiled across DGEA results. As expected, a signalling shift is apparent whereby combined TRA2A/B depletion diminishes AR-V7 transactivation (Figure 6.21A) whilst augmenting AR-FL activity (Figure 6.21B). GSEA analysis confirmed significant negative and positive enrichments in TRA2A/B knockdowns for AR\_V7\_UP and AR\_FL\_UP, respectively (Figure 6.21A, B).

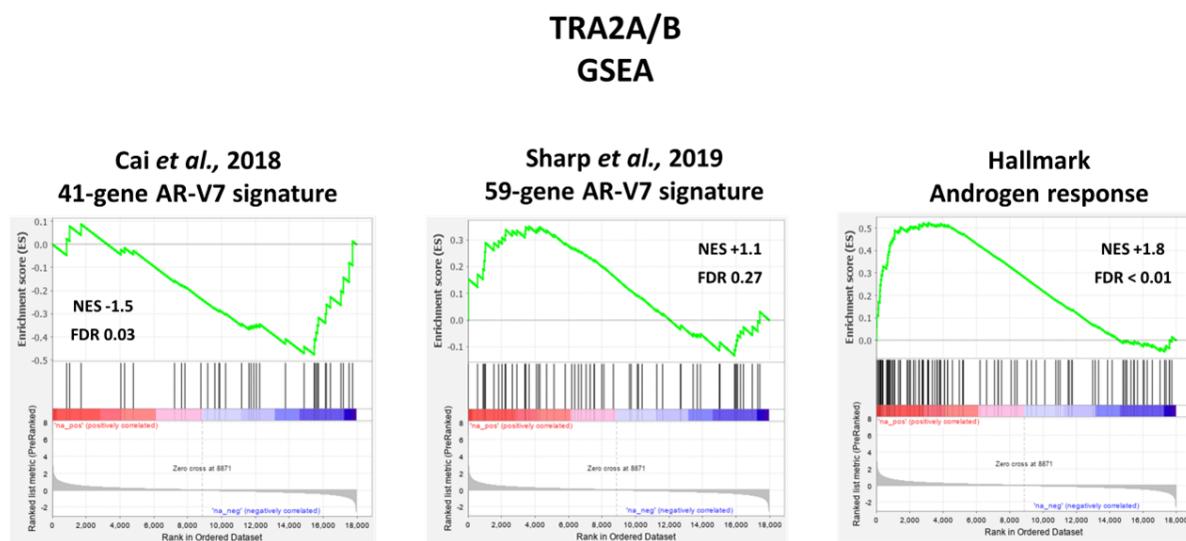


**Figure 6.21 - Combined TRA2A/B depletion drives a switch from AR-V7 to AR-FL signalling in CWR22Rv1**

**A.** Normalised RNA-Seq counts were filtered for genes found in the AR\_V7\_UP 25-gene signature (Hu *et al.*, 2012). Heatmap displays z-score scaled normalised counts for each gene in this signature across each sample.

Plot on the right is GSEA of AR\_V7\_UP for the TRA2A/B vs. NT experimental contrast **B**. The same analysis was performed as in (A), using the AR\_FL\_UP signature. For heatmaps, red and blue indicate positive and negative z-scores, respectively. For GSEA, NES = normalised enrichment score, FDR = GSEA false discovery rate

The 41-gene and 59-gene AR-V7 signatures (Cai *et al.*, 2018; Sharp *et al.*, 2019) were also leveraged for GSEA. As expected, the 41-gene signature, derived from AR-V7 knockdown in CWR22Rv1 before being confirmed in clinical samples (Cai *et al.*, 2018), was significantly negatively enriched in TRA2A/B knockdown samples (Figure 6.22, left panel). Conversely, the 59-gene signature that associates with nuclear AR-V7 protein expression (Sharp *et al.*, 2019) was positively enriched although this was not statistically significant (Figure 6.22, middle panel). The difference in derivation of these gene sets may explain this discrepancy. The authors of the latter signature emphasise in their study that this does not directly encompass an AR-V7 cistrome, rather it represents unifying gene expression characteristics of AR-V7 positive disease (Sharp *et al.*, 2019). It may be that certain genes found in this signature select for AR-V7 expression, potentially by better enabling AR-V7 transactivation rather than being direct expression targets of AR-V7. For example *HOXB13* found in this gene set is a known pioneer factor that facilitates AR-V7 binding to open chromatin (Chen *et al.*, 2018). Finally as predicted, the Hallmark Androgen Response is significantly upregulated by combined TRA2 depletion further indicating potentiation of AR-FL splicing (Figure 6.22, right panel).



**Figure 6.22 - Additional gene set enrichments further demonstrate an AR splicing switch**

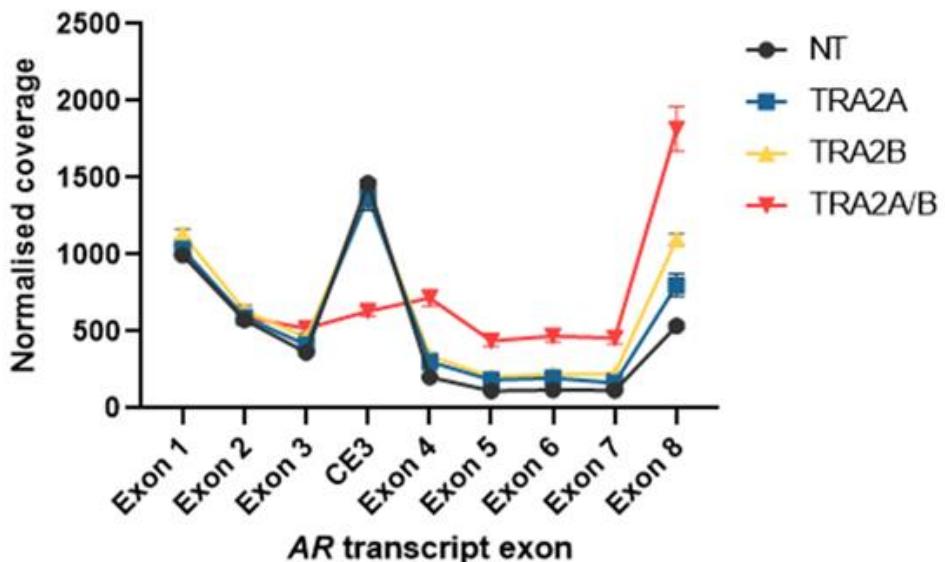
GSEA of published 41-gene and 59-gene AR-V7 expression signatures was performed for DGEA results of TRA2A/B knockdown. GSEA was also performed for the Hallmark Androgen Response

In addition to DGEA, RNA-Seq aligned reads were used to gain a greater understanding of differential *AR* exon splicing resulting from TRA2 knockdowns. qPCR and western blot data indicate that TRA2 depletion is specifically impacting *AR* splicing rather than transcription, and a more detailed examination of RNA-Seq alignments will confirm this. If TRA2 is regulating splicing decisions beyond exon 3, then the first three *AR* exons before this junction should be unaffected. The *JunctionSeq* R Bioconductor package (Hartley and Mullikin, 2016) was used to overlap read alignment Binary Alignment Map (BAM) files with annotated exons and subsequently calculate normalised differential exon usage (DEU) across experimental arms. *JunctionSeq* implements similar normalisation algorithms to *DESeq2* to enable cross-sample comparison of exonic reads (Hartley and Mullikin, 2016).

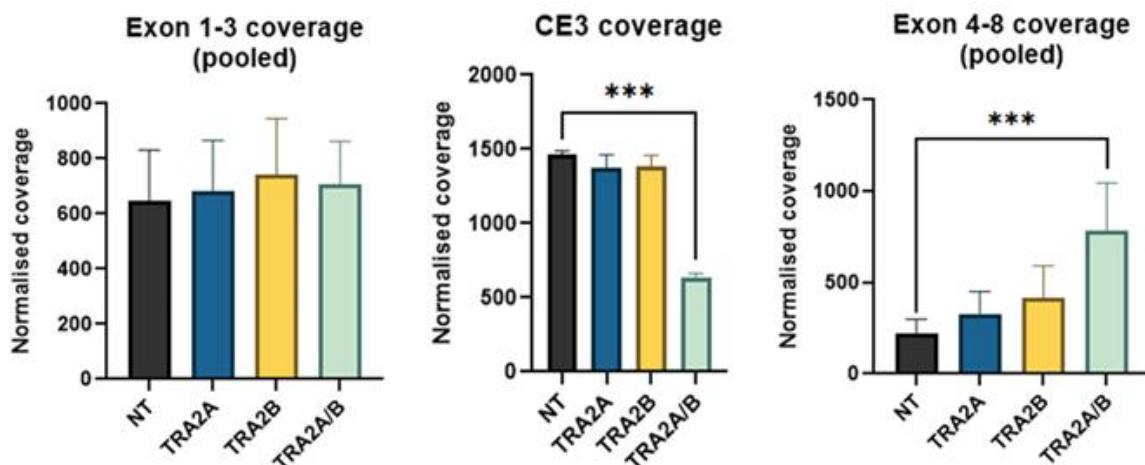
DEU analysis demonstrated striking changes to exon read coverage across the *AR* transcript between experimental arms. Plotting of *JunctionSeq*-normalised RNA-Seq reads across *AR* exons showed a significant decrease in CE3 reads in CWR22Rv1 that had undergone dual TRA2 depletion, which was accompanied by the anticipated increase in exons 4-8 (Figure 6.23A). Crucially, mean coverage of exons 1-3 was unchanged and differences only became apparent beyond this transcript position, vindicating assumptions that TRA2A/B knockdown is altering *AR* splicing rather than gene transcription (Figure 6.23B). This also suggests other facets that may impact expression such as mRNA stability are unlikely, as if AR-V7 transcripts were degraded then exons 1-3 would concomitantly decrease as well.

A

### JunctionSeq AR Differential Exon Usage



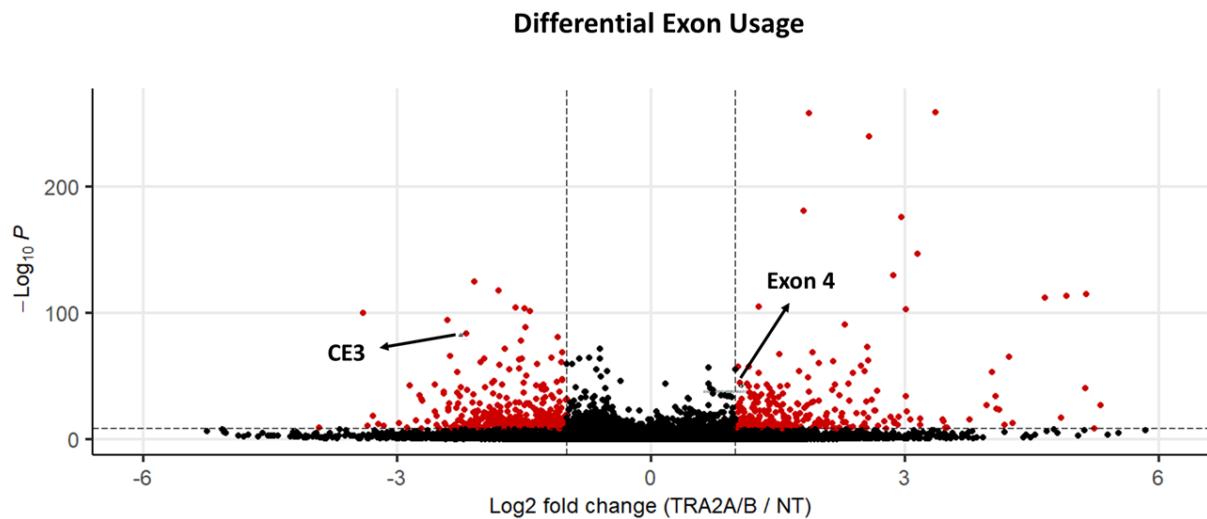
B



**Figure 6.23 - Differential exon usage analysis of RNA-Seq reads confirms a TRA2 control of an AR splicing switch**

A. Differential exon usage (DEU) analysis of RNA-Seq BAM files was performed using *JunctionSeq*. Mean normalised exon counts for each siRNA treatment are plotted  $\pm$  SEM for the AR exons indicated. B. Mean *JunctionSeq*-normalised exon counts for AR exons 1-3 and 4-8 were pooled. Mean normalised exon counts  $\pm$  SEM are plotted across samples for the indicated exons. Significance asterisks denote *JunctionSeq* FDR of DEU contrasts between the indicated samples (\*\* = FDR  $< 0.00001$ ). Not all samples have significance denoted for ease of visualisation

Quite remarkably, global analysis of *JunctionSeq*-calculated DEU alterations revealed that CE3 and exon 4 were amongst some of the most differentially utilised exons across the entire annotated transcriptome (Figure 6.24). Precisely, of all exons analysed by *JunctionSeq* CE3 is the 11<sup>th</sup> most negatively differentially used exon upon TRA2A/B depletion. This further validates that TRA2B and its paralog TRA2A are critical components of AR-V7 generation in this model of CRPC.



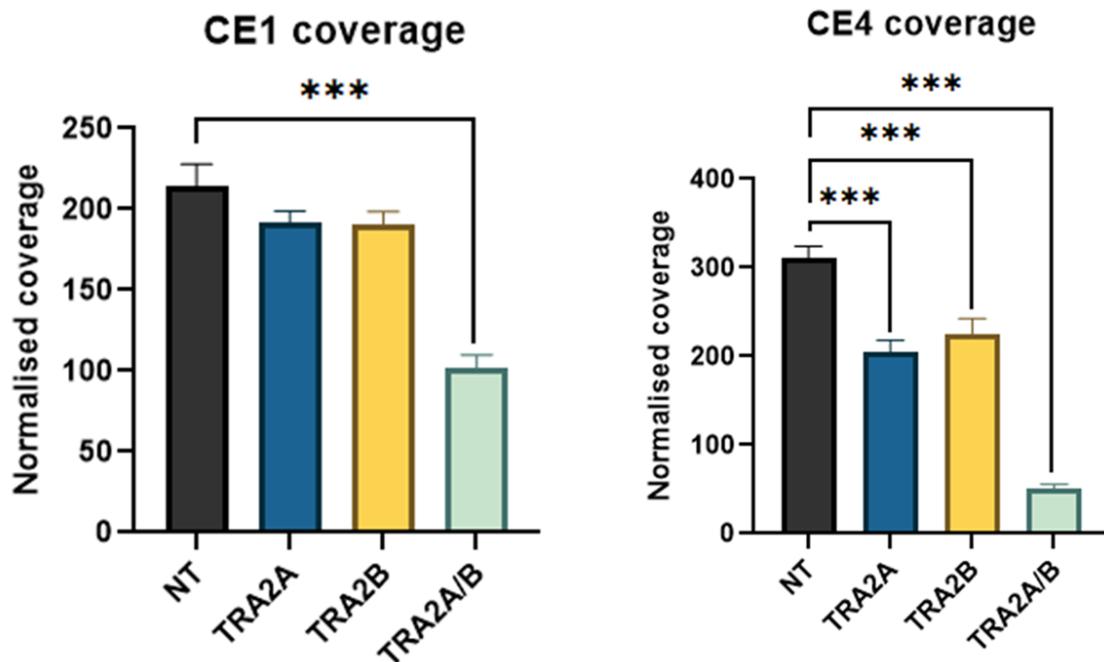
**Figure 6.24 - AR CE3 and exon 4 are amongst the most differentially utilised exons in the annotated transcriptome upon TRA2A/B depletion in CWR22Rv1**

Volcano plot illustrates  $-\log_{10}$  FDR (y-axis) vs log2 fold change (x-axis) of DEU in TRA2A/B vs NT-transfected samples, as calculated by *JunctionSeq*. Only exons differentially used between samples at an FDR < 0.01 are included. Exons highlighted in red meet a cut-off of linear fold change  $\pm 2$  and FDR < 0.00000001. AR CE3 and exon 4 are highlighted

Finally, DEU results were used to examine coverage of other cryptic exons (CEs) that give rise to alternative AR-Vs in CRPC. Although AR-V7 is the most well studied and clinically validated AR-V, other variants are observed in CRPC and analysis of their expression will provide information on TRA2 control across all AR-Vs. CE1, which is spliced into AR-V1 and has been associated with CRPC and abiraterone resistance (Hu *et al.*, 2009; Stuopelyte *et al.*, 2020), also exhibited significant negative differential usage upon combined TRA2 knockdown (Figure 6.25).

CE4 is included in transcripts such as AR-V3, another AR-V linked with emergence of CRPC (Kallio *et al.*; Wüstmann *et al.*, 2023). Similar to other CEs, TRA2A/B depletion significantly diminished CE4 inclusion in transcripts (Figure 6.25). Interestingly, for this cryptic exon lone

TRA2 depletion also mediated reductions in transcript usage (Figure 6.25). This DEU data is consistent with western blot observations that TRA2A/B knockdown significantly depletes AR-Vs as detected by an AR NTD domain antibody, which is reactive to all AR-Vs (Figure 6.11A). As proteomics experiments were targeted at CE3 only, it is interesting that other CEs also exhibit reduced coverage when TRA2 paralogs are depleted. This implies that TRA2 plays a role in *AR* splicing and CE inclusion across multiple AR-Vs.



**Figure 6.25 - Alternative cryptic exon inclusion is also regulated by TRA2**

*JunctionSeq* DEU analysis results for cryptic exons CE1 and CE4, included in AR-Vs such as AR-V1 and AR-V3, are shown as normalised exon counts  $\pm$  SEM for each sample. Significance values denote *JunctionSeq* FDR (\*\*\* = FDR  $< 0.00001$ ). Only results significant at  $\alpha 0.001$  or lower have significance denoted

## 6.6 TRA2 knockdown elicits contrasting proliferative responses in CWR22Rv1 and VCaP

In order to assess a potential therapeutic index of targeting TRA2 as a means to modulate pathogenic splicing, proliferation assays were performed to evaluate toxicity of TRA2 paralog depletion. Furthermore, TRA2 depletion may sensitise cells to NSAA enzalutamide by driving cell signalling away from undruggable AR-V7 back towards a dependency on AR-FL.

CWR22Rv1 were transfected with the same TRA2 siRNAs as previously utilised in media containing 10% FBS, incubated for 24 hours and treated with either 10  $\mu$ M enzalutamide (Enz) or DMSO control. Cells were then fixed at days 0, 2, 4 and 6 and growth over time was measured by sulforhodamine B (SRB) assay. As expected, CWR22Rv1 are highly resistant to enzalutamide, with no apparent difference in proliferation between Enz and DMSO treatment arms observed (Figure 6.26A). First, toxicity of TRA2 depletion in the absence of Enz was established. This revealed that CWR22Rv1 in fact proliferated more effectively upon single TRA2 paralog depletion (Figure 6.26B), whereas cells underwent  $\sim$ 50% reduction in growth with dual TRA2A/B depletion (Figure 6.26B).

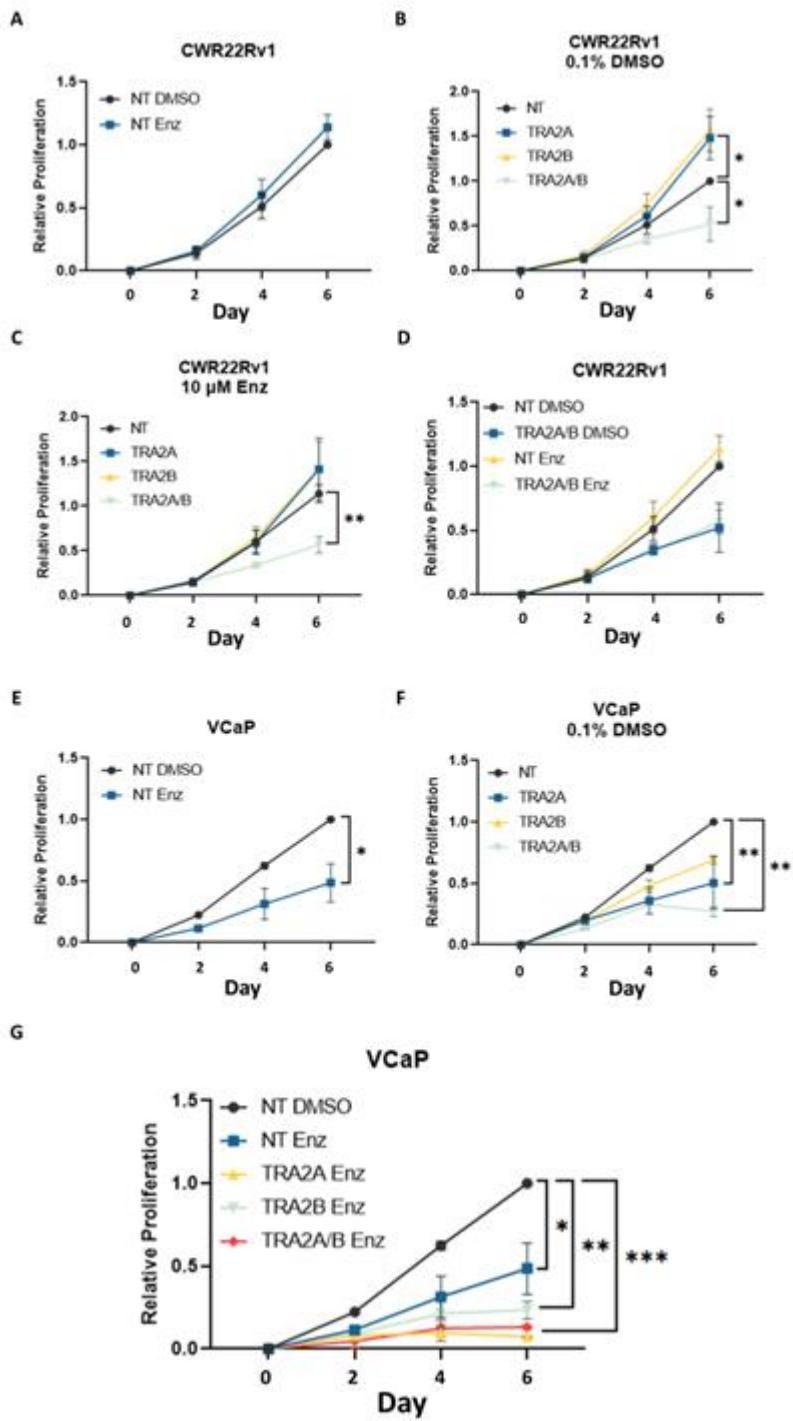
Similar results were observed in CWR22Rv1 treated with 10  $\mu$ M Enz (Figure 6.26C). Although statistically cell growth was inhibited to a greater degree, comparison of NT and TRA2A/B arms demonstrates that in actuality this difference in significance is due to lower data variability in Enz-treated samples, as a negligible difference was seen in effect size (Figure 6.26D). Therefore, TRA2A/B depletion does not appear to sensitise cells to enzalutamide, at least not at a 10  $\mu$ M concentration. The fact that single TRA2 knockdown enhances rather than retards proliferation may allude to the reciprocal upregulation observed in CWR22Rv1, as dual depletion inhibits growth whether in the presence or absence of NSAA enzalutamide.

The same experimental setup was performed in VCaP. VCaP are a more enzalutamide-sensitive PCa line than CWR22Rv1, as reflected by  $\sim$ 50% reductions in growth after 6 days treatment with 10  $\mu$ M Enz (Figure 6.26E). Unlike VCaP, prior experience in the host laboratory has shown that CWR22Rv1 do not modulate their AR-V7 production in response to antiandrogen treatment. Rather, they express constitutively high AR-V7 levels. Therefore, previous demonstration that VCaP cells augment AR-V7 levels in response to enzalutamide (Figure 6.13) meant it was of great interest to test this cell line, to see whether suppressing

this increase in AR-V7 through TRA2 depletion (Figure 6.13) may lead to even greater enzalutamide sensitivity. Interestingly, TRA2 knockdown in VCaP in the absence of Enz revealed a key difference to results in CWR22Rv1 in that lone TRA2A or TRA2B depletion impacted growth, though to a lesser extent than combined TRA2A/B (Figure 6.26F). This is significant due to our previous observation that VCaP do not appear to enact reciprocal TRA2 paralog upregulation upon single knockdown, which may explain this result.

TRA2 knockdown sensitised VCaP to Enz treatment (Figure 6.26G). However, single TRA2A depletion sensitised cells more effectively than TRA2B, and was comparable to combined TRA2 depletion (Figure 6.26G). Previous analyses have shown that lone TRA2A knockdown in VCaP does not quash Enz-mediated AR-V7 elevation as effectively as TRA2B or TRA2A/B (Figure 6.13). Therefore sensitisation may be due to TRA2A depletion modifying other pathways that alter VCaP Enz response independently of AR.

Collective qPCR, western blot and proliferation data demonstrate that VCaP and CWR22Rv1 respond differently to TRA2 knockdown. Though they both undergo reductions in AR-V7, the lack of reciprocal TRA2 upregulation in VCaP and contrasting growth responses suggest that TRA2 paralog behaviour in CRPC is dependent on cellular context. RNA-Seq was not performed in TRA2-depleted VCaP although would indeed be instructive to ascertain the extent of transcriptomic response.



**Figure 6.26 - (previous page) CWR22Rv1 and VCaP cells exhibit contrasting proliferative responses to TRA2 depletion**

**A.** CWR22Rv1 were transfected with non-targeting (NT) siRNA and incubated for 24 hours, before being treated with either 0.1% DMSO or 10  $\mu$ M enzalutamide (Enz, in 0.1% DMSO). Cells were fixed at the indicated timepoints and proliferation was measured by sulforhodamine B (SRB) assay **B.** CWR22Rv1 cells were transfected with the indicated siRNAs, treated with 0.1% DMSO and proliferation was measured as in (A) **C.** CWR22Rv1 cells were treated and assayed as in (B) with 10  $\mu$ M Enz instead of 0.1% DMSO **D.** Proliferation was compared between NT and TRA2A/B siRNA-transfected CWR22Rv1, with and without Enz treatment. Proliferation data is sourced from experiments described in (B) and (C) **E.** VCaP were transfected with NT siRNA, treated with 0.1% DMSO or 10  $\mu$ M Enz and proliferation was measured as in (A) **F.** VCaP were transfected with the indicated siRNAs, treated

with 0.1% DMSO and proliferation was measured as in (A) **G**. Proliferation in NT-transfected, DMSO-treated VCaP cells and each enzalutamide-treated arm was compared. All SRB data was normalised to respective day 0 samples, before being scaled within each biological replicate to the NT/DMSO experimental arm. SRB datapoints were calculated as an average of technical triplicates within each biological replicate. CWR22Rv1 data represents  $n = 3$  biological replicates. VCaP data represents  $n = 2$  biological replicates. Unpaired t-test was used for determination of statistical significance between day 6 datapoints. Only results significant at  $\alpha = 0.05$  or lower have significance denoted (\* =  $p < 0.05$ , \*\* =  $p < 0.01$ , \*\*\* =  $p < 0.001$ )

## 6.7 TRA2 may regulate AR-V splicing at a CE3 splicing enhancer

Evidence from our collective analyses validates TRA2 paralogs as regulators of an *AR* splice switch. Combined TRA2 knockdown results in significant decreases in exon 3/CE3 splice junctions, whereas the exon 3/exon 4 junction is significantly favoured in this context. Furthermore, RNA-Seq DEU analysis of the *AR* gene upon TRA2 depletion provides further evidence for a *bona fide* alteration to splicing events, as it mediates a reduction in cryptic exon usage without affecting preceding exons which would signify a lack of transcriptional suppression.

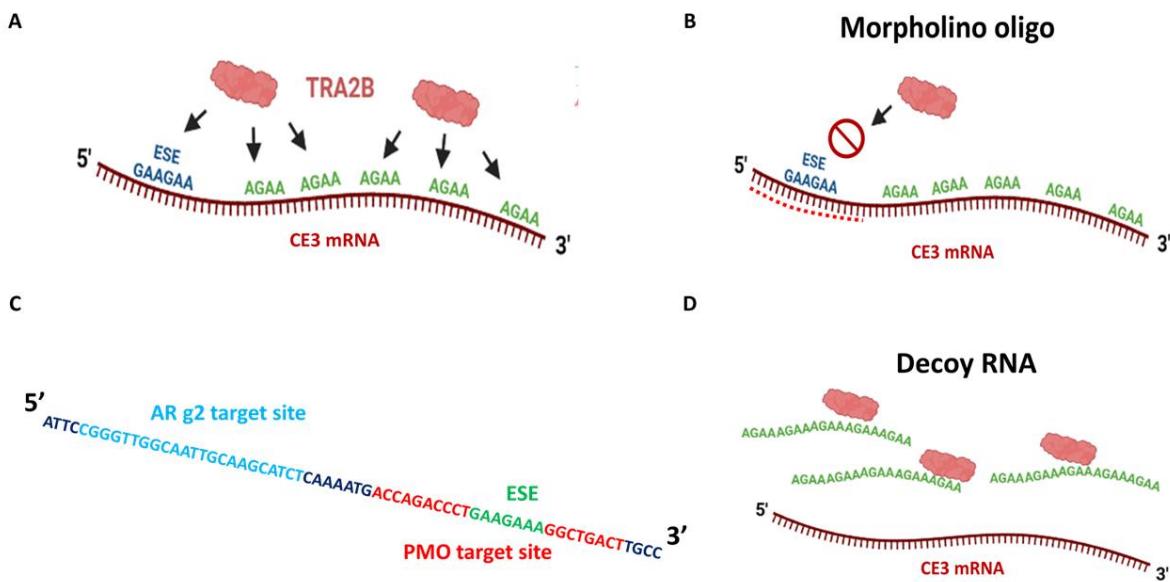
A more detailed analysis of the literature examining TRA2 binding specificities reveals a preference for AGAA RNA motifs (Cléry *et al.*, 2011; Best *et al.*, 2014; Xue, Ma and Zhang, 2023). This tetramer would in theory occur once every 256 nucleotides. A previous study used a cut-off of exons that contain AGAA motifs at 1.5x this rate to select and validate TRA2B binding sites (Storbeck *et al.*, 2014). CE3 and its 3' UTR up to the transcriptional termination site contains 13 such tetramers in a 1,402 nucleotide sequence, representing 2.4x the expected rate of AGAA, therefore CE3 represents an AGAA-rich exon (Figure 6.27A).

Additionally, the GAAGAA hexamer has been noted as a strong exonic splicing enhancer (ESE), and the structure of TRA2B bound to this enhancer has been solved by NMR (Fairbrother *et al.*, 2002; Tsuda *et al.*, 2011). Closer examination of the CE3 mRNA region targeted by AR g2 for dCasRx-APEX2 proteomics uncovered that it is in close proximity to a GAAGAA hexamer found within 70 nucleotides of the 3' splice site, a distance within which ESE activity is optimal (Parmley, Chamary and Hurst, 2006) (Figure 6.27A). Furthermore this region has been identified as an ESE promoting AR-V7 generation, experimental mutation of which abrogates AR-V7 splicing (Liu *et al.*, 2014).

Therefore, two complementary oligonucleotide approaches were used with the aim of modulating these splicing events. The first such approach employed a phosphorodiamidate morpholino oligomer (PMO). PMOs are chemically modified synthetic DNA 25-mers that create a steric blockade, binding complementary RNAs without promoting their degradation to prevent access of other molecules (Figure 6.27B) (Roberts, Langer and Wood, 2020). A PMO was designed, termed CE3 block, which would bind downstream of the AR g2 target region (Figure 6.27C). Should this inhibit AR-V7 splicing, it would provide validation that proteomics

experiments were targeted at a functionally important region whilst also demonstrating an alternative means of splicing modulation.

The second approach was to use decoy RNA oligonucleotides. To date only a single publication has reported on these (Denichenko *et al.*, 2019), whereby RNA oligomers comprising tandem repeats of a known splicing factor binding motif are used to sequester the relevant protein, preventing its association with mRNA targets thus altering splicing patterns (Figure 6.27D). A decoy RNA oligonucleotide was synthesised, termed TRA2 decoy, comprising a 5n repeat of the AGAA tetramer with which the TRA2B RRM specifically forms hydrogen bonds (Cléry *et al.*, 2011; Tsuda *et al.*, 2011). This decoy approach would provide a means to inhibit TRA2 activity without its depletion, specifically hindering its mRNA splicing activity without necessarily impacting other potential functions and protein-protein interactions.



**Figure 6.27 - Oligonucleotide approaches represent a novel means to modulate CRPC splicing activity**

**A.** TRA2 proteins (TRA2B shown in this example) associate with AGAA mRNA motifs. AR CE3 contains a high frequency of these tetramers. Furthermore, the structure of TRA2B bound with the GAAGAA strong exonic splicing enhancer, one of which is located near the CE3 splice site, has been solved **B.** Morpholino oligomers (PMOs) are DNA 25-mers that can stably bind to a complementary RNA region (red dotted line), this blocks access of other molecules including splicing factors without inducing RNA degradation **C.** dCasRx-APEX2 proteomics experiments were directed to the AR g2 target site of CE3 mRNA (turquoise). This is in close proximity to a GAAGAA exonic splicing enhancer (green). Our CE3 block PMO is complementary to a 25-nucleotide sequence (red) that incorporates this splicing enhancer sequence, in theory blocking association of splicing factors that may include TRA2B **D.** Decoy RNA oligonucleotides comprise tandem repeats of RNA motifs recognised by splicing factors, in this instance (AGAA)5n. This may sequester TRA2 away from binding CE3 and mediating its splicing. Figures created using biorender.com

PMOs and decoy RNAs were transfected into CWR22Rv1 and after 48-hours RNA was harvested for RT-qPCR. Synthetic PMOs were 3' conjugated with fluorescein enabling visualisation of transfection, which indicated that methods enabled effective cellular delivery (Figure 6.28A). qPCR analysis of CE3 block-transfected vs control PMO-transfected samples demonstrated a potent and specific reduction in AR-V7 levels (Figure 6.28B). PMO experiments were repeated in VCaP, in which similar results were observed (Figure 6.28C).

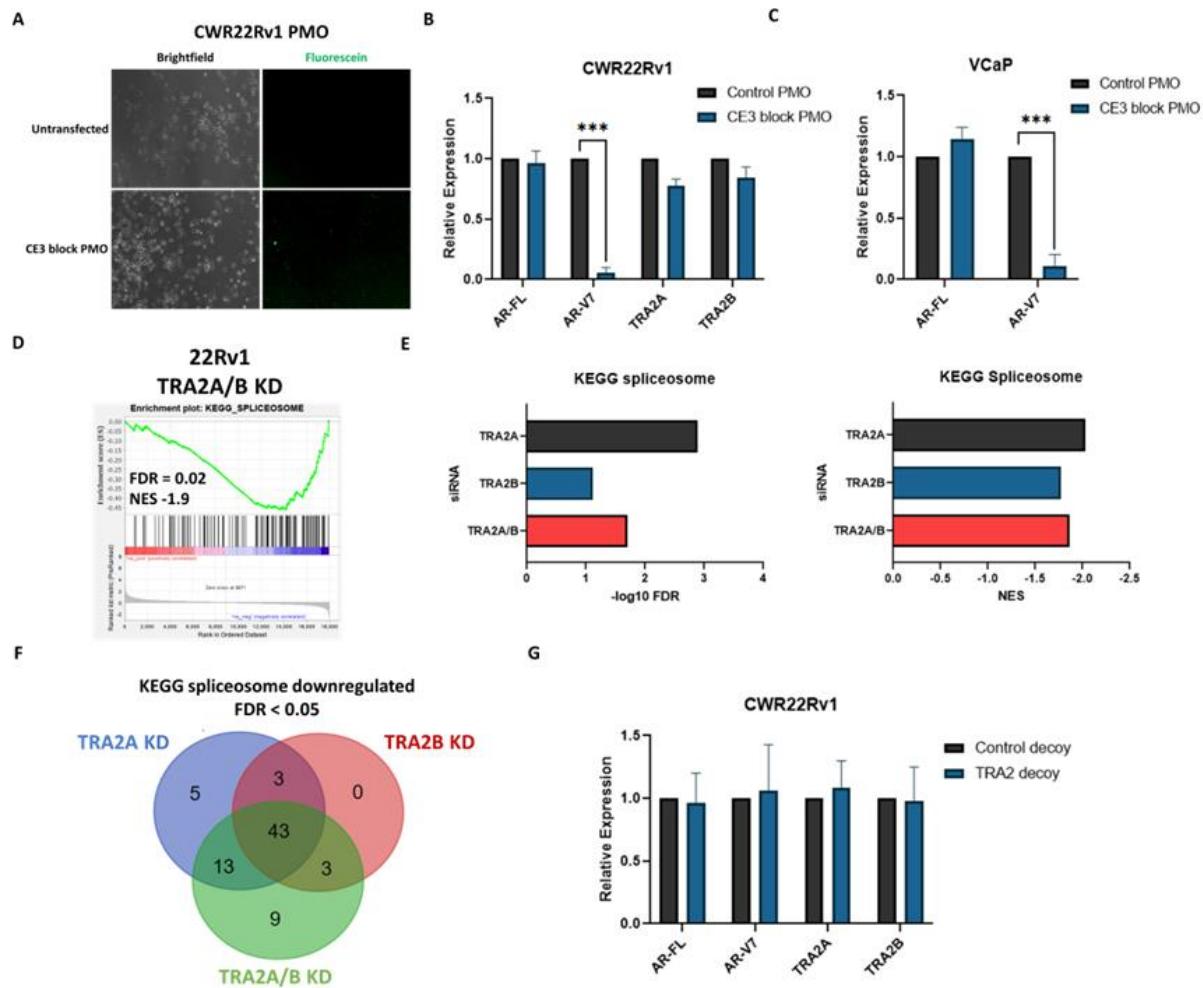
Interestingly, we did not observe a parallel increase in AR-FL splicing in either cell line (Figure 6.28B, C), alluding to mechanistic differences between steric blockade of the splicing enhancer and siRNA depletion of TRA2. This would largely be expected as PMOs will not impact TRA2 protein-protein interactions or interaction with other target splice sites. Additionally, re-examination of DGEA revealed that numerous other splicing factors' expression is decreased upon combined TRA2 depletion as reflected by a significant negative enrichment for the KEGG spliceosome gene set (Figure 6.28D). Therefore TRA2 may also further modify AR splicing by controlling expression of other splicing regulators.

This raises a pertinent question as to whether the AR splicing changes seen upon TRA2A/B knockdown are specific to TRA2 activity at CE3 or merely a reflection of other gene expression changes that more generally alter the spliceosome. However, we have only observed significant changes to AR splicing and a shift from AR-V7 to AR-FL in CWR22Rv1 when both TRA2 paralogs are depleted, and it transpires that similar changes to the KEGG spliceosome gene set are seen whether single or dual TRA2 reduction is enacted (Figure 6.28E). RNA-Seq DGEA was filtered for genes found in the KEGG spliceosome set that were downregulated at an FDR < 0.05 in each experimental arm. Overlap analysis demonstrated that just 9 downregulated genes were unique to dual TRA2A/B depletion (Figure 6.28F).

Of these 9, just three encoded proteins were enriched by our CE3-targeted proteomics approach. SNRNP70, SNRPB and SRSF6. SNRNP70 and SNRPB are components of the core spliceosome thus unlikely to have selectivity for AR-V7 generation (Nikolaou *et al.*, 2022; Lynch). A recent preprint study showed that although SRSF6 expression is associated with CRPC, siRNA depletion had no impact on expression of AR-FL or AR-V7 (Jimenez-Vacas *et al.*, 2023). Finally, of the remaining genes from these 9, only LSM4 and LSM5 significantly correlated with AR-V7 levels in the SU2C/PCF mCRPC cohort leveraged previously (Abida *et al.*, 2019). Both Lsm4 and Lsm5 are also core spliceosomal protein components as part of the

U4/U6.U5 tri-snRNP (Wu and Jiang *et al.*, 2012). Therefore it is likely that the effects of TRA2 depletion are specific to its activity at CE3, and based on binding specificities it is probable it enacts this activity by binding an exonic splicing enhancer which is covered by the CE3 block PMO.

The TRA2 decoy RNA oligonucleotide had no effect on any transcripts assayed (Figure 6.28G). We did not assess cell penetration, however these were delivered by the same means as previous gRNA transfections which have shown high efficiency. In addition, RNAs were 2'-O-methyl modified for stability and incubated for the same timeframe as previously shown to be effective (Denichenko *et al.*, 2019). Given that reciprocal upregulation of TRA2A upon TRA2B depletion arises from reductions in the latter's splicing of the former's poison exon (Best *et al.*, 2014), the fact that both *AR* splicing and TRA2A levels were unchanged would imply the decoy RNA oligonucleotides did not function as intended. As our choice of decoy would in theory bind both TRA2 paralogs, it may be that a significantly greater concentration is required to enable effective sequestration. Alternatively, a different sequence such as tandem repeats of the full GAAGAA ESE may have proven effective, or even use of a 'synthetic CE3' RNA oligo sequence to redirect TRA2 interactions. Use of fluorescently tagged oligos would at the very least enable confirmation of cellular delivery should this approach be investigated further.



**Figure 6.28 - Transfection with CE3 block PMO effectively depletes AR-V7, but not AR-FL, levels to confirm functional relevance of the target mRNA region**

**A.** Live-cell fluorescence imaging was used to confirm that the CE3 block PMO can be effectively transfected into CWR22Rv1. **B.** CWR22Rv1 were transfected with 10  $\mu$ M CE3 block or control PMO. 48 hours later, RNA was harvested and RT-qPCR was used to analyse expression of the indicated mRNAs. qPCR data comprises  $n = 3$  independent biological replicates, plotted as mean  $\pm$  SEM. Unpaired t-test was used for determination of statistical significance ( $*** = p < 0.001$ ). **C.** The experimental setup described in (B) was repeated in VCaP cells. RT-qPCR was used to analyse levels of AR-FL and AR-V7 transcripts. qPCR data comprises  $n = 3$  independent biological replicates, plotted as mean  $\pm$  SEM. Unpaired t-test was used for determination of statistical significance ( $*** = p < 0.001$ ). **D.** GSEA result for the KEGG spliceosome gene set in CWR22Rv1 transfected with TRA2A/B siRNA (NES = normalised enrichment score, FDR = GSEA false discovery rate). **E.** Plot of GSEA  $-\log_{10}$  FDR and normalised enrichment score (NES) for the KEGG spliceosome gene set in each of the indicated RNA-Seq experimental arms. **F.** DGEA results for each of the indicated RNA-Seq experimental arms were filtered for negatively regulated genes (negative log<sub>2</sub> fold change and FDR < 0.05) found in the KEGG spliceosome gene set. The overlap in gene identities between experimental arms was determined. **G.** CWR22Rv1 were transfected with 5  $\mu$ M TRA2 decoy or control decoy RNA oligonucleotides. 48 hours later, RNA was harvested and RT-qPCR was used to analyse expression of the indicated mRNAs. qPCR data comprises  $n = 2$  independent biological replicates, plotted as mean  $\pm$  SEM.

## 6.8 Discussion

Analyses presented in this chapter have culminated in a proof of concept that CE3-targeted dCasRx-APEX2 proteomics methods can be effectively employed for novel splicing factor identification. Proteomics enrichment results, presented in Chapter 5, were cross-referenced with clinical PCa datasets to prioritise selection of candidate AR-V7 splicing regulators (Figure 6.5). Examination of these cohorts presented TRA2B as a lead candidate based on its association with AR-V7 expression and signalling, as well as being associated with poor survival outcomes.

Subsequent experimental validation revealed that TRA2B depletion fails to impact AR-V7 splicing in CWR22Rv1. However, combined knockdown of TRA2B and its paralog TRA2A resulted in dramatic splicing alterations whereby AR-V7 underwent a significant decrease and AR-FL levels were drastically enhanced (Figure 6.10, 6.11). Reciprocal regulation between TRA2 paralogs was also observed in CWR22Rv1, which saw TRA2A levels increase considerably upon knockdown of TRA2B (Figure 6.10, 6.11). This echoed findings from previous work on TRA2 performed in breast cancer cells, in which depletion of both paralogs was required to elicit significant splicing changes in a range of transcripts (Best *et al.*, 2014).

Interestingly, experimental validation performed in a second PCa cell line that expresses AR-V7, VCaP, demonstrated that knockdown of TRA2A or TRA2B alone mediated reductions in AR-V7 generation, although still not to the same extent as combined depletion (Figure 6.12, 6.13). Consistent with this finding, there was no evidence of reciprocal TRA2 paralog upregulation after single TRA2 depletion in VCaP cells (Figure 6.12, 6.13). The reasons for this difference in response to TRA2 knockdown between PCa cell types are not immediately clear. Numerous SR proteins, including TRA2, regulate poison exon inclusion of premature stop codons in a range of other SR family members, controlling their expression (Leclair *et al.*, 2020). Furthermore, the extent of this cross-regulation has been shown to vary considerably across a range of cancer cell lines (Leclair *et al.*, 2020). RNA-Seq analyses, performed here in CWR22Rv1, demonstrated that depletion of either TRA2 paralog significantly downregulated the KEGG spliceosome gene set (Figure 6.28D, E, F). Examination of genes that underwent reduced expression revealed a range of SR splicing factors including *SRSF1*, *SRSF2* and *SRSF6*. Based on our specific analysis of TRA2A and TRA2B, it is apparent that SR protein cross-

regulation functions differently in CWR22V1 than in VCaP cells, and a full appraisal of these differences by carrying out an analogous RNA-Seq experiment in VCaP would be informative.

RNA-Seq DGEA analysis in TRA2-depleted CWR22Rv1 further vindicated that combined TRA2 knockdown results in an *AR* exon splicing switch. AR\_FL\_UP and AR\_V7\_UP gene signatures, representing genes transcriptionally activated by AR-FL and AR-V7, respectively, underwent opposing expression changes upon combined TRA2 paralog depletion. These were consistent with diminished AR-V7, and enhanced AR-FL, levels (Figure 6.21). Moreover, examination of additional AR expression signatures corroborated these observations (Figure 6.22).

In order to confirm that our findings arose from *bona fide* alterations to *AR* splicing as opposed to more general changes in *AR* gene transcription or even mRNA stability, DEU analysis of RNA-Seq data was also performed. Analysis of normalised alignment coverage across exons revealed no measurable changes in *AR* exons 1 - 3 between siRNA treatments (Figure 6.23), implying that transcription and RNA stability are unaffected. Conversely, significant reductions to CE3 were seen upon dual TRA2A/B depletion, whilst exons 4-8 experienced the opposite (Figure 6.23). The distribution of exon coverage changes is exactly as would be expected if TRA2 proteins were controlling a splicing switch beyond exon 3. Furthermore, similar changes to exon usage were seen for AR CE1 and CE4, indicating TRA2 may control splicing of multiple AR-Vs in addition to AR-V7 in CRPC (Figure 6.25). Strikingly, DEU results showed that *AR* CE3 was the 11<sup>th</sup> most differentially negatively used exon upon combined TRA2 depletion across the entire annotated transcriptome (Figure 6.24). This would suggest TRA2 control is particularly relevant to AR-V7 splicing.

As alluded to earlier, based on results in VCaP cells, determination of TRA2 splicing control may be more than merely a function of gene expression. Expression levels of *TRA2B* did correlate with AR-V7 splicing in the clinical cohorts examined in this work, whilst the reciprocal increase of TRA2A upon TRA2B knockdown in CWR22Rv1 would suggest an expression-dependent effect. However, SR protein function is also known to be altered by phosphorylation status (Long *et al.*, 2019). The increased AR-V7 levels in enzalutamide-treated VCaP, which were blocked by TRA2 depletion, cannot apparently be explained by TRA2 expression changes (Figure 6.13). It is possible that post-translational modifications provide an additional layer of TRA2 regulation in CRPC, as TRA2B is known to be phosphorylated by kinases including SRPK1 and CLK1, which modulate its activity (Jamros *et*

*al.*, 2015; Aubol *et al.*, 2014). TRA2B is also subject to regulatory dephosphorylation by protein phosphatase PP1 (Novoyatleva *et al.*, 2008). Inhibitors for all of the aforementioned kinases and phosphatases have been developed, and their effects have been tested preclinically in models of PCa (Mavrou *et al.*, 2015; Uzor *et al.*, 2021; Liu, Han *et al.*, 2015). However, none of these studies examined AR splicing or AR-Vs. Therefore, research into pharmacological inhibition of kinases or phosphatases as a means to module TRA2 activity and its ability to control AR-V7 splicing would be of great interest. Interestingly, examination of CWR22Rv1 RNA-Seq experiments showed that *SRPK1* expression was significantly downregulated upon combined TRA2A/B depletion (data not shown). This may have potentiated the effects of knockdown by altering phosphorylation status of any remaining TRA2 protein, particularly as TRA2B levels were not completely ablated by siRNA (Figure 6.11B).

CWR22v1 cells were generally more tolerant of TRA2 depletion than VCaP (Figure 6.26). Indeed, CWR22Rv1 displayed enhanced proliferation upon single TRA2 paralog depletion, whereas in VCaP, knockdown of TRA2A or TRA2B alone was sufficient for proliferative suppression (Figure 6.26B, F). This may be explained by the observed lack of compensatory TRA2 upregulation seen in the latter cell line, which could enable CWR22Rv1 to more effectively buffer loss of one paralog. Furthermore, TRA2 knockdown failed to sensitise CWR22Rv1 to 10  $\mu$ M enzalutamide (Figure 6.26D). Increased concentrations of enzalutamide could potentially have been tested, however 10  $\mu$ M is typically employed in *in vitro* PCa experiments as it represents a 1,000x higher concentration than 10 nM DHT which is a routinely used maximal concentration for AR stimulation (Jones *et al.*, 2017; Kounatidou *et al.*, 2019). As enzalutamide is a competitive AR inhibitor with just 2-3 fold lower affinity for the receptor than DHT (Tran and Ouk, *et al.*, 2009), this concentration should be more than adequate for abrogation of AR-FL transactivation. Although whether the increased AR-FL levels observed in CWR22Rv1 upon dual TRA2A/B depletion may necessitate higher NSAA concentrations would merit investigation.

AR amplification is a known mechanism of NSAA resistance (McKay *et al.*, 2021), thus the marked induction of AR-FL seen in TRA2A/B-depleted CWR22Rv1 may function as a *de facto* amplification, requiring a higher dose of enzalutamide or use of more potent antiandrogens such as darolutamide (Bastos and Antonarakis, 2019). Other pathways altered by combined TRA2 paralog depletion may also explain maintenance of enzalutamide resistance in

CWR22Rv1. For example, re-examination of RNA-Seq analysis revealed that insulin-like growth factor 1 (IGF1) is significantly upregulated in TRA2 depleted cells (data not shown). IGF1 expression and binding to the IGF1 receptor is implicated in enzalutamide resistance by facilitating ligand-independent AR transactivation (Liu *et al.*, 2023). This has generated interest in treatment regimens combining the anti-IGF1 monoclonal antibody, xentuzumab, with enzalutamide. Such an approach has demonstrated efficacy in enzalutamide-resistant CRPC xenograft models (Weyer-Czernilofsky *et al.*, 2020). However, a phase II clinical trial in mCRPC patients failed to show any survival benefit of xentuzumab in addition to enzalutamide (Hussain *et al.*, 2019), although patients in this trial had already progressed on abiraterone and docetaxel, potentially influencing outcomes. Whilst the impact of IGF1 expression in the context of TRA2 depletion is inconclusive, it does exemplify that transcriptomic effects arising from splicing factor manipulation require careful consideration. It had been hypothesised that knockdown of TRA2 paralogs may reprogram CWR22Rv1 cells back towards enzalutamide vulnerability by favouring AR-FL production over AR-V7. However, if this concomitantly activates other potential mechanisms of resistance such as the aforementioned, then it will fail to achieve this sensitisation.

VCaP cells were sensitised to enzalutamide by lone or combined TRA2 knockdown (Figure 6.26G). Although, this may be due to AR-independent effects of TRA2 depletion, as TRA2A knockdown sensitised cells as effectively as that of TRA2B or TRA2A/B (Figure 6.26G). This is despite TRA2A siRNA failing to elicit as potent an effect on AR-V7 as the other treatment arms tested in this cell line (Figure 6.13). A thorough assessment of potential toxicities arising from TRA2 depletion would enable more confident evaluation of whether TRA2 proteins represent viable therapeutic targets in CRPC. This could be achieved by, for example, assaying proliferative effects of TRA2 depletion in AR-negative PCa cell lines such as PC3 (Tai *et al.*, 2011), or even non-prostate cell types. Pan-cellular toxicity would render a therapeutic index unlikely, although DepMap survival dependency profiles for TRA2B were more selective than that of other splicing factors analysed (Figure 6.7A). However, the observation that dual TRA2 depletion inhibited growth of both cell lines tested in the absence of enzalutamide (Figure 6.26B, F), despite their maintenance of AR-FL expression, would suggest deleterious effects that are independent of AR, which require further study.

Use of a PMO complementary to the CE3 splicing enhancer, CE3 block, resulted in potent abrogation of AR-V7 levels in both CWR22Rv1 and VCaP cells. Although the CE3 block PMO cannot definitively confirm its effects on AR-V7 are due to specific blockade of TRA2 access at the ESE, its complementarity to the GAAGAA enhancer sequence, which is a proven TRA2B binding motif (Tsuda *et al.*, 2011), strengthens evidence. Furthermore, the effect seen by blocking this sequence confirms dCasRx-APEX2 proteomics was centred on a functionally important region for AR-V7 splicing, as the AR g2 gRNA binds a short distance upstream of the CE3 block PMO.

Antisense oligonucleotide approaches such as this could provide an alternative means to modulate oncogenic splicing, obviating the need to inhibit splicing factors altogether. These may show more favourable toxicity profiles by specifically modulating the pathogenic splice site. Currently, two PMOs have received FDA approval for treating Duchenne muscular dystrophy, eteplirsen and golodirsenin, which alter splicing patterns to induce exon skipping of a premature stop codon involved in disease pathogenesis (Aartsma-Rus and Krieg, 2017; Aartsma-Rus and Corey, 2020; Roberts, Langer and Wood, 2020). PMOs have yet to reach clinical approval in cancer, and whether they will prove as efficacious in the oncology setting remains to be seen.

Results presented in this chapter have demonstrated that a dCasRx-APEX2 approach, developed and applied to AR CE3 mRNA in the CWR22Rv1 cell line model of CRPC, enriches a range of proteins that represent candidate AR-V7 splicing factors. Cross-referencing of this protein interactome data with gene expression and survival analyses of PCa clinical cohorts enabled selection of SR protein TRA2B as a putative regulator of AR-V7 generation. Subsequent experimental validation demonstrated that TRA2B and its paralog, TRA2A, are *bona fide* AR-V7 splicing factors. Mechanistically, it is probable, based on their known functions, that TRA2 proteins bind within CE3 to recruit spliceosomal components, mediating inclusion in the AR-V7 mRNA transcript. This splicing regulatory axis merits further investigation as a potential therapeutic target in CRPC, manipulation of which may reprogram disease towards druggable vulnerabilities by ablating production of intractable AR-V7 (Figure 6.29).

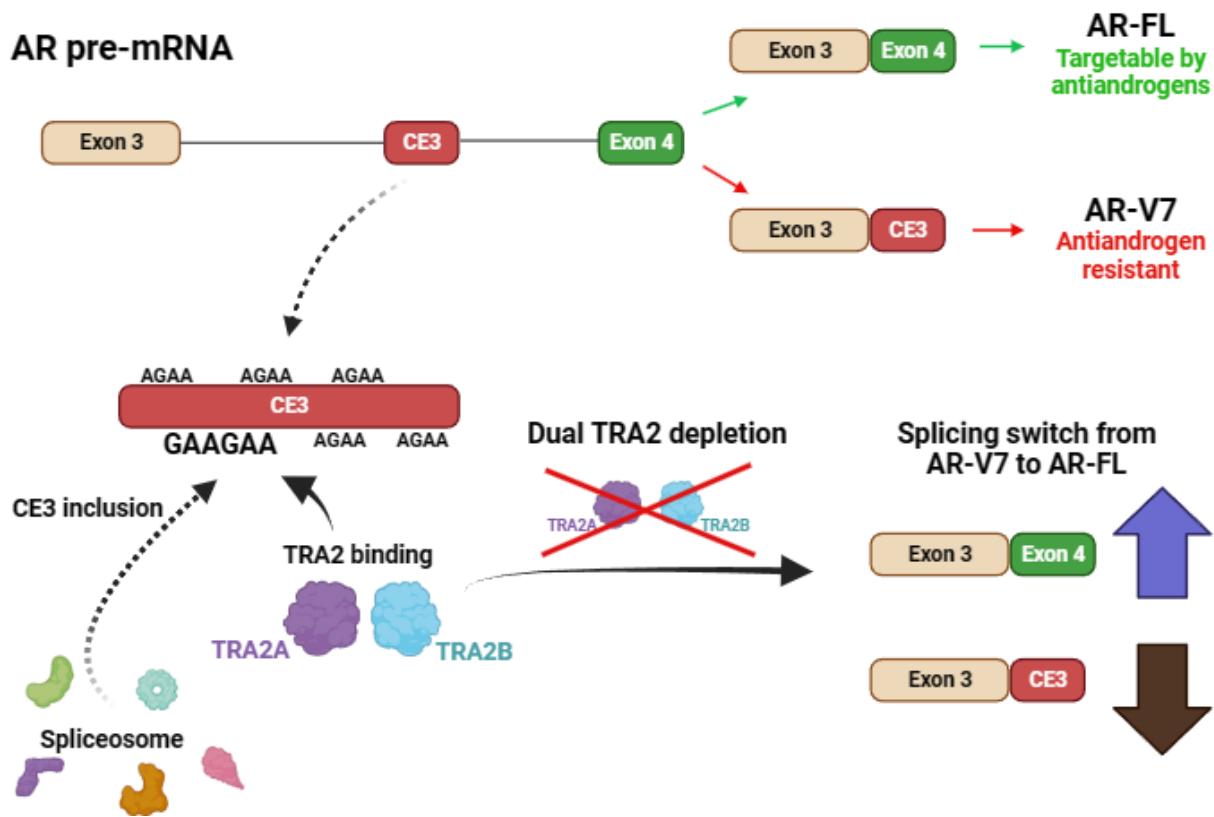


Figure 6.29 - Mechanistic summary of TRA2 proteins' putative role in AR-V7 splicing

AR pre-mRNA is processed to form AR-V7 by splicing of CE3, rather than exon 4, at the 3' end of exon 3. AR-V7 alternative splicing mediates resistance to antiandrogen treatment as the resulting protein lacks the targetable LBD of AR-FL. Evidence presented here suggests that TRA2B, identified through dCasRx-APEX2 targeted proteomics, and its paralog TRA2A are key regulators in AR-V7 generation. TRA2 proteins putatively bind a GAAGAA hexamer exonic splicing enhancer within the AGAA-rich CE3 mRNA sequence, in doing so facilitating recruitment of the spliceosome and CE3 inclusion in the nascent transcript. Depletion of both, but not lone, TRA2 paralogs has been shown here to drive a splicing switch away from AR-V7 production and back to AR-FL generation. AR-FL is druggable by existing antiandrogens, therefore TRA2 splicing factors may be exploited as novel therapeutic targets to reprogram intractable CRPC towards a treatable phenotype (Figure created using biorender.com)

## Chapter 7 Conclusions and future work

The primary aim of this PhD research project was application of novel CRISPR-Cas13 and proximity biotinylation techniques to the discovery of novel splicing factors involved in AR-V7 generation in CRPC. Specifically, the CRISPR-Cas13 subtype RfxCas13d (CasRx), or CasRx, was combined with engineered ascorbate peroxidase APEX2.

Although proteomics workflows were solely developed in CWR22Rv1, leveraging of patient gene expression datasets enhances the wider relevance. Furthermore, this is a powerful approach to novel splicing factor discovery as RNA-protein interactome data synergises to great effect with clinical gene expression data. Association with AR-V7 in the latter alone does not necessarily reflect causation, however a combination of interaction data and clinical correlations is convincing evidence for protein involvement in CE3 inclusion.

For this project, a more thorough validation of TRA2B was pursued as proof of concept that this methodology can provide meaningful information regarding splicing biology. However, other splicing factors identified by this approach may well have demonstrated involvement in similar processes. In our experiments, depletion of THRAP3, which scored strongly in clinical analyses, failed to result in a reduction of AR-V7 splicing (Figure 6.9). However, a recent publication performed a focused siRNA screen of splicing-related genes in CWR22Rv1 and another PCa cell line that expresses AR-V7; LNCaP95. They found that, when calculated as an average between these cell lines, THRAP3 depletion did reduce levels of AR-V7 relative to AR-FL (Paschalidis, Welti *et al.*, 2021). This study did not report individual AR-V7 perturbations for each cell line, thus the effect of THRAP3 knockdown may have been specific to LNCaP95, given we observed no effect in CWR22Rv1 cells. Widening the range of cell lines that were tested for siRNA depletion in this project may have been informative, although given all proteomics experiments were performed in CWR22Rv1 this was our primary validation model.

THRAP3 shares approximately 40% amino acid homology with another RNA processing factor: BCLAF1. In fact, BCLAF2 is an alternative name for THRAP3, and both proteins have been shown to play regulatory roles in splicing complexes. Although they lack a defined RRM they do have RS domains which enable interaction with other spliceosomal components (Varia *et al.*, 2013). Interestingly, a study in osteosarcoma cell line U2OS showed that THRAP3 and

BCLAF1 are able to compensate for one another to maintain splicing and export of numerous mRNAs involved in the DNA damage response (Vohhodina *et al.*, 2017). The authors also observed significant BCLAF1 upregulation upon siRNA knockdown of THRAP3 (Vohhodina *et al.*, 2017). THRAP3-depleted CWR22Rv1 samples from our project were re-analysed by RT-qPCR, which revealed a more than twofold increase in BCLAF1 mRNA upon THRAP3 knockdown, in line with the aforementioned study (data not shown). Therefore, THRAP3 and BCLAF1 may reflect observations in TRA2 experiments, in which the absence of both factors is required to elicit an effect on AR-V7 splicing.

Given THRAP3 presented itself as an encouraging candidate in the clinical PCa cohorts analysed, a THRAP3/BCLAF1 splicing axis warrants further examination as an additional AR-V7 regulatory mechanism. This possibility, as well as observations with TRA2B, illustrate that large-scale approaches to target discovery, such as CRISPR screens, may overlook important targets if such compensatory effects can occur between similar proteins. For example, one possible approach to filtering of proteins identified by dCasRx-APEX2 work would have been a CRISPR or RNAi screen of all genes encoding proteins in our filtered list of 63, however this would have failed to identify TRA2B which has subsequently shown, along with its paralog, to be an important mediator of CE3 inclusion. The proteomics dataset presented here defines a CE3-proximal interactome, which may be used to probe multiple aspects of AR-V7 mRNA splicing biology, and it is entirely possible that other factors to those validated here could show even greater splicing regulatory function.

Use of a control cell line such as LNCaP, which lacks AR-V7 expression (Liu *et al.*, 2014), could have provided additional power to analyses. Targeting of dCasRx-APEX2 in the latter model to CE3 mRNA, which remains transcribed in LNCaP although is not alternatively spliced, may have illuminated which RNA-interacting proteins are specifically associated with CE3 when it is spliced into transcripts. Alternatively, expansion of workflows to multiple AR-V7 expressing cell lines, or analysis of changes to the CE3 interactome between cells treated or untreated with NSAAAs, would provide additional information that could be used to identify common interacting proteins across cell types and conditions. Application of techniques in CWR22Rv1 has proved an invaluable proof of concept that dCasRx-APEX2 is a viable method for definition of splicing factors at defined mRNA regions, and this procedure could undoubtedly be expanded to a range of alternative cell lines.

Expansion to other, more biologically relevant, models of CRPC, such as three-dimensional cultures in Matrigel or *in vivo* models, may also expand understanding of AR-V7 splicing regulators and enhance translational potential. A 2006 study using breast cancer cell line MCF7 compared alternative splicing patterns in two-dimensional cultures with those of spheroids in Matrigel or *in vivo* xenografts, demonstrating that conventional two-dimensional cultures exhibited markedly different splicing patterns to the latter two models, which were largely similar (Li *et al.*, 2006). APEX2 methods have been considered inappropriate for these more complex systems due to BP/H<sub>2</sub>O<sub>2</sub> toxicity and the challenges of reagent delivery in the short timeframes used (Che and Khavari, 2017). Although, a recent publication successfully leveraged APEX2 chemistries in a three-dimensional pluripotent stem cell model to identify regulators of apical-basal cell polarity (S. Wang *et al.*, 2021), however this enzyme is yet to be utilised *in vivo*. Alternative proximity biotinylation enzymes such as BioID, which requires incubation with biotin alone, have also been employed in three-dimensional cultures (L. Wang *et al.*, 2021). Furthermore, a biotin-supplemented diet in murine systems enables *in vivo* proximity biotinylation with either BioID or TurboID enzymes (Murata *et al.*, 2021; Wei *et al.*, 2021). Therefore, although APEX2 has considerable spatiotemporal advantages for studying alternative splicing, fusing other proximity biotinylation enzymes with dCasRx, such as the aforementioned developed from biotin ligases, may enable expansion of methods into more translationally applicable model systems. This would provide significantly greater technical challenges, and whether this would be advantageous beyond work in cell lines, which is subsequently cross-referenced with clinical data as has been performed here, is not immediately clear.

Experimental validation of TRA2 proteins has provided a robust, mechanistic demonstration that alterations to AR-V7 production upon combined TRA2A/B depletion are directly as a result of AR alternative splicing changes. This is in contrast to other publications studying regulators of AR-V7 splicing, such as JMJD6, SFPQ and hnRNPA1, depletion of which reduces expression of both AR-FL and AR-V7 (Paschalis, Welti *et al.*, 2021; Takayama *et al.*, 2017; Nadiminty *et al.*, 2015). Conversely, validation of TRA2 proteins completed here demonstrates highly specific regulation of AR-V7 generation. However, further work is required to ascertain TRA2 protein potential as a therapeutic strategy in CRPC. Dual TRA2A/B

knockdown significantly inhibited proliferation in both PCa cell lines tested, even in the absence of enzalutamide treatment (Figure 6.26).

Despite great preclinical promise, alternative splicing modulation has yet to achieve its clinical potential in cancer. Phase I trials of SF3B splicing complex inhibition using the pladienolide derivatives E7107 and H3B-8800 demonstrated either unacceptable ophthalmologic toxicity or poor efficacy (Bradley and Anczuków, 2023). The SF3B complex is a core spliceosomal protein active at all splicing junctions, whether constitutive or alternative, where it forms part of the U2 snRNP (Will and Lührmann, 2011). Targeting of factors ubiquitously required for splicing is unlikely to provide therapeutic selectivity for pathogenic splicing processes, and the aim of this PhD project has been to identify factors with a degree of specificity for AR-V7 generation. Although TRA2 depletion has shown a substantial ability to alter CE3 exon inclusion to a greater degree than the vast majority of exons in the CWR22Rv1 transcriptome (Figure 6.24), the fact that dual TRA2 paralog knockdown inhibits proliferation warrants further investigation into toxicities arising from loss of TRA2 function. As discussed previously (Chapter 6.8), multiple druggable kinases and phosphatases have been identified that alter the phosphorylation status of TRA2B and have been inhibited preclinically in PCa, though studies did not examine TRA2B and their effect on AR-V7 splicing is unknown. A future avenue of work could be to screen these inhibitors for AR-V7 perturbations, which may combine with phosphoproteomics techniques to assess whether TRA2 proteins can be targeted this way, potentially providing a more therapeutically viable alternative to ablation of proteins.

A potential alternative, or supplement to, inhibition of splicing factors themselves would be delivery of antisense oligonucleotides (ASOs) complementary to CE3, which would block access of proteins to this region. Our work utilised a PMO, CE3 block, to abrogate AR-V7 generation in two independent PCa cell lines: CWR22Rv1 and VCaP (Figure 6.28). PMOs represent an attractive class of antisense splicing modulators as they have excellent stability and do not promote degradation of target transcripts, consequently they have been approved for clinical use in correcting splicing defects that cause Duchenne muscular dystrophy (Roberts, Langer and Wood, 2020). Other classes of ASO could also be applicable, such as RNA-based oligomers which have been used to alter splicing patterns in a range of cancer cell types (Leclair *et al.*, 2020). The RNA ASO nusinersen was the first FDA-approved therapy for spinal muscular atrophy. Nusinersen functions to correct splicing in *SMN2* exon 7 which is,

interestingly, a known splicing target exon of TRA2B (Neil and Bisaccia, 2019; Singh and Singh, 2018). Despite the success of ASOs in the neuromuscular conditions noted, the pharmacokinetics of this approach are prohibitive, as effective ASO delivery into tumours remains a challenge compared to other tissue types (Bradley and Anczuków, 2023).

Methods presented in this work could, in theory, be employed to interrogate local protein interactomes of any RNA region of interest. However, application of these methods is dependent on an appropriate gRNA with complementarity to a region of interest. Testing a panel of gRNAs in 22Rv1(CasRx) only yielded successful *AR* knockdown with one gRNA, AR g2 (Figure 5.2). Two additional *AR* gRNAs were tested more recently, which also failed to elicit an effect (data not shown). The current paucity of gRNA design algorithms for Cas13 is likely to inhibit application of this technology to a wider range of transcripts. Furthermore, gRNA design tools, whether for Cas13 or Cas9, only account for nucleotide composition, with no effective means to confirm pre-existing protein occupancy at any particular site. The authors of the CasRx gRNA design tool used throughout this project proposed that low gRNA design success for certain regions may be due to proteins such as splicing factors or exon junction complexes hindering CasRx access (Wessels, Méndez-Mancilla *et al*, 2020). Although gRNA design was somewhat more successful for targeting of *TP53* mRNA in HEK293FT cells (Figure 4.7), in which CasRx is expressed to a much greater degree, it is of enormous interest that AR g2 was the only gRNA to elicit *AR* knockdown in CWR22Rv1(CasRx). Principally, because alternative splicing is influenced by numerous aspects in addition to predicted splice site recognition by cognate factors, including transcriptional elongation rate and RNA secondary structure, which can in turn be dictated by chromatin compaction and histone modifications. These aspects subsequently dictate splicing factor accessibility at mRNA splice sites (Horn *et al.*, 2023; Marasco and Kornblihtt, 2023). The very fact that the only successful *AR* CasRx gRNA we tested was targeted at CE3, a region undergoing extensive alternative splicing in CWR22Rv1 cells, may allude to inherent accessibility and structural openness of CE3 mRNA to both splicing factors and CasRx.

In summary, this PhD research project has developed a contemporary dCasRx-APEX2 approach to define, for the first time, the protein interactome of *AR* CE3 mRNA in a model of CRPC. In doing so, it has provided a rich source of potential regulatory mediators dictating AR-V7 splicing, of which TRA2B and its paralog TRA2A have been experimentally validated as

*bona fide* novel AR-V7 splicing factors. Although currently, the clinical potential of TRA2 proteins as therapeutic targets that module AR-V7 generation in CRPC requires more extensive functional and toxicological assessment. Finally, the methods presented here demonstrate effective synergy of two contemporary biomolecular techniques, CRISPR-Cas13 and proximity biotinylation, that could have broad relevance in studying protein interactomes at various transcript regions. Therefore, this represents a powerful technique to uncover novel regulatory biology dictating RNA function.

## Chapter 8 References

Aaron, L., Franco, O. & Hayward, S.W. (2016) 'Review of Prostate Anatomy and Embryology and the Etiology of BPH', *The Urologic clinics of North America*, 43(3), pp. 279–288.

Aartsma-Rus, A. & Corey, D.R. (2020) 'The 10th Oligonucleotide Therapy Approved: Golodirsen for Duchenne Muscular Dystrophy', *Nucleic Acid Therapeutics*, 30(2), p. 67.

Aartsma-Rus, A. & Krieg, A.M. (2017) 'FDA Approves Eteplirsen for Duchenne Muscular Dystrophy: The Next Chapter in the Eteplirsen Saga', *Nucleic Acid Therapeutics*, 27(1), p. 1.

Abdulkadir, S.A., Magee, J.A., Peters, T.J., Kaleem, Z., Naughton, C.K., Humphrey, P.A. & Milbrandt, J. (2002) 'Conditional Loss of Nkx3.1 in Adult Mice Induces Prostatic Intraepithelial Neoplasia', *Molecular and Cellular Biology*, 22(5), pp. 1495–1503.

Abida, W., Cyrta, J., Heller, G., Prandi, D., Armenia, J., Coleman, I., Cieslik, M., Benelli, M., Robinson, D., Van Allen, E.M., Sboner, A., Fedrizzi, T., Mosquera, J.M., Robinson, B.D., De Sarkar, N., Kunju, L.P., Tomlins, S., Wu, Y.M., Nava Rodrigues, D., et al. (2019) 'Genomic correlates of clinical outcome in advanced prostate cancer', *Proceedings of the National Academy of Sciences of the United States of America*, 116(23), pp. 11428–11436.

Abreu, A.P. & Kaiser, U.B. (2016) 'Pubertal development and regulation', *The lancet. Diabetes & endocrinology*, 4(3), pp. 254–264.

Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B.T., Shmakov, S., Makarova, K.S., Semenova, E., Minakhin, L., Severinov, K., Regev, A., Lander, E.S., Koonin, E.V. & Zhang, F. (2016) 'C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector', *Science*, 353(6299), p. aaf5573.

Adamo, P. & Ladomery, M.R. (2016) 'The oncogene ERG: a key factor in prostate cancer', *Oncogene*, 35(4), pp. 403–414.

Adhikari, S., Nice, E.C., Deutsch, E.W., Lane, L., Omenn, G.S., Pennington, S.R., Paik, Y.-K., Overall, C.M., Corrales, F.J., Cristea, I.M., Van Eyk, J.E., Uhlén, M., Lindskog, C., Chan, D.W., Bairoch, A., Waddington, J.C., Justice, J.L., LaBaer, J., Rodriguez, H., et al. (2020) 'A high-stringency blueprint of the human proteome', *Nature Communications*, 11(1), p. 5301.

Adler, D., Lindstrom, A., Ellinger, J., Rogenhofer, S., Buettner, R., Perner, S. & Wernert, N. (2012) 'The peripheral zone of the prostate is more prone to tumor development than the transitional zone: is the ETS family the key?', *Molecular Medicine Reports*, 5(2), pp. 313–316.

Aebersold, R., Agar, J.N., Amster, I.J., Baker, M.S., Bertozzi, C.R., Boja, E.S., Costello, C.E., Cravatt, B.F., Fenselau, C., Garcia, B.A., Ge, Y., Gunawardena, J., Hendrickson, R.C., Hergenrother, P.J., Huber, C.G., Ivanov, A.R., Jensen, O.N., Jewett, M.C., Kelleher, N.L., et al. (2018) 'How many human proteoforms are there?', *Nature chemical biology*, 14(3), pp. 206–214.

Agirre, E., Oldfield, A.J., Bellora, N., Segelle, A. & Luco, R.F. (2021) 'Splicing-associated chromatin signatures: a combinatorial and position-dependent role for histone marks in splicing definition', *Nature Communications*, 12(1), p. 682.

Agoulnik, I.U., Vaid, A., Nakka, M., Alvarado, M., Bingman, W.E., III, Erdem, H., Frolov, A., Smith, C.L., Ayala, G.E., Ittmann, M.M. & Weigel, N.L. (2006) 'Androgens Modulate Expression of Transcription Intermediary Factor 2, an Androgen Receptor Coactivator whose Expression Level Correlates with Early Biochemical Recurrence in Prostate Cancer', *Cancer Research*, 66(21), pp. 10594–10602.

Ai, Y., Liang, D. & Wilusz, J.E. (2022) 'CRISPR/Cas13 effectors have differing extents of off-target effects that limit their utility in eukaryotic cells', *Nucleic Acids Research*, 50(11), p. e65.

Aird, D., Teng, T., Huang, C.-L., Pazolli, E., Banka, D., Cheung-Ong, K., Eifert, C., Furman, C., Wu, Z.J., Seiler, M., Buonamici, S., Fekkes, P., Karr, C., Palacino, J., Park, E., Smith, P.G., Yu, L., Mizui, Y., Warmuth, M., et al. (2019) 'Sensitivity to splicing modulation of BCL2 family genes defines cancer therapeutic strategies for splicing modulators', *Nature Communications*, 10(1), p. 137.

Akaza, H., Hinotsu, S., Usami, M., Arai, Y., Kanetake, H., Naito, S., Hirao, Y., & Study Group for the Combined Androgen Blockade Therapy of Prostate Cancer (2009) 'Combined androgen blockade with bicalutamide for advanced prostate cancer: long-term follow-up of a phase 3, double-blind, randomized study for survival', *Cancer*, 115(15), pp. 3437–3445.

Alimirah, F., Chen, J., Xin, H. & Choubey, D. (2006) 'Androgen receptor auto-regulates its expression by a negative feedback loop through upregulation of IFI16 protein', *FEBS letters*, 580(6), pp. 1659–1664.

Alizadeh, M. & Alizadeh, S. (2014) 'Survey of Clinical and Pathological Characteristics and Outcomes of Patients With Prostate Cancer', *Global Journal of Health Science*, 6(7), pp. 49–57.

Allen, D., Rosenberg, M. & Hendel, A. (2021) 'Using Synthetically Engineered Guide RNAs to Enhance CRISPR Genome Editing Systems in Mammalian Cells', *Frontiers in Genome Editing*, 2p. 617910.

An, J., Wang, C., Deng, Y., Yu, L. & Huang, H. (2014) 'Destruction of Full-Length Androgen Receptor by Wild-Type SPOP, but Not Prostate-Cancer-Associated Mutants', *Cell reports*, 6(4), pp. 657–669.

Anczuków, O. & Krainer, A.R. (2016) 'Splicing-factor alterations in cancers', *RNA*, 22(9), pp. 1285–1301.

Anderson, J., Abrahamsson, P.-A., Crawford, D., Miller, K. & Tombal, B. (2008) 'Management of advanced prostate cancer: can we improve on androgen deprivation therapy?', *BJU International*, 101(12), pp. 1497–1501.

Andreoiu, M. & Cheng, L. (2010) 'Multifocal prostate cancer: biologic, prognostic, and therapeutic implications', *Human Pathology*, 41(6), pp. 781–793.

Andrews, S. (2010) *Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data*. [Online] [online]. Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (Accessed 7 June 2023).

Antonarakis, E.S., Lu, C., Wang, H., Luber, B., Nakazawa, M., Roeser, J.C., Chen, Yan, Mohammad, T.A., Chen, Yidong, Fedor, H.L., Lotan, T.L., Zheng, Q., De Marzo, A.M., Isaacs, J.T., Isaacs, W.B., Nadal, R., Paller, C.J., Denmeade, S.R., Carducci, M.A., et al. (2014) 'AR-V7 and Resistance to Enzalutamide and Abiraterone in Prostate Cancer', *New England Journal of Medicine*, 371(11), pp. 1028–1038.

Antony, L., van der Schoor, F., Dalrymple, S.L. & Isaacs, J.T. (2014) 'Androgen Receptor (AR) Suppresses Normal Human Prostate Epithelial Cell Proliferation via AR/β-catenin/TCF-4 Complex Inhibition of c-MYC Transcription', *The Prostate*, 74(11), pp. 1118–1131.

Araujo, J.C., Trudel, G.C., Saad, F., Armstrong, A.J., Yu, E.Y., Bellmunt, J., Wilding, G., McCaffrey, J., Serrano, S.V., Matveev, V.B., Efsthathiou, E., Oudard, S., Morris, M.J., Sizer, B., Goebell, P.J., Heidenreich, A., de Bono, J.S., Begbie, S., Hong, J.H., et al. (2013) 'Docetaxel and dasatinib or placebo in men with metastatic castration-resistant prostate cancer (READY): a randomised, double-blind phase 3 trial', *The Lancet Oncology*, 14(13), pp. 1307–1316.

Argos, P. (1990) 'An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion', *Journal of Molecular Biology*, 211(4), pp. 943–958.

Armandari, I., Hamid, A.R., Verhaegh, G. & Schalken, J. (2014) 'Intratumoral steroidogenesis in castration-resistant prostate cancer: a target for therapy', *Prostate International*, 2(3), pp. 105–113.

Armstrong, A.J., Lin, P., Tombal, B., Saad, F., Higano, C.S., Joshua, A.M., Parli, T., Rosbrook, B., van Os, S. & Beer, T.M. (2020) 'Five-year Survival Prediction and Safety Outcomes with Enzalutamide in Men with Chemotherapy-naïve Metastatic Castration-resistant Prostate Cancer from the PREVAIL Trial', *European Urology*, 78(3), pp. 347–357.

Armstrong, A.J., Szmulewitz, R.Z., Petrylak, D.P., Holzbeierlein, J., Villers, A., Azad, A., Alcaraz, A., Alekseev, B., Iguchi, T., Shore, N.D., Rosbrook, B., Sugg, J., Baron, B., Chen, L. & Stenzl, A. (2019) 'ARCHES: A Randomized, Phase III Study of Androgen Deprivation Therapy With Enzalutamide or Placebo in Men With Metastatic Hormone-Sensitive Prostate Cancer', *Journal of Clinical Oncology*, 37(32), pp. 2974–2986.

Arora, R., Koch, M.O., Eble, J.N., Ulbright, T.M., Li, L. & Cheng, L. (2004) 'Heterogeneity of Gleason grade in multifocal adenocarcinoma of the prostate', *Cancer*, 100(11), pp. 2362–2366.

Artemyeva-Isman, O.V. & Porter, A.C.G. (2021) 'U5 snRNA Interactions With Exons Ensure Splicing Precision', *Frontiers in Genetics*, 12.

Asim, M., Siddiqui, I., Hafeez, B., Baniahmad, A. & Mukhtar, H. (2008) 'Src kinase potentiates androgen receptor transactivation function and invasion of androgen-independent prostate cancer C4-2 cells', *Oncogene*, 27(25), pp. 3596–3604.

Askew, E.B., Minges, J.T., Hnat, A.T. & Wilson, E.M. (2012) 'Structural features discriminate androgen receptor N/C terminal and coactivator interactions', *Molecular and cellular endocrinology*, 348(2), pp. 403–410.

Ateeq, B., Tomlins, S.A., Laxman, B., Asangani, I.A., Cao, Q., Cao, X., Li, Y., Wang, X., Feng, F.Y., Pienta, K.J., Varambally, S. & Chinnaiyan, A.M. (2011) 'Therapeutic Targeting of SPINK1-Positive Prostate Cancer', *Science Translational Medicine*, 3(72), pp. 72ra17-72ra17.

Aubol, B.E., Plocinik, R.M., Keshwani, M.M., McGlone, M.L., Hagopian, J.C., Ghosh, G., Fu, X.-D. & Adams, J.A. (2014) 'N-terminus of the protein kinase CLK1 induces SR protein hyperphosphorylation', *Biochemical Journal*, 462(1), pp. 143–152.

Auchus, R.J., Yu, M.K., Nguyen, S. & Mundle, S.D. (2014) 'Use of Prednisone With Abiraterone Acetate in Metastatic Castration-Resistant Prostate Cancer', *The Oncologist*, 19(12), pp. 1231–1240.

Augello, M.A., Liu, D., Deonarine, L.D., Robinson, B.D., Huang, D., Stelloo, S., Blattner, M., Doane, A.S., Wong, E.W.P., Chen, Y., Rubin, M.A., Beltran, H., Elemento, O., Bergman, A.M., Zwart, W., Sboner, A., Dephoure, N. & Barbieri, C.E. (2019) 'CHD1 Loss Alters AR Binding at Lineage-Specific Enhancers and Modulates Distinct Transcriptional Programs to Drive Prostate Tumorigenesis', *Cancer Cell*, 35(4), pp. 603-617.e8.

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Schaeffer, R.D., Millán, C., Park, H., Adams, C., Glassman, C.R., DeGiovanni, A., Pereira, J.H., Rodrigues, A.V., van Dijk, A.A., Ebrecht, A.C., et al. (2021) 'Accurate prediction of protein structures and interactions using a three-track neural network', *Science*, 373(6557), pp. 871–876.

Bagi, C.M. (2003) 'Skeletal implications of prostate cancer', *Journal of Musculoskeletal & Neuronal Interactions*, 3(2), pp. 112–117.

Barbieri, C.E., Baca, S.C., Lawrence, M.S., Demichelis, F., Blattner, M., Theurillat, J.-P., White, T.A., Stojanov, P., Van Allen, E., Stransky, N., Nickerson, E., Chae, S.-S., Boysen, G., Auclair, D., Onofrio, R.C., Park, K., Kitabayashi, N., MacDonald, T.Y., Sheikh, K., et al. (2012) 'Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer', *Nature Genetics*, 44(6), pp. 685–689.

Barnes, C.E., English, D.M., Broderick, M., Collins, M.O. & Cowley, S.M. (2022) 'Proximity-dependent biotin identification (BioID) reveals a dynamic LSD1-CoREST interactome during embryonic stem cell differentiation', *Molecular Omics*, 18(1), pp. 31–44.

Baron, S., Manin, M., Beaudoin, C., Leotoing, L., Communal, Y., Veyssiere, G. & Morel, L. (2004) 'Androgen receptor mediates non-genomic activation of phosphatidylinositol 3-OH kinase in androgen-sensitive epithelial cells', *The Journal of Biological Chemistry*, 279(15), pp. 14579–14586.

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippe, K.H., Sherman, P.M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C.L.,

Serova, N., Davis, S. & Soboleva, A. (2013) 'NCBI GEO: archive for functional genomics data sets—update', *Nucleic Acids Research*, 41(D1), pp. D991–D995.

Barrie, S.E., Potter, G.A., Goddard, P.M., Haynes, B.P., Dowsett, M. & Jarman, M. (1994) 'Pharmacology of novel steroidal inhibitors of cytochrome P45017 $\alpha$  (17 $\alpha$ -hydroxylase/C17–20 lyase)', *The Journal of Steroid Biochemistry and Molecular Biology*, 50(5), pp. 267–273.

Basil, P., Robertson, M.J., Bingman, W.E., Dash, A.K., Krause, W.C., Shafi, A.A., Piyarathna, B., Coarfa, C. & Weigel, N.L. (2022) 'Cistrome and transcriptome analysis identifies unique androgen receptor (AR) and AR-V7 splice variant chromatin binding and transcriptional activities', *Scientific Reports*, 12(1), p. 5351.

Bastos, D.A. & Antonarakis, E.S. (2019) 'Darolutamide For Castration-Resistant Prostate Cancer', *OncoTargets and therapy*, 12pp. 8769–8777.

Beato, M., Herrlich, P. & Schütz, G. (1995) 'Steroid hormone receptors: Many Actors in search of a plot', *Cell*, 83(6), pp. 851–857.

Beil, B., Screamton, G. & Stamm, S. (1997) 'Molecular cloning of htra2-beta-1 and htra2-beta-2, two human homologs of tra-2 generated by alternative splicing', *DNA and cell biology*, 16(6), pp. 679–690.

Beltran, H., Yelensky, R., Frampton, G.M., Park, K., Downing, S.R., MacDonald, T.Y., Jarosz, M., Lipson, D., Tagawa, S.T., Nanus, D.M., Stephens, P.J., Mosquera, J.M., Cronin, M.T. & Rubin, M.A. (2013) 'Targeted Next-generation Sequencing of Advanced Prostate Cancer Identifies Potential Therapeutic Targets and Disease Heterogeneity', *European urology*, 63(5), pp. 920–926.

Benjamini, Y. & Hochberg, Y. (1995) 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing', *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), pp. 289–300.

Berget, S.M., Moore, C. & Sharp, P.A. (1977) 'Spliced segments at the 5' terminus of adenovirus 2 late mRNA\*', *Proceedings of the National Academy of Sciences*, 74(8), pp. 3171–3175.

Berglund, J.A., Abovich, N. & Rosbash, M. (1998) 'A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition', *Genes & Development*, 12(6), pp. 858–867.

Bernasocchi, T. & Theurillat, J.-P.P. (2022) 'SPOP-mutant prostate cancer: Translating fundamental biology into patient care', *Cancer Letters*, 529pp. 11–18.

Best, A., Dagliesh, C., Ehrmann, I., Kheirollahi-Kouhestani, M., Tyson-Capper, A. & Elliott, D.J. (2013) 'Expression of Tra2  $\beta$  in Cancer Cells as a Potential Contributory Factor to Neoplasia and Metastasis', *International Journal of Cell Biology*, 2013p. 843781.

Best, A., James, K., Dalgliesh, C., Hong, E., Kheirollahi-Kouhestani, M., Curk, T., Xu, Y., Danilenko, M., Hussain, R., Keavney, B., Wipat, A., Klinck, R., Cowell, I.G., Cheong Lee, K., Austin, C.A., Venables, J.P., Chabot, B., Santibanez Koref, M., Tyson-Capper, A., et al. (2014)

'Human Tra2 proteins jointly control a CHEK1 splicing switch among alternative and constitutive target exons', *Nature Communications*, 5p. 4760.

Bevan, C.L., Hoare, S., Claessens, F., Heery, D.M. & Parker, M.G. (1999) 'The AF1 and AF2 Domains of the Androgen Receptor Interact with Distinct Regions of SRC1', *Molecular and Cellular Biology*, 19(12), pp. 8383–8392.

Bhanalaph, T., Varkarakis, M.J. & Murphy, G.P. (1974) 'Current status of bilateral adrenalectomy or advanced prostatic carcinoma.', *Annals of Surgery*, 179(1), pp. 17–23.

Bhatia-Gaur, R., Donjacour, A.A., Sciavolino, P.J., Kim, M., Desai, N., Young, P., Norton, C.R., Gridley, T., Cardiff, R.D., Cunha, G.R., Abate-Shen, C. & Shen, M.M. (1999) 'Roles for Nkx3.1 in prostate development and cancer', *Genes & Development*, 13(8), pp. 966–977.

Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'Donovan, C. & Apweiler, R. (2009) 'QuickGO: a web-based tool for Gene Ontology searching', *Bioinformatics*, 25(22), pp. 3045–3046.

Blighe, K., Rana, S. & Lewis, M. (2021) *EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling*.

Bohl, C.E., Gao, W., Miller, D.D., Bell, C.E. & Dalton, J.T. (2005) 'Structural basis for antagonism and resistance of bicalutamide in prostate cancer', *Proceedings of the National Academy of Sciences of the United States of America*, 102(17), pp. 6201–6206.

Bohl, C.E., Wu, Z., Miller, D.D., Bell, C.E. & Dalton, J.T. (2007) 'Crystal Structure of the T877A Human Androgen Receptor Ligand-binding Domain Complexed to Cyproterone Acetate Provides Insight for Ligand-induced Conformational Changes and Structure-based Drug Design', *The Journal of biological chemistry*, 282(18), pp. 13648–13655.

Bolte, S. & Cordelières, F.P. (2006) 'A guided tour into subcellular colocalization analysis in light microscopy', *Journal of Microscopy*, 224(3), pp. 213–232.

de Bono, J.S., Logothetis, C.J., Molina, A., Fizazi, K., North, S., Chu, L., Chi, K.N., Jones, R.J., Goodman, O.B., Saad, F., Staffurth, J.N., Mainwaring, P., Harland, S., Flraig, T.W., Hutson, T.E., Cheng, T., Patterson, H., Hainsworth, J.D., Ryan, C.J., et al. (2011) 'Abiraterone and Increased Survival in Metastatic Prostate Cancer', *New England Journal of Medicine*, 364(21), pp. 1995–2005.

Bookstein, R., MacGrogan, D., Hilsenbeck, S.G., Sharkey, F. & Allred, D.C. (1993) 'p53 is mutated in a subset of advanced-stage prostate cancers', *Cancer Research*, 53(14), pp. 3369–3373.

Bostwick, D.G. (1995) 'High grade prostatic intraepithelial neoplasia. The most likely precursor of prostate cancer', *Cancer*, 75(S7), pp. 1823–1836.

Bostwick, D.G. & Brawer, M.K. (1987) 'Prostatic Intra-Epithelial Neoplasia and Early Invasion in Prostate Cancer', *Cancer*, 59(4), pp. 788–794.

Bostwick, D.G., Burke, H.B., Djakiew, D., Euling, S., Ho, S., Landolph, J., Morrison, H., Sonawane, B., Shifflett, T., Waters, D.J. & Timms, B. (2004) 'Human prostate cancer risk factors', *Cancer*, 101(S10), pp. 2371–2490.

Bostwick, D.G. & Qian, J. (2004) 'High-grade prostatic intraepithelial neoplasia', *Modern Pathology*, 17(3), pp. 360–379.

Boukis, L.A., Liu, N., Furuyama, S. & Bruzik, J.P. (2004) 'Ser/Arg-rich Protein-mediated Communication between U1 and U2 Small Nuclear Ribonucleoprotein Particles\*', *Journal of Biological Chemistry*, 279(28), pp. 29647–29653.

Bowen, C., Ostrowski, M.C., Leone, G. & Gelmann, E.P. (2019) 'Loss of PTEN Accelerates NKX3.1 Degradation to Promote Prostate Cancer Progression', *Cancer Research*, 79(16), pp. 4124–4134.

Bowen, C., Zheng, T. & Gelmann, E.P. (2015) 'NKX3.1 Suppresses TMPRSS2–ERG Gene Rearrangement and Mediates Repair of Androgen Receptor–Induced DNA Damage', *Cancer Research*, 75(13), pp. 2686–2698.

Bowers, R.F. (1962) 'Adrenalectomy for Hopeless Carcinoma of the Prostate', *Archives of Surgery*, 84(4), pp. 421–424.

Boysen, G., Barbieri, C.E., Prandi, D., Blattner, M., Chae, S.-S., Dahija, A., Nataraj, S., Huang, D., Marotz, C., Xu, L., Huang, J., Lecca, P., Chhangawala, S., Liu, D., Zhou, P., Sboner, A., de Bono, J.S., Demichelis, F., Houvras, Y., et al. (2015) 'SPOP mutation leads to genomic instability in prostate cancer' Joaquín M Espinosa (ed.), *eLife*, 4p. e09207.

Bracken, C.P., Wall, S.J., Barré, B., Panov, K.I., Ajuh, P.M. & Perkins, N.D. (2008) 'Regulation of Cyclin D1 RNA Stability by SNIP1', *Cancer Research*, 68(18), pp. 7621–7628.

Bradley, R.K. & Añczuków, O. (2023) 'RNA splicing dysregulation and the hallmarks of cancer', *Nature Reviews Cancer*, 23(3), pp. 135–155.

Branlant, C., Krol, A., Ebel, J.P., Lazar, E., Haendler, B. & Jacob, M. (1982) 'U2 RNA shares a structural domain with U1, U4, and U5 RNAs.', *The EMBO Journal*, 1(10), pp. 1259–1265.

Branon, T.C., Bosch, J.A., Sanchez, A.D., Udeshi, N.D., Svinkina, T., Carr, S.A., Feldman, J.L., Perrimon, N. & Ting, A.Y. (2018) 'Efficient proximity labeling in living cells and organisms with TurboID', *Nature Biotechnology*, 36(9), pp. 880–887.

Brawer, M.K. (2005) 'Prostatic Intraepithelial Neoplasia: An Overview', *Reviews in Urology*, 7(Suppl 3), pp. S11–S18.

Brinkmann, A.O., Lindh, L.M., Breedveld, D.I., Mulder, E. & Van Der Molen, H.J. (1983) 'Cyproterone acetate prevents translocation of the androgen receptor in the rat prostate', *Molecular and Cellular Endocrinology*, 32(2–3), pp. 117–129.

Bryant, R. J., Cross, N. A., Eaton, C. I., Hamdy, F. C. & Cunliffe, V. T. (2007) 'EZH2 promotes proliferation and invasiveness of prostate cancer cells', *The Prostate*, 67(5), pp. 547–556.

Bryce, A. & Ryan, C.J. (2012) 'Development and Clinical Utility of Abiraterone Acetate as an Androgen Synthesis Inhibitor', *Clinical Pharmacology & Therapeutics*, 91(1), pp. 101–108.

Brzozowski, A.M., Pike, A.C.W., Dauter, Z., Hubbard, R.E., Bonn, T., Engström, O., Öhman, L., Greene, G.L., Gustafsson, J.-Å. & Carlquist, M. (1997) 'Molecular basis of agonism and antagonism in the oestrogen receptor', *Nature*, 389(6652), pp. 753–758.

Bubendorf, L., Kononen, J., Koivisto, P., Schraml, P., Moch, H., Gasser, T.C., Willi, N., Mihatsch, M.J., Sauter, G. & Kallioniemi, O.P. (1999) 'Survey of gene amplifications during prostate cancer progression by high-throughout fluorescence in situ hybridization on tissue microarrays', *Cancer Research*, 59(4), pp. 803–806.

Bubendorf, L., Schöpfer, A., Wagner, U., Sauter, G., Moch, H., Willi, N., Gasser, T.C. & Mihatsch, M.J. (2000) 'Metastatic patterns of prostate cancer: An autopsy study of 1,589 patients', *Human Pathology*, 31(5), pp. 578–583.

Buchman, A.B., Brogan, D.J., Sun, R., Yang, T., Hsu, P.D. & Akbari, O.S. (2020) 'Programmable RNA Targeting Using CasRx in Flies', *The CRISPR Journal*, 3(3), pp. 164–176.

Burkhart, D.L., Morel, K.L., Sheahan, A.V., Richards, Z.A. & Ellis, L. (2019) 'The Role of RB in Prostate Cancer Progression', in Scott M. Dehm & Donald J. Tindall (eds.) *Prostate Cancer: Cellular and Genetic Mechanisms of Disease Development and Progression*. Advances in Experimental Medicine and Biology. [Online]. Cham: Springer International Publishing. pp. 301–318.

Butler, W. & Huang, J. (2021) 'Neuroendocrine cells of the prostate: Histology, biological functions, and molecular mechanisms', *Precision Clinical Medicine*, 4(1), pp. 25–34.

Cáceres, J.F., Misteli, T., Screamton, G.R., Spector, D.L. & Krainer, A.R. (1997) 'Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity', *The Journal of Cell Biology*, 138(2), pp. 225–238.

Cai, C., Chen, Sen, Ng, P., Bubley, G.J., Nelson, P.S., Mostaghel, E.A., Marck, B., Matsumoto, A.M., Simon, N.I., Wang, H., Chen, Shaoyong & Balk, S.P. (2011) 'Intratumoral De Novo Steroid Synthesis Activates Androgen Receptor in Castration Resistant Prostate Cancer and is Upregulated by Treatment with CYP17A1 Inhibitors', *Cancer research*, 71(20), pp. 6503–6513.

Cai, C., He, H.H., Chen, Sen, Coleman, I., Wang, H., Fang, Z., Chen, Shaoyong, Nelson, P.S., Liu, X.S., Brown, M. & Balk, S.P. (2011) 'Androgen Receptor Gene Expression in Prostate Cancer is Directly Suppressed by the Androgen Receptor Through Recruitment of Lysine Specific Demethylase 1', *Cancer cell*, 20(4), pp. 457–471.

Cai, L., Tsai, Y.-H., Wang, P., Wang, J., Li, D., Fan, H., Zhao, Y., Bareja, R., Lu, R., Wilson, E.M., Sboner, A., Whang, Y.E., Zheng, D., Parker, J.S., Earp, H.S. & Wang, G.G. (2018) 'ZFX mediates non-canonical oncogenic functions of the androgen receptor splice variant 7 (AR-V7) in castrate-resistant prostate cancer', *Molecular cell*, 72(2), pp. 341–354.e6.

Cairns, P., Esteller, M., Herman, J.G., Schoenberg, M., Jeronimo, C., Sanchez-Cespedes, M., Chow, N.-H., Grasso, M., Wu, L., Westra, W.B. & Sidransky, D. (2001) 'Molecular Detection of

Prostate Cancer in Urine by GSTP1 Hypermethylation1', *Clinical Cancer Research*, 7(9), pp. 2727–2730.

Cairns, P., Okami, K., Halachmi, S., Halachmi, N., Esteller, M., Herman, J.G., Jen, J., Isaacs, W.B., Bova, G.S. & Sidransky, D. (1997) 'Frequent Inactivation of PTEN/MMAC1 in Primary Prostate Cancer', *Cancer Research*, 57(22), pp. 4997–5000.

Callewaert, L., Van Tilborgh, N. & Claessens, F. (2006) 'Interplay between Two Hormone-Independent Activation Domains in the Androgen Receptor', *Cancer Research*, 66(1), pp. 543–553.

Cao, B., Qi, Y., Zhang, G., Xu, D., Zhan, Y., Alvarez, X., Guo, Z., Fu, X., Plymate, S.R., Sartor, O., Zhang, H. & Dong, Y. (2014) 'Androgen receptor splice variants activating the full-length receptor in mediating resistance to androgen-directed therapy', *Oncotarget*, 5(6), pp. 1646–1656.

Cao, C., Li, A., Xu, C., Wu, B., Liu, J. & Liu, Y. (2023) 'Enhancement of protein translation by CRISPR/dCasRx coupled with SINEB2 repeat of noncoding RNAs', *Nucleic Acids Research*, 51(6), p. e33.

Carl, P., Laummonerie, C. & Mutterer, J. (2004) *Colocalization Finder*.

Cassidy, L., Kaulich, P.T., Maaß, S., Bartel, J., Becher, D. & Tholey, A. (2021) 'Bottom-up and top-down proteomic approaches for the identification, characterization, and quantification of the low molecular weight proteome with focus on short open reading frame-encoded peptides', *PROTEOMICS*, 21(23–24), p. 2100008.

Cato, L., de Tribolet-Hardy, J., Lee, I., Rottenberg, J.T., Coleman, I., Melchers, D., Houtman, R., Xiao, T., Li, W., Uo, T., Sun, S., Kuznik, N.C., Göppert, B., Ozgun, F., van Royen, M.E., Houtsmuller, A.B., Vadhi, R., Rao, P.K., Li, L., et al. (2019) 'ARv7 Represses Tumor-Suppressor Genes in Castration-Resistant Prostate Cancer', *Cancer cell*, 35(3), pp. 401–413.e6.

Chamberlain, N.L., Driver, E.D. & Miesfeld, R.L. (1994) 'The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function', *Nucleic Acids Research*, 22(15), pp. 3181–3186.

Chan, S., Sridhar, P., Kirchner, R., Lock, Y.J., Herbert, Z., Buonamici, S., Smith, P., Lieberman, J. & Petrocca, F. (2017) 'Basal-A Triple-Negative Breast Cancer Cells Selectively Rely on RNA Splicing for Survival', *Molecular Cancer Therapeutics*, 16(12), pp. 2849–2861.

Chan, S.C., Li, Y. & Dehm, S.M. (2012) 'Androgen receptor splice variants activate androgen receptor target genes and support aberrant prostate cancer cell growth independent of canonical androgen receptor nuclear localization signal', *The Journal of Biological Chemistry*, 287(23), pp. 19736–19749.

Chandrasekar, T., Yang, J.C., Gao, A.C. & Evans, C.P. (2015) 'Mechanisms of resistance in castration-resistant prostate cancer (CRPC)', *Translational Andrology and Urology*, 4(3), pp. 36580–36380.

Chapman-Smith, A. & Cronan Jr, J.E. (1999) 'Molecular Biology of Biotin Attachment to Proteins', *The Journal of Nutrition*, 129(2), pp. 477S-484S.

Charenton, C., Wilkinson, M.E. & Nagai, K. (2019) 'Mechanism of 5' splice site transfer for human spliceosome activation', *Science*, 364(6438), pp. 362–367.

Charlet-B, N., Singh, G., Cooper, T.A. & Logan, P. (2002) 'Dynamic Antagonism between ETR-3 and PTB Regulates Cell Type-Specific Alternative Splicing', *Molecular Cell*, 9(3), pp. 649–658.

Che, Y. & Khavari, P.A. (2017) 'Research Techniques Made Simple: Emerging Methods to Elucidate Protein Interactions through Spatial Proximity', *Journal of Investigative Dermatology*, 137(12), pp. e197–e203.

Chen, C.D., Welsbie, D.S., Tran, C., Baek, S.H., Chen, R., Vessella, R., Rosenfeld, M.G. & Sawyers, C.L. (2004) 'Molecular determinants of resistance to antiandrogen therapy', *Nature Medicine*, 10(1), pp. 33–39.

Chen, C.-L. & Perrimon, N. (2017) 'Proximity-dependent labeling methods for proteomic profiling in living cells', *Wiley interdisciplinary reviews. Developmental biology*, 6(4), p. 10.1002/wdev.272.

Chen, S., Gulla, S., Cai, C. & Balk, S.P. (2012) 'Androgen receptor serine 81 phosphorylation mediates chromatin binding and transcriptional activation', *The Journal of Biological Chemistry*, 287(11), pp. 8571–8583.

Chen, S., Xu, Y., Yuan, X., Bubley, G.J. & Balk, S.P. (2006) 'Androgen receptor phosphorylation and stabilization in prostate cancer by cyclin-dependent kinase 1', *Proceedings of the National Academy of Sciences of the United States of America*, 103(43), pp. 15969–15974.

Chen, X., Wei, S., Ji, Y., Guo, X. & Yang, F. (2015) 'Quantitative proteomics using SILAC: Principles, applications, and developments', *PROTEOMICS*, 15(18), pp. 3175–3192.

Chen, X., Zaro, J. & Shen, W.-C. (2013) 'Fusion Protein Linkers: Property, Design and Functionality', *Advanced drug delivery reviews*, 65(10), pp. 1357–1369.

Chen, Y. & Wang, X. (2022) 'Evaluation of efficiency prediction algorithms and development of ensemble model for CRISPR/Cas9 gRNA selection', *Bioinformatics*, 38(23), pp. 5175–5181.

Chen, Z., Wu, D., Thomas-Ahner, J.M., Lu, C., Zhao, P., Zhang, Q., Geraghty, C., Yan, P.S., Hankey, W., Sunkel, B., Cheng, X., Antonarakis, E.S., Wang, Q.-E., Liu, Z., Huang, T.H.-M., Jin, V.X., Clinton, S.K., Luo, J., Huang, J., et al. (2018) 'Diverse AR-V7 cistromes in castration-resistant prostate cancer are governed by HoxB13', *Proceedings of the National Academy of Sciences*, 115(26), pp. 6810–6815.

Cheng, L., Pretlow, T.G., Abdul-Karim, F.W., Kung, H.-J., Tsai, M.-L., Dawson, D.V., Linehan, W.M., Song, S.-Y., Park, W.-S., Moon, Y.-W., Emmert-Buck, M.R., Liotta, L.A. & Zhuang, Z. (1998) 'Evidence of Independent Origin of Multiple Tumors From Patients With Prostate Cancer', *JNCI: Journal of the National Cancer Institute*, 90(3), pp. 233–237.

Chiou, N.-T., Shankarling, G. & Lynch, K.W. (2013) 'hnRNP L and hnRNP A1 induce extended U1 snRNA interactions with an exon to repress spliceosome assembly', *Molecular Cell*, 49(5), pp. 972–982.

Cho, K.F., Branen, T.C., Udeshi, N.D., Myers, S.A., Carr, S.A. & Ting, A.Y. (2020) 'Proximity labeling in mammalian cells with Turbolid and split-Turbolid', *Nature Protocols*, 15(12), pp. 3971–3999.

Cho, S., Hoang, A., Sinha, R., Zhong, X.-Y., Fu, X.-D., Krainer, A.R. & Ghosh, G. (2011) 'Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly', *Proceedings of the National Academy of Sciences of the United States of America*, 108(20), pp. 8233–8238.

Chodak, G.W. (2005) 'Maximum Androgen Blockade: A Clinical Update', *Reviews in Urology*, 7(Suppl 5), pp. S13–S17.

Cieslik, M., Chugh, R., Wu, Y.-M., Wu, M., Brennan, C., Lonigro, R., Su, F., Wang, R., Siddiqui, J., Mehra, R., Cao, X., Lucas, D., Chinnaiyan, A.M. & Robinson, D. (2015) 'The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing', *Genome Research*, 25(9), pp. 1372–1381.

Cinar, B., Mukhopadhyay, N.K., Meng, G. & Freeman, M.R. (2007) 'Phosphoinositide 3-Kinase-independent Non-genomic Signals Transit from the Androgen Receptor to Akt1 in Membrane Raft Microdomains \*', *Journal of Biological Chemistry*, 282(40), pp. 29584–29593.

Claessens, F., Alen, P., Devos, A., Peeters, B., Verhoeven, G. & Rombauts, W. (1996) 'The Androgen-specific Probasin Response Element 2 Interacts Differentially with Androgen and Glucocorticoid Receptors\*', *Journal of Biological Chemistry*, 271(32), pp. 19013–19016.

Claessens, F., Denayer, S., Van Tilborgh, N., Kerkhofs, S., Helsen, C. & Haelens, A. (2008) 'Diverse roles of androgen receptor (AR) domains in AR-mediated signaling', *Nuclear Receptor Signaling*, 6p. e008.

Cléry, A., Jayne, S., Benderska, N., Dominguez, C., Stamm, S. & Allain, F.H.-T. (2011) 'Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2-β1', *Nature Structural & Molecular Biology*, 18(4), pp. 443–450.

Cleutjens, K.B., van der Korput, H.A., van Eekelen, C.C., van Rooij, H.C., Faber, P.W. & Trapman, J. (1997) 'An androgen response element in a far upstream enhancer region is essential for high, androgen-regulated activity of the prostate-specific antigen promoter', *Molecular Endocrinology (Baltimore, Md.)*, 11(2), pp. 148–161.

Clinckemalie, L., Spans, L., Dubois, V., Laurent, M., Helsen, C., Joniau, S. & Claessens, F. (2013) 'Androgen Regulation of the TMPRSS2 Gene and the Effect of a SNP in an Androgen Response Element', *Molecular Endocrinology*, 27(12), pp. 2028–2040.

Clinckemalie, L., Vanderschueren, D., Boonen, S. & Claessens, F. (2012) 'The hinge region in androgen receptor control', *Molecular and Cellular Endocrinology*, 358(1), pp. 1–8.

Clough, E. & Barrett, T. (2016) 'The Gene Expression Omnibus Database', *Methods in Molecular Biology* (Clifton, N.J.), 1418pp. 93–110.

Cohen-Eliav, M., Golan-Gerstl, R., Siegfried, Z., Andersen, C.L., Thorsen, K., Ørntoft, T.F., Mu, D. & Karni, R. (2013) 'The splicing factor SRSF6 is amplified and is an oncoprotein in lung and colon cancers', *The Journal of Pathology*, 229(4), pp. 630–639.

Colaprico, A., Silva, T.C., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T.S., Malta, T.M., Pagnotta, S.M., Castiglioni, I., Ceccarelli, M., Bontempi, G. & Noushmehr, H. (2016) 'TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data', *Nucleic Acids Research*, 44(8), p. e71.

Conteduca, V., Wetterskog, D., Sharabiani, M.T.A., Grande, E., Fernandez-Perez, M.P., Jayaram, A., Salvi, S., Castellano, D., Romanel, A., Lolli, C., Casadio, V., Gurioli, G., Amadori, D., Font, A., Vazquez-Estevez, S., González Del Alba, A., Mellado, B., Fernandez-Calvo, O., Méndez-Vidal, M.J., et al. (2017) 'Androgen receptor gene status in plasma DNA associates with worse outcome on enzalutamide or abiraterone for castration-resistant prostate cancer: a multi-institution correlative biomarker study', *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, 28(7), pp. 1508–1516.

Cooke, B.A. & Sullivan, M.H. (1985) 'The mechanisms of LHRH agonist action in gonadal tissues', *Molecular and Cellular Endocrinology*, 41(2–3), pp. 115–122.

Cordin, O. & Beggs, J.D. (2013) 'RNA helicases in splicing', *RNA Biology*, 10(1), pp. 83–95.

Corrionero, A., Miñana, B. & Valcárcel, J. (2011) 'Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A', *Genes & Development*, 25(5), pp. 445–459.

Coutinho, I., Day, T.K., Tilley, W.D. & Selth, L.A. (2016) 'Androgen receptor signaling in castration-resistant prostate cancer: a lesson in persistence', *Endocrine-Related Cancer*, 23(12), pp. T179–T197.

Cox, D.B.T., Gootenberg, J.S., Abudayyeh, O.O., Franklin, B., Kellner, M.J., Joung, J. & Zhang, F. (2017) 'RNA Editing with CRISPR-Cas13', *Science (New York, N.Y.)*, 358(6366), pp. 1019–1027.

Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N. & Mann, M. (2014) 'Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ', *Molecular & Cellular Proteomics : MCP*, 13(9), pp. 2513–2526.

Cox, J. & Mann, M. (2008) 'MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification', *Nature Biotechnology*, 26(12), pp. 1367–1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V. & Mann, M. (2011a) 'Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment', *Journal of Proteome Research*, 10(4), pp. 1794–1805.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V. & Mann, M. (2011b) 'Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment', *Journal of Proteome Research*, 10(4), pp. 1794–1805.

Craft, N., Shostak, Y., Carey, M. & Sawyers, C.L. (1999) 'A mechanism for hormone-independent prostate cancer through modulation of androgen receptor signaling by the HER-2/neu tyrosine kinase', *Nature Medicine*, 5(3), pp. 280–285.

Crawford, E.D. (2004) 'Hormonal Therapy in Prostate Cancer: Historical Approaches', *Reviews in Urology*, 6(Suppl 7), pp. S3–S11.

Crawford, E.D., Eisenberger, M.A., McLeod, D.G., Spaulding, J.T., Benson, R., Dorr, F.A., Blumenstein, B.A., Davis, M.A. & Goodman, P.J. (1989) 'A controlled trial of leuprolide with and without flutamide in prostatic carcinoma', *The New England Journal of Medicine*, 321(7), pp. 419–424.

Crawford, J.B. & Patton, J.G. (2006) 'Activation of alpha-tropomyosin exon 2 is regulated by the SR protein 9G8 and heterogeneous nuclear ribonucleoproteins H and F', *Molecular and Cellular Biology*, 26(23), pp. 8791–8802.

Crona, D.J. & Whang, Y.E. (2017) 'Androgen Receptor-Dependent and -Independent Mechanisms Involved in Prostate Cancer Therapy Resistance', *Cancers*, 9(6), p. 67.

Culig, Z. (2016) 'Androgen Receptor Coactivators in Regulation of Growth and Differentiation in Prostate Cancer', *Journal of Cellular Physiology*, 231(2), pp. 270–274.

Cunha, G.R., Alarid, E.T., Turner, T., Donjacour, A.A., Boutin, E.L. & Foster, B.A. (1992) 'Normal and abnormal development of the male urogenital tract. Role of androgens, mesenchymal–epithelial interactions, and growth factors', *Journal of Andrology*, 13(6), pp. 465–475.

Cunha, G.R., Cao, M., Derpinghaus, A. & Baskin, L.S. (2021) 'Human urogenital sinus mesenchyme is an inducer of prostatic epithelial development', *American Journal of Clinical and Experimental Urology*, 9(4), p. 329.

Cunha, G.R., Vezina, C.M., Isaacson, D., Ricke, W.A., Timms, B.G., Cao, M., Franco, O. & Baskin, L.S. (2018) 'Development of the human prostate', *Differentiation; research in biological diversity*, 103pp. 24–45.

Cutress, M.L., Whitaker, H.C., Mills, I.G., Stewart, M. & Neal, D.E. (2008) 'Structural basis for the nuclear import of the human androgen receptor', *Journal of Cell Science*, 121(7), pp. 957–968.

Dalal, K., Roshan-Moniri, M., Sharma, A., Li, H., Ban, F., Hessein, M., Hsing, M., Singh, K., LeBlanc, E., Dehm, S., Tomlinson Guns, E.S., Cherkasov, A. & Rennie, P.S. (2014) 'Selectively Targeting the DNA-binding Domain of the Androgen Receptor as a Prospective Therapy for Prostate Cancer', *The Journal of Biological Chemistry*, 289(38), pp. 26417–26429.

Danckwardt, S., Hentze, M.W. & Kulozik, A.E. (2008) '3' end mRNA processing: molecular mechanisms and implications for health and disease', *The EMBO Journal*, 27(3), pp. 482–498.

Dauwalder, B., Amaya-Manzanares, F. & Mattox, W. (1996) 'A human homologue of the *Drosophila* sex determination factor *transformer-2* has conserved splicing regulatory functions.', *Proceedings of the National Academy of Sciences of the United States of America*, 93(17), pp. 9004–9009.

Davey, R.A. & Grossmann, M. (2016) 'Androgen Receptor Structure, Function and Biology: From Bench to Bedside', *The Clinical Biochemist Reviews*, 37(1), pp. 3–15.

Davis, C.A., Grate, L., Spingola, M. & Ares Jr, M. (2000) 'Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast', *Nucleic Acids Research*, 28(8), pp. 1700–1706.

Davis, I.D., Martin, A.J., Stockler, M.R., Begbie, S., Chi, K.N., Chowdhury, S., Coskinas, X., Frydenberg, M., Hague, W.E., Horvath, L.G., Joshua, A.M., Lawrence, N.J., Marx, G., McCaffrey, J., McDermott, R., McJannett, M., North, S.A., Parnis, F., Parulekar, W., et al. (2019) 'Enzalutamide with Standard First-Line Therapy in Metastatic Prostate Cancer', *New England Journal of Medicine*, 381(2), pp. 121–131.

De Coster, R., Mahler, C., Denis, L., Coene, M.C., Caers, I., Amery, W., Haelterman, C. & Beerens, D. (1987) 'Effects of high-dose ketoconazole and dexamethasone on ACTH-stimulated adrenal steroidogenesis in orchiectomized prostatic cancer patients', *Acta Endocrinologica*, 115(2), pp. 265–271.

Debes, J.D., Comuzzi, B., Schmidt, L.J., Dehm, S.M., Culig, Z. & Tindall, D.J. (2005) 'p300 regulates androgen receptor-independent expression of prostate-specific antigen in prostate cancer cells treated chronically with interleukin-6', *Cancer Research*, 65(13), pp. 5965–5973.

Debes, J.D., Schmidt, L.J., Huang, H. & Tindall, D.J. (2002) 'p300 mediates androgen-independent transactivation of the androgen receptor by interleukin 6', *Cancer Research*, 62(20), pp. 5632–5636.

Dehm, S.M., Schmidt, L.J., Heemers, H.V., Vessella, R.L. & Tindall, D.J. (2008) 'Splicing of a novel AR exon generates a constitutively active androgen receptor that mediates prostate cancer therapy resistance', *Cancer research*, 68(13), pp. 5469–5477.

Del Re, M., Conteduca, V., Crucitta, S., Gurioli, G., Casadei, C., Restante, G., Schepisi, G., Lolli, C., Cucchiara, F., Danesi, R. & De Giorgi, U. (2021) 'Androgen receptor gain in circulating free DNA and splicing variant 7 in exosomes predict clinical outcome in CRPC patients treated with abiraterone and enzalutamide', *Prostate Cancer and Prostatic Diseases*, 24(2), pp. 524–531.

Denayer, S., Helsen, C., Thorrez, L., Haelens, A. & Claessens, F. (2010) 'The Rules of DNA Recognition by the Androgen Receptor', *Molecular Endocrinology*, 24(5), pp. 898–913.

Deng, Q., Wu, Y., Zhang, Z., Wang, Y., Li, M., Liang, H. & Gui, Y. (2017) 'Androgen Receptor Localizes to Plasma Membrane by Binding to Caveolin-1 in Mouse Sertoli Cells', *International Journal of Endocrinology*, 2017p. e3985916.

Denichenko, P., Mogilevsky, M., Cléry, A., Welte, T., Biran, J., Shimshon, O., Barnabas, G.D., Danan-Gothold, M., Kumar, S., Yavin, E., Levanon, E.Y., Allain, F.H., Geiger, T., Levkowitz, G.

& Karni, R. (2019) 'Specific inhibition of splicing factor activity by decoy RNA oligonucleotides', *Nature Communications*, 10p. 1590.

Diao, Y., Wu, D., Dai, Z., Kang, H., Wang, Z. & Wang, X. (2015) 'Prognostic value of transformer 2 $\beta$  expression in prostate cancer', *International Journal of Clinical and Experimental Pathology*, 8(6), pp. 6967–6973.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T.R. (2013) 'STAR: ultrafast universal RNA-seq aligner', *Bioinformatics (Oxford, England)*, 29(1), pp. 15–21.

Doesburg, P., Kuil, C.W., Berrevoets, C.A., Steketee, K., Faber, P.W., Mulder, E., Brinkmann, A.O. & Trapman, J. (1997) 'Functional in Vivo Interaction between the Amino-Terminal, Transactivation Domain and the Ligand Binding Domain of the Androgen Receptor', *Biochemistry*, 36(5), pp. 1052–1064.

Dong, J., Wu, Z., Wang, D., Pascal, L.E., Nelson, J.B., Wipf, P. & Wang, Z. (2019) HSP70 binds to the androgen receptor N-terminal domain and modulates the receptor function in prostate cancer cells. *Molecular Cancer Therapeutics*. 18 (1), 39–50.

Donjacour, A.A., Thomson, A.A. & Cunha, G.R. (2003) 'FGF-10 plays an essential role in the growth of the fetal prostate', *Developmental Biology*, 261(1), pp. 39–54.

Doudna, J.A. & Charpentier, E. (2014) 'The new frontier of genome engineering with CRISPR-Cas9', *Science*, 346(6213), p. 1258096.

Du, M., Jillette, N., Zhu, J.J., Li, S. & Cheng, A.W. (2020) 'CRISPR artificial splicing factors', *Nature Communications*, 11(1), p. 2973.

Du, Z., Su, H., Wang, W., Ye, L., Wei, H., Peng, Z., Anishchenko, I., Baker, D. & Yang, J. (2021) 'The trRosetta server for fast and accurate protein structure prediction', *Nature Protocols*, 16(12), pp. 5634–5651.

Duan, L., Chen, Z., Lu, Jun, Liang, Y., Wang, M., Roggero, C.M., Zhang, Q.-J., Gao, J., Fang, Y., Cao, J., Lu, Jian, Zhao, H., Dang, A., Pong, R.-C., Hernandez, E., Chang, C.-M., Hoang, D.T., Ahn, J.-M., Xiao, G., et al. (2019) 'Histone lysine demethylase KDM4B regulates the alternative splicing of the androgen receptor in response to androgen deprivation', *Nucleic Acids Research*, 47(22), pp. 11623–11636.

Dubbink, H.J., Hersmus, R., Pike, A.C.W., Molier, M., Brinkmann, A.O., Jenster, G. & Trapman, J. (2006) 'Androgen receptor ligand-binding domain interaction and nuclear receptor specificity of FXXLF and LXXLL motifs as determined by L/F swapping', *Molecular Endocrinology (Baltimore, Md.)*, 20(8), pp. 1742–1755.

Dubbury, S.J., Bourtz, P.L. & Sharp, P.A. (2018) 'CDK12 regulates DNA repair genes by suppressing intronic polyadenylation', *Nature*, 564(7734), pp. 141–145.

Dunn, K.W., Kamocka, M.M. & McDonald, J.H. (2011) 'A practical guide to evaluating colocalization in biological microscopy', *American Journal of Physiology - Cell Physiology*, 300(4), pp. C723–C742.

Dvinge, H. (2018) 'Regulation of alternative mRNA splicing: old players and new perspectives', *FEBS Letters*, 592(17), pp. 2987–3006.

Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. & Hood, L. (1980) 'Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways', *Cell*, 20(2), pp. 313–319.

Eastham, J.A., Stapleton, A.M., Gousse, A.E., Timme, T.L., Yang, G., Slawin, K.M., Wheeler, T.M., Scardino, P.T. & Thompson, T.C. (1995) 'Association of p53 mutations with metastatic prostate cancer', *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 1(10), pp. 1111–1118.

Ebling, F.J.P. (2005) 'The neuroendocrine timing of puberty', *Reproduction*, 129(6), pp. 675–683.

Edwards, J., Krishna, N.S., Grigor, K.M. & Bartlett, J.M.S. (2003) 'Androgen receptor gene amplification and protein expression in hormone refractory prostate cancer', *British Journal of Cancer*, 89(3), pp. 552–556.

Edwards, J., Traynor, P., Munro, A.F., Pirret, C.F., Dunne, B. & Bartlett, J.M.S. (2006) 'The Role of HER1-HER4 and EGFRvIII in Hormone-Refractory Prostate Cancer', *Clinical Cancer Research*, 12(1), pp. 123–130.

Egevad, L., Granfors, T., Karlberg, L., Bergh, A. & Stattin, P. (2002) 'Prognostic value of the Gleason score in prostate cancer', *BJU International*, 89(6), pp. 538–542.

Eisai Inc. (2009) *A Phase I Open-Label, Single-Arm, Dose Escalation Study of E7107 Administered Intravenously on Days 1 and 8 Every 21 Days to Patients With Solid Tumors*.

Eisai Limited (2009) *A Phase I Open-Label, Single-Arm, Dose Escalation Study of E7107 Administered IV (Bolus) on Days 1, 8, and 15 Every 28 Days to Patients With Solid Tumours*.

Eisermann, K., Broderick, C.J., Bazarov, A., Moazam, M.M. & Fraizer, G.C. (2013) 'Androgen up-regulates vascular endothelial growth factor expression in prostate cancer cells via an Sp1 binding site', *Molecular Cancer*, 12(1), p. 7.

Elgehama, A., Sun, L., Yu, B., Guo, W. & Xu, Q. (2021) 'Selective targeting of the androgen receptor-DNA binding domain by the novel antiandrogen SBF-1 and inhibition of the growth of prostate cancer cells', *Investigational New Drugs*, 39(2), pp. 442–457.

Ellwood-Yen, K., Graeber, T.G., Wongvipat, J., Iruela-Arispe, M.L., Zhang, J., Matusik, R., Thomas, G.V. & Sawyers, C.L. (2003) 'Myc-driven murine prostate cancer shares molecular features with human prostate tumors', *Cancer Cell*, 4(3), pp. 223–238.

Emmert-Buck, M.R., Vocke, C.D., Pozzatti, R.O., Duray, P.H., Jennings, S.B., Florence, C.D., Zhuang, Z., Bostwick, D.G., Liotta, L.A. & Linehan, W.M. (1995) 'Allelic Loss on Chromosome 8p12–21 in Microdissected Prostatic Intraepithelial Neoplasia', *Cancer Research*, 55(14), pp. 2959–2962.

Eperon, I.C., Makarova, O.V., Mayeda, A., Munroe, S.H., Cáceres, J.F., Hayward, D.G. & Krainer, A.R. (2000) 'Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1', *Molecular and Cellular Biology*, 20(22), pp. 8303–8318.

Epigenetix, Inc. (2022) *A Phase 1 Study of EP31670, a Dual BET and CBP/p300 Inhibitor in Patients With Targeted Advanced Solid Tumors*.

Erdős, G., Pajkos, M. & Dosztányi, Z. (2021) 'IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation', *Nucleic Acids Research*, 49(W1), pp. W297–W303.

Eri, L.M. & Tveten, K.J. (1993) 'A prospective, placebo-controlled study of the antiandrogen Casodex as treatment for patients with benign prostatic hyperplasia', *The Journal of Urology*, 150(1), pp. 90–94.

Eskens, F.A.L.M., Ramos, F.J., Burger, H., O'Brien, J.P., Piera, A., de Jonge, M.J.A., Mizui, Y., Wiemer, E.A.C., Carreras, M.J., Baselga, J. & Tabernero, J. (2013) 'Phase I Pharmacokinetic and Pharmacodynamic Study of the First-in-Class Spliceosome Inhibitor E7107 in Patients with Advanced Solid Tumors', *Clinical Cancer Research*, 19(22), pp. 6296–6304.

ESSA Pharmaceuticals (2022a) *A Phase 1, Open-label Study to Evaluate the Safety, Pharmacokinetics, and Anti-Tumor Activity of Oral EPI-7386 in Patients With Castration-Resistant Prostate Cancer*.

ESSA Pharmaceuticals (2018) *A Phase 1/2 Open-Label Study to Assess the Safety, Pharmacokinetics, and Anti-Tumor Activity of Oral EPI-506 in Patients With Metastatic Castration-Resistant Prostate Cancer*.

ESSA Pharmaceuticals (2022b) *A Phase 1/2 Study of EPI-7386 in Combination With Enzalutamide Compared With Enzalutamide Alone in Subjects With Metastatic Castration-Resistant Prostate Cancer*.

Ewels, P., Magnusson, M., Lundin, S. & Käller, M. (2016) 'MultiQC: summarize analysis results for multiple tools and samples in a single report', *Bioinformatics (Oxford, England)*, 32(19), pp. 3047–3048.

Ewing, C.M., Ray, A.M., Lange, E.M., Zuhlke, K.A., Robbins, C.M., Tembe, W.D., Wiley, K.E., Isaacs, S.D., Johng, D., Wang, Y., Bizon, C., Yan, G., Gielzak, M., Partin, A.W., Shanmugam, V., Izatt, T., Sinari, S., Craig, D.W., Zheng, S.L., et al. (2012) 'Germline Mutations in HOXB13 and Prostate-Cancer Risk', *The New England journal of medicine*, 366(2), pp. 141–149.

Fairbrother, W.G., Yeh, R.-F., Sharp, P.A. & Burge, C.B. (2002) 'Predictive identification of exonic splicing enhancers in human genes', *Science (New York, N.Y.)*, 297(5583), pp. 1007–1013.

Fan, L., Zhang, F., Xu, S., Cui, X., Hussain, A., Fazli, L., Gleave, M., Dong, X. & Qi, J. (2018) 'Histone demethylase JMJD1A promotes alternative splicing of AR variant 7 (AR-V7) in

prostate cancer cells', *Proceedings of the National Academy of Sciences of the United States of America*, 115(20), pp. E4584–E4593.

Fang, Y., Fliss, A.E., Robins, D.M. & Caplan, A.J. (1996) 'Hsp90 Regulates Androgen Receptor Hormone Binding Affinity in Vivo\*', *Journal of Biological Chemistry*, 271(45), pp. 28697–28702.

Farla, P., Hersmus, R., Trapman, J. & Houtsmuller, A.B. (2005) 'Antiandrogens prevent stable DNA-binding of the androgen receptor', *Journal of Cell Science*, 118(18), pp. 4187–4198.

Fazal, F.M., Han, S., Parker, K.R., Kaewsapsak, P., Xu, J., Boettiger, A.N., Chang, H.Y. & Ting, A.Y. (2019) 'Atlas of Subcellular RNA Localization Revealed by APEX-Seq', *Cell*, 178(2), pp. 473–490.e26.

Fetzer, S., Lauber, J., Will, C.L. & Lührmann, R. (1997) 'The [U4/U6.U5] tri-snRNP-specific 27K protein is a novel SR protein that can be phosphorylated by the snRNP-associated protein kinase', *RNA (New York, N.Y.)*, 3(4), pp. 344–355.

Fizazi, K., Scher, H.I., Molina, A., Logothetis, C.J., Chi, K.N., Jones, R.J., Staffurth, J.N., North, S., Vogelzang, N.J., Saad, F., Mainwaring, P., Harland, S., Goodman, O.B., Sternberg, C.N., Li, J.H., Kheoh, T., Haqq, C.M. & de Bono, J.S. (2012) 'Abiraterone acetate for treatment of metastatic castration-resistant prostate cancer: final overall survival analysis of the COU-AA-301 randomised, double-blind, placebo-controlled phase 3 study', *The Lancet Oncology*, 13(10), pp. 983–992.

Fizazi, K., Shore, N., Tammela, T.L., Ulys, A., Vjaters, E., Polyakov, S., Jievaltas, M., Luz, M., Alekseev, B., Kuss, I., Kappeler, C., Snapir, A., Sarapohja, T. & Smith, M.R. (2019) 'Darolutamide in Nonmetastatic, Castration-Resistant Prostate Cancer', *New England Journal of Medicine*, 380(13), pp. 1235–1246.

Flavin, R., Pettersson, A., Hendrickson, W.K., Fiorentino, M., Finn, S., Kunz, L., Judson, G.L., Lis, R., Bailey, D., Fiore, C., Nuttall, E., Martin, N.E., Stack, E., Penney, K.L., Rider, J.R., Sinnott, J., Sweeney, C., Sesso, H.D., Fall, K., et al. (2014) 'SPINK1 Protein Expression and Prostate Cancer Progression', *Clinical Cancer Research*, 20(18), pp. 4904–4911.

Fleckner, J., Zhang, M., Valcárcel, J. & Green, M.R. (1997) 'U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction.', *Genes & Development*, 11(14), pp. 1864–1872.

Folco, E.G., Coil, K.E. & Reed, R. (2011) 'The anti-tumor drug E7107 reveals an essential role for SF3b in remodeling U2 snRNP to expose the branch point-binding region', *Genes & Development*, 25(5), pp. 440–444.

Förch, P. & Valcárcel, J. (2003) 'Splicing regulation in Drosophila sex determination', *Progress in Molecular and Subcellular Biology*, 31pp. 127–151.

Fourcade, R.-O. & McLeod, D. (2004) 'Tolerability of Antiandrogens in the Treatment of Prostate Cancer', *UroOncology*, 4(1), pp. 5–13.

Francis, J.C. & Swain, A. (2018) 'Prostate Organogenesis', *Cold Spring Harbor Perspectives in Medicine*, 8(7), p. a030353.

Frigo, D.E., Sherk, A.B., Wittmann, B.M., Norris, J.D., Wang, Q., Joseph, J.D., Toner, A.P., Brown, M. & McDonnell, D.P. (2009) 'Induction of Krüppel-Like Factor 5 Expression by Androgens Results in Increased CXCR4-Dependent Migration of Prostate Cancer Cells in Vitro', *Molecular Endocrinology*, 23(9), pp. 1385–1396.

Frønsdal, K., Engedal, N., Slagsvold, T. & Saatcioglu, F. (1998) 'CREB Binding Protein Is a Coactivator for the Androgen Receptor and Mediates Cross-talk with AP-1\*', *Journal of Biological Chemistry*, 273(48), pp. 31853–31859.

Gacci, M., Corona, G., Vignozzi, L., Salvi, M., Serni, S., De Nunzio, C., Tubaro, A., Oelke, M., Carini, M. & Maggi, M. (2015) 'Metabolic syndrome and benign prostatic enlargement: a systematic review and meta-analysis', *BJU international*, 115(1), pp. 24–31.

Gaddipati, J.P., McLeod, D.G., Heidenberg, H.B., Sesterhenn, I.A., Finger, M.J., Moul, J.W. & Srivastava, S. (1994) 'Frequent Detection of Codon 877 Mutation in the Androgen Receptor Gene in Advanced Prostate Cancers1', *Cancer Research*, 54(11), pp. 2861–2864.

Gagliardi, M. & Matarazzo, M.R. (2016) 'RIP: RNA Immunoprecipitation', in Chiara Lanzuolo & Beatrice Bodega (eds.) *Polycomb Group Proteins: Methods and Protocols*. Methods in Molecular Biology. [Online]. New York, NY: Springer. pp. 73–86.

Gandaglia, G., Abdollah, F., Schiffmann, J., Trudeau, V., Shariat, S.F., Kim, S.P., Perrotte, P., Montorsi, F., Briganti, A., Trinh, Q.-D., Karakiewicz, P.I. & Sun, M. (2014) 'Distribution of metastatic sites in patients with prostate cancer: A population-based analysis', *The Prostate*, 74(2), pp. 210–216.

Gandaglia, G., Briganti, A., Gontero, P., Mondaini, N., Novara, G., Salonia, A., Sciarra, A. & Montorsi, F. (2013) 'The role of chronic prostatic inflammation in the pathogenesis and progression of benign prostatic hyperplasia (BPH)', *BJU international*, 112(4), pp. 432–441.

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C. & Schultz, N. (2013) 'Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal', *Science Signaling*, 6(269), pp. pl1–pl1.

Gao, X.D., Rodríguez, T.C. & Sontheimer, E.J. (2019) 'Adapting dCas9-APEX2 for subnuclear proteomic profiling', *Methods in enzymology*, 616pp. 365–383.

Gao, X.D., Tu, L.-C., Mir, A., Rodriguez, T., Ding, Y., Leszyk, J., Dekker, J., Shaffer, S.A., Zhu, L.J., Wolfe, S.A. & Sontheimer, E.J. (2018) 'C-BERST: Defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2', *Nature methods*, 15(6), pp. 433–436.

Gao, L., Zhang, W., Zhang, J., Liu, J., Sun, F., Liu, H., Hu, J., Wang, X., Xin, Wang, Xueli, Su, P., Chen, S., Qu, S., Shi, B., Xiong, X., Chen, W., Dong, X. & Han, B. (2021) KIF15-mediated stabilization of Ar and ar-V7 contributes to enzalutamide resistance in prostate cancer. *Cancer Research*. 81 (4), 1026–1039.

Gelmann, E. (2002) 'Molecular Biology of the Androgen Receptor', *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 20pp. 3001–15.

Geng, C., Rajapakshe, K., Shah, S.S., Shou, J., Eedunuri, V.K., Foley, C., Fiskus, W., Rajendran, M., Chew, S.A., Zimmermann, M., Bond, R., He, B., Coarfa, C. & Mitsiades, N. (2014) 'Androgen receptor is the key transcriptional mediator of the tumor suppressor SPOP in prostate cancer', *Cancer research*, 74(19), pp. 5631–5643.

Geuens, T., Bouhy, D. & Timmerman, V. (2016) 'The hnRNP family: insights into their role in health and disease', *Human Genetics*, 135pp. 851–867.

Gillingham, A.K., Bertram, J., Begum, F. & Munro, S. (2019) 'In vivo identification of GTPase interactors by mitochondrial relocalization and proximity biotinylation', *eLife*, 8p. e45916.

Gim, H.J., Park, J., Jung, M.E. & Houk, K.N. (2021) 'Conformational dynamics of androgen receptors bound to agonists and antagonists', *Scientific Reports*, 11(1), p. 15887.

Glazko, G. & Mushegian, A. (2010) 'Measuring gene expression divergence: the distance to keep', *Biology Direct*, 5p. 51.

Gleason, D.F. (1966) 'Classification of prostatic carcinomas', *Cancer Chemotherapy Reports*, 50(3), pp. 125–128.

Gleason, D.F. & Mellinger, G.T. (1974) 'Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging', *The Journal of Urology*, 111(1), pp. 58–64.

Godoy, A.S., Sotomayor, P.C., Villagran, M., Yacoub, R., Montecinos, V.P., McNerney, E.M., Moser, M., Foster, B.A. & Onate, S.A. (2012) 'Altered corepressor SMRT expression and recruitment to target genes as a mechanism that change the response to androgens in prostate cancer progression', *Biochemical and Biophysical Research Communications*, 423(3), pp. 564–570.

Golan-Gerstl, R., Cohen, M., Shilo, A., Suh, S.-S., Bakàcs, A., Coppola, L. & Karni, R. (2011) 'Splicing factor hnRNP A2/B1 regulates tumor suppressor gene splicing and is an oncogenic driver in glioblastoma', *Cancer Research*, 71(13), pp. 4464–4472.

Goldspiel, B.R. & Kohler, D.R. (1990) 'Flutamide: an antiandrogen for advanced prostate cancer', *DICP: the annals of pharmacotherapy*, 24(6), pp. 616–623.

Gomella, L.G. (2009) 'Effective Testosterone Suppression for Prostate Cancer: Is There a Best Castration Therapy?', *Reviews in Urology*, 11(2), pp. 52–60.

Gordon, V., Bhadel, S., Wunderlich, W., Zhang, J., Ficarro, S.B., Mollah, S.A., Shabanowitz, J., Hunt, D.F., Xenarios, I., Hahn, W.C., Conaway, M., Carey, M.F. & Gioeli, D. (2010) 'CDK9 regulates AR promoter selectivity and cell growth through serine 81 phosphorylation', *Molecular Endocrinology (Baltimore, Md.)*, 24(12), pp. 2267–2280.

Gormley, G.J., Stoner, E., Bruskewitz, R.C., Imperato-McGinley, J., Walsh, P.C., McConnell, J.D., Andriole, G.L., Geller, J., Bracken, B.R., Tenover, J.S., Vaughan, E.D., Pappas, F., Taylor, A.,

Binkowitz, B. & Ng, J. (1992) 'The Effect of Finasteride in Men with Benign Prostatic Hyperplasia', *New England Journal of Medicine*, 327(17), pp. 1185–1191.

Goulet, A. & Cambillau, C. (2022) 'Present Impact of AlphaFold2 Revolution on Structural Biology, and an Illustration With the Structure Prediction of the Bacteriophage J-1 Host Adhesion Device', *Frontiers in Molecular Biosciences*, 9.

Grant, M.K.O., Shapiro, S.L., Ashe, K.H., Liu, P. & Zahs, K.R. (2019) 'A Cautionary Tale: Endogenous Biotinylated Proteins and Exogenously-Introduced Protein A Cause Antibody-Independent Artefacts in Western Blot Studies of Brain-Derived Proteins', *Biological Procedures Online*, 21(1), p. 6.

Grasso, C.S., Wu, Y.-M., Robinson, D.R., Cao, X., Dhanasekaran, S.M., Khan, A.P., Quist, M.J., Jing, X., Lonigro, R.J., Brenner, J.C., Asangani, I.A., Ateeq, B., Chun, S.Y., Siddiqui, J., Sam, L., Anstett, M., Mehra, R., Prensner, J.R., Palanisamy, N., et al. (2012) 'The Mutational Landscape of Lethal Castrate Resistant Prostate Cancer', *Nature*, 487(7406), p. 239.

Graw, S., Tang, J., Zafar, M.K., Byrd, A.K., Bolden, C., Peterson, E.C. & Byrum, S.D. (2020) 'proteiNorm – A User-Friendly Tool for Normalization and Analysis of TMT and Label-Free Protein Quantification', *ACS Omega*, 5(40), pp. 25625–25633.

Gregory, C.W., He, B., Johnson, R.T., Ford, O.H., Mohler, J.L., French, F.S. & Wilson, E.M. (2001) 'A mechanism for androgen receptor-mediated prostate cancer recurrence after androgen deprivation therapy', *Cancer Research*, 61(11), pp. 4315–4319.

Gregory, C.W., He, B. & Wilson, E.M. (2001) 'The putative androgen receptor-A form results from in vitro proteolysis', *Journal of Molecular Endocrinology*, 27(3), pp. 309–319.

Gregory, C.W., Whang, Y.E., McCall, W., Fei, X., Liu, Y., Ponguta, L.A., French, F.S., Wilson, E.M. & Earp, H.S. (2005) 'Heregulin-induced activation of HER2 and HER3 increases androgen receptor transactivation and CWR-R1 human recurrent prostate cancer cell growth', *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 11(5), pp. 1704–1712.

Grellscheid, S., Dalgiesh, C., Storbeck, M., Best, A., Liu, Y., Jakubik, M., Mende, Y., Ehrmann, I., Curk, T., Rossbach, K., Bourgeois, C.F., Stévenin, J., Grellscheid, D., Jackson, M.S., Wirth, B. & Elliott, D.J. (2011) 'Identification of Evolutionarily Conserved Exons as Regulated Targets for the Splicing Activator Tra2 $\beta$  in Development', *PLoS Genetics*, 7(12), p. e1002390.

Grellscheid, S.N., Dalgiesh, C., Rozanska, A., Grellscheid, D., Bourgeois, C.F., Stévenin, J. & Elliott, D.J. (2011) 'Molecular design of a splicing switch responsive to the RNA binding protein Tra2 $\beta$ ', *Nucleic Acids Research*, 39(18), pp. 8092–8104.

Grignon, D.J. (2004) 'Unusual subtypes of prostate cancer', *Modern Pathology*, 17(3), pp. 316–327.

Grino, P.B., Griffin, J.E. & Wilson, J.D. (1990) 'Testosterone at high concentrations interacts with the human androgen receptor similarly to dihydrotestosterone', *Endocrinology*, 126(2), pp. 1165–1172.

Groth, D., Hartmann, S., Klie, S. & Selbig, J. (2013) 'Principal components analysis', *Methods in Molecular Biology* (Clifton, N.J.), 930pp. 527–547.

Guiochon-Mantel, A., Delabre, K., Lescop, P. & Milgrom, E. (1996) 'Intracellular traffic of steroid hormone receptors', *The Journal of Steroid Biochemistry and Molecular Biology*, 56(1), pp. 3–9.

Guo, H.-B., Perminov, A., Bekele, S., Kedziora, G., Farajollahi, S., Varaljay, V., Hinkle, K., Molinero, V., Meister, K., Hung, C., Dennis, P., Kelley-Loughnane, N. & Berry, R. (2022) 'AlphaFold2 models indicate that protein sequence determines both structure and dynamics', *Scientific Reports*, 12.

Guo, Z., Dai, B., Jiang, T., Xu, K., Xie, Y., Kim, O., Nesheiwat, I., Kong, X., Melamed, J., Handratta, V.D., Njar, V.C.O., Brodie, A.M.H., Yu, L.-R., Veenstra, T.D., Chen, H. & Qiu, Y. (2006) 'Regulation of androgen receptor activity by tyrosine phosphorylation', *Cancer Cell*, 10(4), pp. 309–319.

Guo, Z., Yang, X., Sun, F., Jiang, R., Linn, D.E., Chen, Hege, Chen, Hegang, Kong, X., Melamed, J., Tepper, C.G., Kung, H.-J., Brodie, A.M.H., Edwards, J. & Qiu, Y. (2009) 'A Novel Androgen Receptor Splice Variant Is Up-regulated during Prostate Cancer Progression and Promotes Androgen Depletion–Resistant Growth', *Cancer Research*, 69(6), pp. 2305–2313.

Gupta, R., Somyajit, K., Narita, T., Maskey, E., Stanlie, A., Kremer, M., Typas, D., Lammers, M., Mailand, N., Nussenzweig, A., Lukas, J. & Choudhary, C. (2018) 'DNA Repair Network Analysis Reveals Shieldin as a Key Regulator of NHEJ and PARP Inhibitor Sensitivity', *Cell*, 173(4), pp. 972–988.e23.

Gurel, B., Iwata, T., Koh, C., Jenkins, R.B., Lan, F., Van Dang, C., Hicks, J.L., Morgan, J., Cornish, T.C., Sutcliffe, S., Isaacs, W.B., Luo, J. & De Marzo, A.M. (2008) 'Nuclear MYC Protein Overexpression is an Early Alteration in Human Prostate Carcinogenesis', *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, 21(9), pp. 1156–1167.

Haapala, K., Hytytin, E.-R., Roiha, M., Laurila, M., Rantala, I., Helin, H.J. & Koivisto, P.A. (2001) 'Androgen Receptor Alterations in Prostate Cancer Relapsed during a Combined Androgen Blockade by Orchiectomy and Bicalutamide', *Laboratory Investigation*, 81(12), pp. 1647–1651.

Haelens, A., Tanner, T., Denayer, S., Callewaert, L. & Claessens, F. (2007) 'The Hinge Region Regulates DNA Binding, Nuclear Translocation, and Transactivation of the Androgen Receptor', *Cancer Research*, 67(9), pp. 4514–4523.

Hägglöf, C. & Bergh, A. (2012) 'The Stroma—A Key Regulator in Prostate Function and Malignancy', *Cancers*, 4(2), pp. 531–548.

Hammerich, K., Ayala, G. & Wheeler, T. (2008) 'Anatomy of the prostate gland and surgical pathology of prostate cancer', *Prostate Cancer*,

Han, J., Xiong, J., Wang, D. & Fu, X.-D. (2011) 'Pre-mRNA splicing: where and when in the nucleus', *Trends in cell biology*, 21(6), pp. 336–343.

Han, S., Udeshi, N.D., Deerinck, T.J., Svinkina, T., Ellisman, M.H., Carr, S.A. & Ting, A.Y. (2017) 'Proximity Biotinylation as a Method for Mapping Proteins Associated with mtDNA in Living Cells', *Cell chemical biology*, 24(3), pp. 404–414.

Han, S., Zhao, B.S., Myers, S.A., Carr, S.A., He, C. & Ting, A.Y. (2020) 'RNA-protein interaction mapping via MS2- or Cas13-based APEX targeting', *Proceedings of the National Academy of Sciences of the United States of America*, 117(36), pp. 22068–22079.

Hanahan, D. (2022) 'Hallmarks of Cancer: New Dimensions', *Cancer Discovery*, 12(1), pp. 31–46.

Hanahan, D. & Weinberg, R.A. (2011) 'Hallmarks of Cancer: The Next Generation', *Cell*, 144(5), pp. 646–674.

Hanahan, D. & Weinberg, R.A. (2000) 'The Hallmarks of Cancer', *Cell*, 100(1), pp. 57–70.

Hara, T., Kouno, J., Nakamura, K., Kusaka, M. & Yamaoka, M. (2005) 'Possible role of adaptive mutation in resistance to antiandrogen in prostate cancer cells', *The Prostate*, 65(3), pp. 268–275.

Hara, T., Miyazaki, J., Araki, H., Yamaoka, M., Kanzaki, N., Kusaka, M. & Miyamoto, M. (2003) 'Novel mutations of androgen receptor: a possible mechanism of bicalutamide withdrawal syndrome', *Cancer Research*, 63(1), pp. 149–153.

Harrison, J.H., Thorn, G.W. & Jenkins, D. (1953) 'Total Adrenalectomy for Reactivated Carcinoma of the Prostate', *New England Journal of Medicine*, 248(3), pp. 86–92.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., Barnes, I., Bignell, A., Boychenko, V., Hunt, T., Kay, M., Mukherjee, G., Rajan, J., Despacio-Reyes, G., Saunders, G., et al. (2012) 'GENCODE: the reference human genome annotation for The ENCODE Project', *Genome Research*, 22(9), pp. 1760–1774.

Hartley, S.W. & Mullikin, J.C. (2016) 'Detection and visualization of differential splicing in RNA-Seq data with JunctionSeq', *Nucleic Acids Research*, 44(15), p. e127.

Hartley, S.W. & Mullikin, J.C. (2015) 'QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments', *BMC bioinformatics*, 16(1), p. 224.

Hasegawa, M., Miura, T., Kuzuya, K., Inoue, A., Won Ki, S., Horinouchi, S., Yoshida, T., Kunoh, T., Koseki, K., Mino, K., Sasaki, R., Yoshida, M. & Mizukami, T. (2011) 'Identification of SAP155 as the Target of GEX1A (Herboxidiene), an Antitumor Natural Product', *ACS Chemical Biology*, 6(3), pp. 229–233.

He, B., Gampe, R.T., Kole, A.J., Hnat, A.T., Stanley, T.B., An, G., Stewart, E.L., Kalman, R.I., Minges, J.T. & Wilson, E.M. (2004) 'Structural Basis for Androgen Receptor Interdomain and

Coactivator Interactions Suggests a Transition in Nuclear Receptor Activation Function Dominance', *Molecular Cell*, 16(3), pp. 425–438.

He, B., Kemppainen, J.A. & Wilson, E.M. (2000) 'FXXLF and WXXLF Sequences Mediate the NH<sub>2</sub>-terminal Interaction with the Ligand Binding Domain of the Androgen Receptor \*', *Journal of Biological Chemistry*, 275(30), pp. 22986–22994.

Heber, D., Dodson, R., Stoskopf, C., Peterson, M. & Swerdloff, R.S. (1982) 'Pituitary desensitization and the regulation of pituitary gonadotropin-releasing hormone (GnRH) receptors following chronic administration of a superactive GnRH analog and testosterone', *Life Sciences*, 30(26), pp. 2301–2308.

Heery, D.M., Kalkhoven, E., Hoare, S. & Parker, M.G. (1997) 'A signature motif in transcriptional co-activators mediates binding to nuclear receptors', *Nature*, 387(6634), pp. 733–736.

Hemavant Sciences GmbH (2022) *An Open-label, Multicenter Phase 1 Trial to Evaluate the Safety, Pharmacokinetics and Pharmacodynamics of Splicing Modulator H3B-8800 for Subjects With Myelodysplastic Syndromes, Acute Myeloid Leukemia, and Chronic Myelomonocytic Leukemia*.

Hernández, J., Bechara, E., Schlesinger, D., Delgado, J., Serrano, L. & Valcárcel, J. (2016) 'Tumor suppressor properties of the splicing regulatory factor RBM10', *RNA biology*, 13(4), pp. 466–472.

Herzel, L., Ottoz, D.S.M., Alpert, T. & Neugebauer, K.M. (2017) 'Splicing and transcription touch base: co-transcriptional spliceosome assembly and function', *Nature Reviews Molecular Cell Biology*, 18(10), pp. 637–650.

Higano, C. (2019) Enzalutamide, apalutamide, or darolutamide: Are apples or bananas best for patients? *Nature Reviews Urology*. 16 (6), 335–336.

Higdon, R. & Kolker, E. (2007) 'A predictive model for identifying proteins by a single peptide match', *Bioinformatics*, 23(3), pp. 277–280.

Hirst, C.J., Cabrera, C. & Kirby, M. (2012) 'Epidemiology of castration resistant prostate cancer: A longitudinal analysis using a UK primary care database', *Cancer Epidemiology*, 36(6), pp. e349–e353.

Hoang, D.T., Iczkowski, K.A., Kilari, D., See, W. & Nevalainen, M.T. (2016) 'Androgen receptor-dependent and -independent mechanisms driving prostate cancer progression: Opportunities for therapeutic targeting from multiple angles', *Oncotarget*, 8(2), pp. 3724–3745.

Hobson, B.D., Choi, S.J., Mosharov, E.V., Soni, R.K., Sulzer, D. & Sims, P.A. (2022) 'Subcellular proteomics of dopamine neurons in the mouse brain' Andrew B West & Gary L Westbrook (eds.), *eLife*, 11p. e70921.

Hohmann, T. & Dehghani, F. (2019) 'The Cytoskeleton—A Complex Interacting Meshwork', *Cells*, 8(4), p. 362.

Hong, D.S., Kurzrock, R., Naing, A., Wheler, J.J., Falchook, G.S., Schiffman, J.S., Faulkner, N., Pilat, M.J., O'Brien, J. & LoRusso, P. (2014) 'A phase I, open-label, single-arm, dose-escalation study of E7107, a precursor messenger ribonucleic acid (pre-mRNA) splicesome inhibitor administered intravenously on days 1 and 8 every 21 days to patients with solid tumors', *Investigational New Drugs*, 32(3), pp. 436–444.

Hong, N.H., Moigne, R.L., Banuelos, C.A., Mawji, N.R., Tam, T., Wang, J., Andersen, R.J., Cesano, A., Sadar, M.D., Zhou, H.-J. & Virsik, P. (2020) 'Abstract 1953: Pre-clinical development of the second-generation N-terminal domain androgen receptor inhibitor, EPI-7386, for the treatment of prostate cancer', *Cancer Research*, 80(16\_Supplement), p. 1953.

Hoogland, A.M., Kweldam, C.F. & van Leenders, G.J.L.H. (2014) 'Prognostic Histopathological and Molecular Markers on Prostate Cancer Needle-Biopsies: A Review', *BioMed Research International*, 2014, p. 341324.

Horn, T., Gosliga, A., Li, C., Enculescu, M. & Legewie, S. (2023) 'Position-dependent effects of RNA-binding proteins in the context of co-transcriptional splicing', *npj Systems Biology and Applications*, 9(1), pp. 1–22.

Hsu, F.-N., Chen, M.-C., Chiang, M.-C., Lin, E., Lee, Y.-T., Huang, P.-H., Lee, G.-S. & Lin, H. (2011) 'Regulation of androgen receptor and prostate cancer growth by cyclin-dependent kinase 5', *The Journal of Biological Chemistry*, 286(38), pp. 33141–33149.

Hu, R., Dunn, T.A., Wei, S., Isharwal, S., Veltri, R.W., Humphreys, E., Han, M., Partin, A.W., Vessella, R.L., Isaacs, W.B., Bova, G.S. & Luo, J. (2009) 'Ligand-independent androgen receptor variants derived from splicing of cryptic exons signify hormone-refractory prostate cancer', *Cancer Research*, 69(1), pp. 16–22.

Hu, R., Isaacs, W.B. & Luo, J. (2011) 'A snapshot of the expression signature of androgen receptor splicing variants and their distinctive transcriptional activities', *The Prostate*, 71(15), pp. 1656–1667.

Hu, R., Lu, C., Mostaghel, E.A., Yegnasubramanian, S., Gurel, M., Tannahill, C., Edwards, J., Isaacs, W.B., Nelson, P.S., Bluemn, E., Plymate, S.R. & Luo, J. (2012) 'Distinct Transcriptional Programs Mediated by the Ligand-Dependent Full-Length Androgen Receptor and Its Splice Variants in Castration-Resistant Prostate Cancer', *Cancer Research*, 72(14), pp. 3457–3462.

Huang, J., Zhao, R., Qin, S., Yang, S., Li, W., Mo, J., Wang, F., Du, Y., Weng, X. & Zhou, X. (2021) '4-Thiouridine-Enhanced Peroxidase-Generated Biotinylation of RNA', *ChemBioChem*, 22(1), pp. 212–216.

Huang, S., Gulzar, Z.G., Salari, K., Lapointe, J., Brooks, J.D. & Pollack, J.R. (2012) 'Recurrent deletion of CHD1 in prostate cancer with relevance to cell invasiveness', *Oncogene*, 31(37), pp. 4164–4170.

Huang, W., Shostak, Y., Tarr, P., Sawyers, C. & Carey, M. (1999) 'Cooperative Assembly of Androgen Receptor into a Nucleoprotein Complex That Regulates the Prostate-specific Antigen Enhancer\*', *Journal of Biological Chemistry*, 274(36), pp. 25756–25768.

Hudson, D.L., Guy, A.T., Fry, P., O'Hare, M.J., Watt, F.M. & Masters, J.R. (2001) 'Epithelial cell differentiation pathways in the human prostate: identification of intermediate phenotypes by keratin expression', *The Journal of Histochemistry and Cytochemistry: Official Journal of the Histochemistry Society*, 49(2), pp. 271–278.

Huggins, C. & Hodges, C.V. (1941) 'Studies on Prostatic Cancer. I. The Effect of Castration, of Estrogen and of Androgen Injection on Serum Phosphatases in Metastatic Carcinoma of the Prostate\*', *Cancer Research*, 1(4), pp. 293–297.

Huggins, C., Stevens, R.E., Jr. & Hodges, C.V. (1941) 'STUDIES ON PROSTATIC CANCER: II. THE EFFECTS OF CASTRATION ON ADVANCED CARCINOMA OF THE PROSTATE GLAND', *Archives of Surgery*, 43(2), pp. 209–223.

Hui, J., Stangl, K., Lane, W.S. & Bindereif, A. (2003) 'HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats', *Nature Structural Biology*, 10(1), pp. 33–37.

Humphrey, P.A. (2004) 'Gleason grading and prognostic factors in carcinoma of the prostate', *Modern Pathology*, 17(3), pp. 292–306.

Hung, V., Lam, S.S., Udeshi, N.D., Svinkina, T., Guzman, G., Mootha, V.K., Carr, S.A. & Ting, A.Y. (2017) 'Proteomic mapping of cytosol-facing outer mitochondrial and ER membranes in living human cells by proximity biotinylation' David Pagliarini (ed.), *eLife*, 6p. e24463.

Hung, V., Udeshi, N.D., Lam, S.S., Loh, K.H., Cox, K.J., Pedram, K., Carr, S.A. & Ting, A.Y. (2016) 'Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2', *Nature protocols*, 11(3), pp. 456–475.

Hunter, I., Hay, C.W., Esswein, B., Watt, K. & McEwan, I.J. (2018) 'Tissue control of androgen action: The ups and downs of androgen receptor expression', *Molecular and Cellular Endocrinology*, 465pp. 27–35.

Hussain, S.A., Maroto, P., Climent, M.Á., Bianchini, D., Jones, R.H., Lin, C.-C., Wang, S.-S., Dean, E., Crossley, K., Schlieker, L., Bogenrieder, T. & De Bono, J.S. (2019) 'Targeting IGF-1/2 with xentuzumab (Xe) plus enzalutamide (En) in metastatic castration-resistant prostate cancer (mCRPC) after progression on docetaxel chemotherapy (DCt) and abiraterone (Abi): Randomized phase II trial results.', *Journal of Clinical Oncology*, 37(15\_suppl), pp. 5030–5030.

Huynh, N., Depner, N., Larson, R. & King-Jones, K. (2020) 'A versatile toolkit for CRISPR-Cas13-based RNA manipulation in *Drosophila*', *Genome Biology*, 21(1), p. 279.

Iborra, S., Hirschfeld, M., Jaeger, M., Zur Hausen, A., Braicu, I., Sehouli, J., Gitsch, G. & Stickeler, E. (2013) 'Alterations in expression pattern of splicing factors in epithelial ovarian cancer and its clinical impact', *International Journal of Gynecological Cancer: Official Journal of the International Gynecological Cancer Society*, 23(6), pp. 990–996.

Ibrahim, E.C., Schaal, T.D., Hertel, K.J., Reed, R. & Maniatis, T. (2005) 'Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers', *Proceedings of the National Academy of Sciences of the United States of America*, 102(14), pp. 5002–5007.

Imamura, Y. & Sadar, M.D. (2016) 'Androgen receptor targeted therapies in castration-resistant prostate cancer: Bench to clinic', *International Journal of Urology*, 23(8), pp. 654–665.

Ino, Y., Arakawa, N., Ishiguro, H., Uemura, H., Kubota, Y., Hirano, H. & Toda, T. (2016) 'Phosphoproteome analysis demonstrates the potential role of THRAP3 phosphorylation in androgen-independent prostate cancer cell growth', *PROTEOMICS*, 16(7), pp. 1069–1078.

Irvine, R.A., Yu, M.C., Ross, R.K. & Coetzee, G.A. (1995) 'The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer', *Cancer Research*, 55(9), pp. 1937–1940.

Isaacs, J.T. (1984) 'Antagonistic effect of androgen on prostatic cell death', *The Prostate*, 5(5), pp. 545–557.

Isaacs, J.T., D'Antonio, J.M., Chen, S., Antony, L., Dalrymple, S.P., Ndikuyeze, G.H., Luo, J. & Denmeade, S.R. (2012) 'Adaptive Auto-Regulation of Androgen Receptor Provides a Paradigm Shifting Rationale for Bipolar Androgen Therapy (BAT) for Castrate Resistant Human Prostate Cancer', *The Prostate*, 72(14), pp. 1491–1505.

Isaacs, W.B., Carter, B.S. & Ewing, C.M. (1991) 'Wild-Type p53 Suppresses Growth of Human Prostate Cancer Cells Containing Mutant p53 Alleles1', *Cancer Research*, 51(17), pp. 4716–4720.

Ishida, T. & Kinoshita, K. (2007) 'PrDOS: prediction of disordered protein regions from amino acid sequence', *Nucleic Acids Research*, 35(Web Server issue), pp. W460-464.

Ittmann, M. (2018) 'Anatomy and Histology of the Human and Murine Prostate', *Cold Spring Harbor Perspectives in Medicine*, 8(5), p. a030346.

Iwata, T., Schultz, D., Hicks, J., Hubbard, G.K., Mutton, L.N., Lotan, T.L., Bethel, C., Lotz, M.T., Yegnasubramanian, S., Nelson, W.G., Dang, C.V., Xu, M., Anele, U., Koh, C.M., Bieberich, C.J. & De Marzo, A.M. (2010) 'MYC Overexpression Induces Prostatic Intraepithelial Neoplasia and Loss of Nkx3.1 in Mouse Luminal Epithelial Cells', *PLoS ONE*, 5(2), p. e9427.

Jamaspishvili, T., Berman, D.M., Ross, A.E., Scher, H.I., De Marzo, A.M., Squire, J.A. & Lotan, T.L. (2018) 'Clinical implications of PTEN loss in prostate cancer', *Nature reviews. Urology*, 15(4), pp. 222–234.

James, N.D., Sydes, M.R., Clarke, N.W., Mason, M.D., Dearnaley, D.P., Spears, M.R., Ritchie, A.W.S., Parker, C.C., Russell, J.M., Attard, G., de Bono, J., Cross, W., Jones, R.J., Thalmann, G., Amos, C., Matheson, D., Millman, R., Alzoueibi, M., Beesley, S., et al. (2016) 'Addition of docetaxel, zoledronic acid, or both to first-line long-term hormone therapy in prostate cancer (STAMPEDE): survival results from an adaptive, multiarm, multistage, platform randomised controlled trial', *Lancet (London, England)*, 387(10024), pp. 1163–1177.

Jamison, S.F., Pasman, Z., Wang, J., Will, C., Lührmann, R., Manley, J.L. & Garcia-Blanco, M.A. (1995) 'U1 snRNP-ASF/SF2 interaction and 5' splice site recognition: characterization of required elements.', *Nucleic Acids Research*, 23(16), pp. 3260–3267.

Jamros, M.A., Aubol, B.E., Keshwani, M.M., Zhang, Z., Stamm, S. & Adams, J.A. (2015) 'Intra-domain Cross-talk Regulates Serine-arginine Protein Kinase 1-dependent Phosphorylation and Splicing Function of Transformer 2 $\beta$ 1\*', *Journal of Biological Chemistry*, 290(28), pp. 17269–17281.

Jenster, G., van der Korput, H.A., Trapman, J. & Brinkmann, A.O. (1995) 'Identification of two transcription activation units in the N-terminal domain of the human androgen receptor', *The Journal of Biological Chemistry*, 270(13), pp. 7341–7346.

Jeong, S. (2017) 'SR Proteins: Binders, Regulators, and Connectors of RNA', *Molecules and Cells*, 40(1), pp. 1–9.

Jernberg, E., Bergh, A. & Wikström, P. (2017) 'Clinical relevance of androgen receptor alterations in prostate cancer', *Endocrine Connections*, 6(8), pp. R146–R161.

Jiang, F. & Doudna, J.A. (2017) 'CRISPR–Cas9 Structures and Mechanisms', *Annual Review of Biophysics*, 46(1), pp. 505–529.

Jiang, G., Stalewski, J., Galyean, R., Dykert, J., Schteingart, C., Broqua, P., Aebi, A., Aubert, M.L., Semple, G., Robson, P., Akinsanya, K., Haigh, R., Riviere, P., Trojnar, J., Junien, J.L. & Rivier, J.E. (2001) 'GnRH antagonists: a new generation of long acting analogues incorporating p-ureido-phenylalanines at positions 5 and 6', *Journal of Medicinal Chemistry*, 44(3), pp. 453–467.

Jimenez-Vacas, J., Montero-Hidalgo, A., Gómez, E., Saez-Martinez, P., Perez-Gomez, J., Fuentes-Fayos, A., Blázquez-Encinas, R., Sanchez-Sánchez, R., Gonzalez-Serrano, T., Castro, E., López-Soto, P., Carrasco-Valiente, J., Sarmento-Cabral, A., Martinez-Fuentes, A., Eyras, E., Castaño, J., Sharp, A., Olmos, D., Gahete, M., et al. (2023) *The splicing factor SRSF6 regulates AR activity and represents a potential therapeutic target in prostate cancer*.

Jin, H.-J., Kim, J. & Yu, J. (2013) 'Androgen receptor genomic regulation', *Translational Andrology and Urology*, 2(3), pp. 158–177.

Johnson, D.B. & Sonthalia, S. (2022) 'Flutamide', in *StatPearls*. [Online]. Treasure Island (FL): StatPearls Publishing.

Jones, A.L. & Chinegwundoh, F. (2014) 'Update on prostate cancer in black men within the UK', *ecancermedicalscience*, 8p. 455.

Jones, D., Noble, M., Wedge, S.R., Robson, C.N. & Gaughan, L. (2017) 'Aurora A regulates expression of AR-V7 in models of castrate resistant prostate cancer', *Scientific Reports*, 7p. 40957.

Jorquerá, R., Ortiz, R., Ossandon, F., Cárdenas, J.P., Sepúlveda, R., González, C. & Holmes, D.S. (2016) 'SinEx DB: a database for single exon coding sequences in mammalian genomes', *Database: The Journal of Biological Databases and Curation*, 2016p. baw095.

Joseph, J.D., Lu, N., Qian, J., Sensintaffar, J., Shao, G., Brigham, D., Moon, M., Maneval, E.C., Chen, I., Darimont, B. & Hager, J.H. (2013) 'A clinically relevant androgen receptor mutation

confers resistance to second-generation antiandrogens enzalutamide and ARN-509', *Cancer Discovery*, 3(9), pp. 1020–1029.

Judge, A.D., Bola, G., Lee, A.C.H. & MacLachlan, I. (2006) 'Design of Noninflammatory Synthetic siRNA Mediating Potent Gene Silencing in Vivo', *Molecular Therapy*, 13(3), pp. 494–505.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., et al. (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature*, 596(7873), pp. 583–589.

Jurica, M.S. & Moore, M.J. (2003) 'Pre-mRNA Splicing: Awash in a Sea of Proteins', *Molecular Cell*, 12(1), pp. 5–14.

Juul, A., Müller, J. & Skakkebaek, N.E. (1997) 'Prostate specific antigen in boys with precocious puberty before and during gonadal suppression by GnRH agonist treatment', *European Journal of Endocrinology*, 136(4), pp. 401–405.

Kaida, D., Motoyoshi, H., Tashiro, E., Nojima, T., Hagiwara, M., Ishigami, K., Watanabe, H., Kitahara, T., Yoshida, T., Nakajima, H., Tani, T., Horinouchi, S. & Yoshida, M. (2007) 'Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA', *Nature Chemical Biology*, 3(9), pp. 576–583.

Kallio, H.M.L., Hieta, R., Latonen, L., Brofeldt, A., Annala, M., Kivinummi, K., Tammela, T.L., Nykter, M., Isaacs, W.B., Lilja, H.G., Bova, G.S. & Visakorpi, T. (2018) 'Constitutively active androgen receptor splice variants AR-V3, AR-V7 and AR-V9 are co-expressed in castration-resistant prostate cancer metastases', *British Journal of Cancer*, 119(3), pp. 347–356.

Kallio, P.J., Poukka, H., Moilanen, A., Jänne, O.A. & Palvimo, J.J. (1995) 'Androgen receptor-mediated transcriptional regulation in the absence of direct interaction with a specific DNA element.', *Molecular Endocrinology*, 9(8), pp. 1017–1028.

Kandels-Lewis, S. & Séraphin, B. (1993) 'Involvement of U6 snRNA in 5' Splice Site Selection', *Science*, 262(5142), pp. 2035–2039.

Kang, H.-Y., Cho, C.-L., Huang, K.-L., Wang, J.-C., Hu, Y.-C., Lin, H.-K., Chang, C. & Huang, K.-E. (2004) 'Nongenomic Androgen Activation of Phosphatidylinositol 3-Kinase/Akt Signaling Pathway in MC3T3-E1 Osteoblasts', *Journal of Bone and Mineral Research*, 19(7), pp. 1181–1190.

Kanopka, A., Mühlemann, O. & Akusjärvi, G. (1996) 'Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA', *Nature*, 381(6582), pp. 535–538.

Karantanos, T., Corn, P.G. & Thompson, T.C. (2013) 'Prostate cancer progression after androgen deprivation therapy: mechanisms of castrate resistance and novel therapeutic approaches', *Oncogene*, 32(49), pp. 5501–5511.

Kasper, S. (2008) 'Exploring the Origins of the Normal Prostate and Prostate Cancer Stem Cell', *Stem Cell Reviews*, 4(3), p. 193.

Kawamura, N., Nimura, K., Saga, K., Ishibashi, A., Kitamura, K., Nagano, H., Yoshikawa, Y., Ishida, K., Nonomura, N., Arisawa, M., Luo, J. & Kaneda, Y. (2019) 'SF3B2-Mediated RNA Splicing Drives Human Prostate Cancer Progression', *Cancer Research*, 79(20), pp. 5204–5217.

Kelley, M.L., Strezoska, Ž., He, K., Vermeulen, A. & Smith, A. van B. (2016) 'Versatility of chemically synthesized guide RNAs for CRISPR-Cas9 genome editing', *Journal of Biotechnology*, 233pp. 74–83.

Kellokumpu-Lehtinen, P., Santti, R. & Pelliniemi, L.J. (1980) 'Correlation of early cytodifferentiation of the human fetal prostate and Leydig cells', *The Anatomical Record*, 196(3), pp. 263–273.

Kemppainen, J.A., Lane, M.V., Sar, M. & Wilson, E.M. (1992) 'Androgen receptor phosphorylation, turnover, nuclear transport, and transcriptional activation. Specificity for steroids and antihormones', *The Journal of Biological Chemistry*, 267(2), pp. 968–974.

Khan, K.A., Ng, M.K. & Cheung, P. (2020) 'The Use of Mononucleosome Immunoprecipitation for Analysis of Combinatorial Histone Post-translational Modifications and Purification of Nucleosome-Interacting Proteins', *Frontiers in Cell and Developmental Biology*, 8.

Kido, K., Yamanaka, S., Nakano, S., Motani, K., Shinohara, S., Nozawa, A., Kosako, H., Ito, S. & Sawasaki, T. (2020) 'AirID, a novel proximity biotinylation enzyme, for analysis of protein–protein interactions' Philip A Cole, Volker Dötsch, Volker Dötsch, & John E Cronan (eds.), *eLife*, 9p. e54983.

Kim, D.I., Jensen, S.C., Noble, K.A., Kc, B., Roux, K.H., Motamedchaboki, K. & Roux, K.J. (2016) 'An improved smaller biotin ligase for BioID proximity labeling', *Molecular Biology of the Cell*, 27(8), pp. 1188–1196.

Kim, H.-J., Park, J.-W., Cho, Y.-S., Cho, C.-H., Kim, J.-S., Shin, H.-W., Chung, D.H., Kim, S.J. & Chun, Y.-S. (2013) 'Pathogenic role of HIF-1 $\alpha$  in prostate hyperplasia in the presence of chronic inflammation', *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1832(1), pp. 183–194.

Kim, M.S., Je, E.M., Oh, J.E., Yoo, N.J. & Lee, S.H. (2013) 'Mutational and expressional analyses of SPOP, a candidate tumor suppressor gene, in prostate, gastric and colorectal cancers', *APMIS*, 121(7), pp. 626–633.

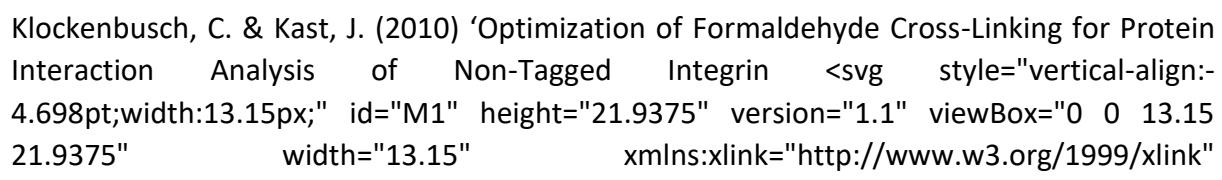
Kim, S., Au, C.C., Jamalruddin, M.A.B., Abou-Ghali, N.E., Mukhtar, E., Portella, L., Berger, A., Worroll, D., Vatsa, P., Rickman, D.S., Nanus, D.M. & Giannakakou, P. (2022) 'AR-V7 exhibits non-canonical mechanisms of nuclear import and chromatin engagement in castrate-resistant prostate cancer', *eLife*, 11p. e73396.

Kim, W., Zhang, L., Wilton, J.H., Fetterly, G., Mohler, J.L., Weinberg, V., Morse, A., Szmulewitz, R.Z., Friedlander, T.W., Fong, L., Lin, A.M., Harzstark, A.L., Molina, A., Small, E.J. & Ryan, C.J. (2014) 'Sequential Use of the Androgen Synthesis Inhibitors Ketoconazole and Abiraterone Acetate in Castration-Resistant Prostate Cancer and the Predictive Value of Circulating Androgens', *Clinical Cancer Research*, 20(24), pp. 6269–6276.

Kinter, K.J. & Anekar, A.A. (2022) 'Biochemistry, Dihydrotestosterone', in *StatPearls*. [Online]. Treasure Island (FL): StatPearls Publishing.

Kirby, M., Hirst, C. & Crawford, E.D. (2011) 'Characterising the castration-resistant prostate cancer population: a systematic review', *International Journal of Clinical Practice*, 65(11), pp. 1180–1192.

Klausen, M.S., Jespersen, M.C., Nielsen, H., Jensen, K.K., Jurtz, V.I., Sønderby, C.K., Sommer, M.O.A., Winther, O., Nielsen, M., Petersen, B. & Marcatili, P. (2019) 'NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning', *Proteins*, 87(6), pp. 520–527.

Klockenbusch, C. & Kast, J. (2010) 'Optimization of Formaldehyde Cross-Linking for Protein Interaction Analysis of Non-Tagged Integrin  width="13.15" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns="http://www.w3.org/2000/svg" style="vertical-align: 4.698pt; width: 13.15px; height: 21.9375px;"/> 1', *BioMed Research International*, 2010p. e927585.

Klotz, L., Boccon-Gibod, L., Shore, N.D., Andreou, C., Persson, B.-E., Cantor, P., Jensen, J.-K., Olesen, T.K. & Schröder, F.H. (2008) 'The efficacy and safety of degarelix: a 12-month, comparative, randomized, open-label, parallel-group phase III study in patients with prostate cancer', *BJU International*, 102(11), pp. 1531–1538.

Klotz, L. & Schellhammer, P. (2005) 'Combined androgen blockade: the case for bicalutamide', *Clinical Prostate Cancer*, 3(4), pp. 215–219.

Koh, C.M., Bieberich, C.J., Dang, C.V., Nelson, W.G., Yegnasubramanian, S. & De Marzo, A.M. (2010) 'MYC and Prostate Cancer', *Genes & Cancer*, 1(6), pp. 617–628.

Koivisto, P., Kononen, J., Palmberg, C., Tammela, T., Hytinen, E., Isola, J., Trapman, J., Cleutjens, K., Noordzij, A., Visakorpi, T. & Kallioniemi, O.-P. (1997) 'Androgen Receptor Gene Amplification: A Possible Molecular Mechanism for Androgen Deprivation Therapy Failure in Prostate Cancer1', *Cancer Research*, 57(2), pp. 314–319.

Kokontis, J., Takakura, K., Hay, N. & Liao, S. (1994) 'Increased Androgen Receptor Activity and Altered c-myc Expression in Prostate Cancer Cells after Long-Term Androgen Deprivation1', *Cancer Research*, 54(6), pp. 1566–1573.

Kolde, R. (2019) *pheatmap: Pretty Heatmaps*.

Kolvenbag, G., Furr, B.J.A. & Blackledge, G.R.P. (1998) 'Receptor affinity and potency of non-steroidal antiandrogens: translation of preclinical findings into clinical activity', *Prostate Cancer and Prostatic Diseases*, 1(6), pp. 307–314.

Kondo, Y., Oubridge, C., van Roon, A.-M.M. & Nagai, K. (2015) 'Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition' Timothy Nilsen (ed.), *eLife*, 4p. e04986.

Konermann, S., Lotfy, P., Brideau, N.J., Oki, J., Shokhirev, M.N. & Hsu, P.D. (2018) 'Transcriptome engineering with RNA-targeting Type VI-D CRISPR effectors', *Cell*, 173(3), pp. 665–676.e14.

Korpal, M., Korn, J.M., Gao, X., Rakiec, D.P., Ruddy, D.A., Doshi, S., Yuan, J., Kovats, S.G., Kim, S., Cooke, V.G., Monahan, J.E., Stegmeier, F., Roberts, T.M., Sellers, W.R., Zhou, W. & Zhu, P. (2013) 'An F876L mutation in androgen receptor confers genetic and phenotypic resistance to MDV3100 (enzalutamide)', *Cancer Discovery*, 3(9), pp. 1030–1043.

Kotake, Y., Sagane, K., Owa, T., Mimori-Kiyosue, Y., Shimizu, H., Uesugi, M., Ishihama, Y., Iwata, M. & Mizui, Y. (2007) 'Splicing factor SF3b as a target of the antitumor natural product pladienolide', *Nature Chemical Biology*, 3(9), pp. 570–575.

Kounatidou, E., Nakjang, S., McCracken, S.R.C., Dehm, S.M., Robson, C.N., Jones, D. & Gaughan, L. (2019) 'A novel CRISPR-engineered prostate cancer cell line defines the AR-V transcriptome and identifies PARP inhibitor sensitivities', *Nucleic Acids Research*, 47(11), pp. 5634–5647.

Kousteni, S., Bellido, T., Plotkin, L.I., O'Brien, C.A., Bodenner, D.L., Han, L., Han, K., DiGregorio, G.B., Katzenellenbogen, J.A., Katzenellenbogen, B.S., Roberson, P.K., Weinstein, R.S., Jilka, R.L. & Manolagas, S.C. (2001) 'Nongenotropic, Sex-Nonspecific Signaling through the Estrogen or Androgen Receptors: Dissociation from Transcriptional Activity', *Cell*, 104(5), pp. 719–730.

Krajewska, M., Dries, R., Grassetti, A.V., Dust, S., Gao, Y., Huang, H., Sharma, B., Day, D.S., Kwiatkowski, N., Pomaville, M., Dodd, O., Chipumuro, E., Zhang, T., Greenleaf, A.L., Yuan, G.-C., Gray, N.S., Young, R.A., Geyer, M., Gerber, S.A., et al. (2019) 'CDK12 loss in cancer cells affects DNA damage response genes through premature cleavage and polyadenylation', *Nature Communications*, 10(1), p. 1757.

Krey, J.F., Scheffer, D.I., Choi, D., Reddy, A., David, L.L., Corey, D.P. & Barr-Gillespie, P.G. (2018) 'Mass spectrometry quantitation of proteins from small pools of developing auditory and vestibular cells', *Scientific Data*, 5(1), p. 180128.

Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K. & Moult, J. (2021) 'Critical Assessment of Methods of Protein Structure Prediction (CASP) – Round XIV', *Proteins*, 89(12), pp. 1607–1617.

Kubitz, L., Bitsch, S., Zhao, X., Schmitt, K., Deweid, L., Roehrig, A., Barazzzone, E.C., Valerius, O., Kolmar, H. & Béthune, J. (2022) 'Engineering of ultraID, a compact and hyperactive enzyme for proximity-dependent biotinylation in living cells', *Communications Biology*, 5p. 657.

Kumar, A., Clerici, M., Muckenfuss, L.M., Passmore, L.A. & Jinek, M. (2019) Mechanistic insights into mRNA 3'-end processing. *Current Opinion in Structural Biology*, 59, 143–150.

Kumar, R., Zakharov, M.N., Khan, S.H., Miki, R., Jang, H., Toraldo, G., Singh, R., Bhasin, S. & Jasuja, R. (2011) 'The Dynamic Structure of the Estrogen Receptor', *Journal of Amino Acids*, 2011, p. e812540.

Kumar, V.L. & Majumder, P.K. (1995) 'Prostate gland: Structure, functions and regulation', *International Urology and Nephrology*, 27(3), pp. 231–243.

Kurita, T., Medina, R.T., Mills, A.A. & Cunha, G.R. (2004) 'Role of p63 and basal cells in the prostate', *Development (Cambridge, England)*, 131(20), pp. 4955–4964.

Kushawah, G., Hernandez-Huertas, L., Abugattas-Nuñez del Prado, J., Martinez-Morales, J.R., DeVore, M.L., Hassan, H., Moreno-Sanchez, I., Tomas-Gallardo, L., Diaz-Moscoso, A., Monges, D.E., Guelfo, J.R., Theune, W.C., Brannan, E.O., Wang, W., Corbin, T.J., Moran, A.M., Sánchez Alvarado, A., Málaga-Trillo, E., Takacs, C.M., et al. (2020) 'CRISPR-Cas13d Induces Efficient mRNA Knockdown in Animal Embryos', *Developmental Cell*, 54(6), pp. 805–817.e7.

Kučák, D., Melo, L., Schroeder, F., Jelic-Matošević, Z., Mutter, N., Bertoša, B. & Barišić, I. (2022) 'CATANA: an online modelling environment for proteins and nucleic acid nanostructures', *Nucleic Acids Research*, 50(W1), pp. W152–W158.

Kuznik, N.C., Solozobova, V., Jung, N., Gräßle, S., Lei, Q., Lewandowski, E.M., Munuganti, R., Zoubeidi, A., Chen, Y., Bräse, S. & Cato, A.C.B. (2021) 'Development of a Benzothiazole Scaffold-Based Androgen Receptor N-Terminal Inhibitor for Treating Androgen-Responsive Prostate Cancer', *ACS Chemical Biology*, 16(11), pp. 2103–2108.

Kwok, J., O'Shea, M., Hume, D.A. & Lengeling, A. (2017) 'Jmjd6, a JmjC Dioxygenase with Many Interaction Partners and Pleiotropic Functions', *Frontiers in Genetics*, 8.

Kwon, Y.W., Jo, H.-S., Bae, S., Seo, Y., Song, P., Song, M. & Yoon, J.H. (2021) 'Application of Proteomics in Cancer: Recent Trends and Approaches for Biomarkers Discovery', *Frontiers in Medicine*, 8.

Kyprianou, N. & Isaacs, J.T. (1988) 'Activation of Programmed Cell Death in the Rat Ventral Prostate after Castration\*', *Endocrinology*, 122(2), pp. 552–562.

Labbé, D.P., Sweeney, C.J., Brown, M., Galbo, P., Rosario, S., Wadosky, K.M., Ku, S.-Y., Sjöström, M., Alshalalfa, M., Erho, N., Davicioni, E., Karnes, R.J., Schaeffer, E.M., Jenkins, R.B., Den, R.B., Ross, A.E., Bowden, M., Huang, Y., Gray, K.P., et al. (2017) 'TOP2A and EZH2 Provide Early Detection of an Aggressive Prostate Cancer Subgroup', *Clinical Cancer Research*, 23(22), pp. 7072–7083.

Labrie, F. (1993) 'Mechanism of action and pure antiandrogenic properties of flutamide', *Cancer*, 72(12 Suppl), pp. 3816–3827.

Labrie, F., Dupont, A., Belanger, A., Cusan, L., Lacourciere, Y., Monfette, G., Laberge, J.G., Emond, J.P., Fazekas, A.T., Raynaud, J.P. & Husson, J.M. (1982) 'New hormonal therapy in

prostatic carcinoma: combined treatment with an LHRH agonist and an antiandrogen', *Clinical and Investigative Medicine. Medecine Clinique Et Experimentale*, 5(4), pp. 267–275.

Labrie, F., Luu-The, V., Labrie, C. & Simard, J. (2001) 'DHEA and its transformation into androgens and estrogens in peripheral target tissues: introcrinology', *Frontiers in Neuroendocrinology*, 22(3), pp. 185–212.

Lai, K.-P., Yamashita, S., Vitkus, S., Shyr, C.-R., Yeh, S. & Chang, C. (2012) 'Suppressed Prostate Epithelial Development with Impaired Branching Morphogenesis in Mice Lacking Stromal Fibromuscular Androgen Receptor', *Molecular Endocrinology*, 26(1), pp. 52–66.

Lallous, N., Snow, O., Sanchez, C., Parra Nuñez, A.K., Sun, B., Hussain, A., Lee, J., Morin, H., Leblanc, E., Gleave, M.E. & Cherkasov, A. (2021) 'Evaluation of Darolutamide (ODM201) Efficiency on Androgen Receptor Mutants Reported to Date in Prostate Cancer Patients', *Cancers*, 13(12), p. 2939.

Lam, B.J., Bakshi, A., Ekinci, F.Y., Webb, J., Graveley, B.R. & Hertel, K.J. (2003) 'Enhancer-dependent 5'-Splice Site Control of fruitless Pre-mRNA Splicing \*', *Journal of Biological Chemistry*, 278(25), pp. 22740–22747.

Lam, S.S., Martell, J.D., Kamer, K.J., Deerinck, T.J., Ellisman, M.H., Mootha, V.K. & Ting, A.Y. (2015) 'Directed evolution of APEX2 for electron microscopy and proteomics', *Nature methods*, 12(1), pp. 51–54.

Lamb, C.A., Nühlen, S., Judith, D., Frith, D., Snijders, A.P., Behrends, C. & Tooze, S.A. (2016) 'TBC1D14 regulates autophagy via the TRAPP complex and ATG9 traffic', *The EMBO Journal*, 35(3), pp. 281–301.

Lara, P.N., Longmate, J., Evans, C.P., Quinn, D.I., Twardowski, P., Chattha, G., Posadas, E., Stadler, W. & Gandara, D.R. (2009) 'A phase II trial of the Src-kinase inhibitor AZD0530 in patients with advanced castration-resistant prostate cancer: a California Cancer Consortium Study', *Anti-cancer drugs*, 20(3), pp. 179–184.

Latil, A., Bièche, I., Vidaud, D., Lidereau, R., Berthon, P., Cussenot, O. & Vidaud, M. (2001) 'Evaluation of Androgen, Estrogen (ER $\alpha$  and ER $\beta$ ), and Progesterone Receptor Expression in Human Prostate Cancer by Real-Time Quantitative Reverse Transcription-Polymerase Chain Reaction Assays', *Cancer Research*, 61(5), pp. 1919–1926.

Lavery, D.N. & McEwan, I.J. (2008) 'Structural characterization of the native NH<sub>2</sub>-terminal transactivation domain of the human androgen receptor: a collapsed disordered conformation underlies structural plasticity and protein-induced folding', *Biochemistry*, 47(11), pp. 3360–3369.

Lazar, C., Gatto, L., Ferro, M., Bruley, C. & Burger, T. (2016) 'Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies', *Journal of Proteome Research*, 15(4), pp. 1116–1125.

Le Moigne, R., Pearson, P., Lauriault, V., Hong, N.H., Virsik, P., Zhou, H.-J. & Cesano, A. (2021) 'Preclinical and clinical pharmacology of EPI-7386, an androgen receptor N-terminal domain

inhibitor for castration-resistant prostate cancer.', *Journal of Clinical Oncology*, 39(6\_suppl), pp. 119–119.

Le Moigne, R., Zhou, H.-J., Obst, J.K., Banuelos, C.A., Jian, K., Williams, D., Virsik, P., Andersen, R.J., Sadar, M., Perabo, F. & Chi, K.N. (2019) 'Lessons learned from the metastatic castration-resistant prostate cancer phase I trial of EPI-506, a first-generation androgen receptor N-terminal domain inhibitor.', *Journal of Clinical Oncology*, 37(7\_suppl), pp. 257–257.

Leclair, N.K., Brugiolo, M., Urbanski, L., Lawson, S.C., Thakar, K., Yurieva, M., George, J., Hinson, J.T., Cheng, A., Graveley, B.R. & Anczuków, O. (2020) 'POISON EXON SPLICING REGULATES A COORDINATED NETWORK OF SR PROTEIN EXPRESSION DURING DIFFERENTIATION AND TUMORIGENESIS', *Molecular cell*, 80(4), pp. 648-665.e9.

Lee, H.-G., Jo, J., Hong, H.-H., Kim, K.K., Park, J.-K., Cho, S.-J. & Park, C. (2016) 'State-of-the-art housekeeping proteins for quantitative western blotting: Revisiting the first draft of the human proteome', *Proteomics*, 16(13), pp. 1863–1867.

Lee, I.I., Kuznik, N.C., Rottenberg, J.T., Brown, M. & Cato, A.C. (2019) Bag1l: A promising therapeutic target for androgen receptor-dependent prostate cancer. *Journal of Molecular Endocrinology*. 62 (4), .

van Leenders, G.J.L.H., Gage, W.R., Hicks, J.L., van Balken, B., Aalders, T.W., Schalken, J.A. & De Marzo, A.M. (2003) 'Intermediate Cells in Human Prostate Epithelium Are Enriched in Proliferative Inflammatory Atrophy', *The American Journal of Pathology*, 162(5), pp. 1529–1537.

Lei, Q., Jiao, J., Xin, L., Chang, C.-J., Wang, S., Gao, J., Gleave, M.E., Witte, O.N., Liu, X. & Wu, H. (2006) 'NKX3.1 stabilizes p53, inhibits AKT activation, and blocks prostate cancer initiation caused by PTEN loss', *Cancer Cell*, 9(5), pp. 367–378.

Leitzmann, M.F. & Rohrmann, S. (2012) 'Risk factors for the onset of prostatic cancer: age, location, and behavioral correlates', *Clinical Epidemiology*, 4pp. 1–11.

Leslie, S.W., Soon-Sutton, T.L., Sajjad, H. & Siref, L.E. (2022) 'Prostate Cancer', in *StatPearls*. [Online]. Treasure Island (FL): StatPearls Publishing.

Leung, J.K. & Sadar, M.D. (2017) 'Non-Genomic Actions of the Androgen Receptor in Prostate Cancer', *Frontiers in Endocrinology*, 8p. 2.

Levine, A.C., Wang, J.P., Ren, M., Eliashvili, E., Russell, D.W. & Kirschenbaum, A. (1996) 'Immunohistochemical localization of steroid 5 alpha-reductase 2 in the human male fetal reproductive tract and adult prostate', *The Journal of Clinical Endocrinology and Metabolism*, 81(1), pp. 384–389.

Li, C., Kato, M., Shiue, L., Shively, J.E., Ares, M., Jr. & Lin, R.-J. (2006) 'Cell Type and Culture Condition-Dependent Alternative Splicing in Human Breast Cancer Cells Revealed by Splicing-Sensitive Microarrays', *Cancer Research*, 66(4), pp. 1990–1999.

Li, Jiexin, Chen, Z., Chen, F., Xie, G., Ling, Y., Peng, Y., Lin, Y., Luo, N., Chiang, C.-M. & Wang, H. (2020) 'Targeted mRNA demethylation using an engineered dCas13b-ALKH5 fusion protein', *Nucleic Acids Research*, 48(10), pp. 5684–5694.

Li, Junsheng, Lu, M., Zhang, P., Hou, E., Li, T., Liu, X., Xu, X., Wang, Z., Fan, Y., Zhen, X., Li, R., Liu, P., Yu, Y., Hang, J. & Qiao, J. (2020) 'Aberrant spliceosome expression and altered alternative splicing events correlate with maturation deficiency in human oocytes', *Cell Cycle*, 19(17), pp. 2182–2194.

Li, Jianzhuo, Fu, X., Cao, S., Li, Jing, Xing, S., Li, D., Dong, Y., Cardin, D., Park, H.-W., Mauvais-Jarvis, F. & Zhang, H. (2018) 'Membrane-associated androgen receptor (AR) potentiates its transcriptional activities by activating heat shock protein 27 (HSP27)', *Journal of Biological Chemistry*, 293(33), pp. 12719–12729.

Li, M. & Smyth, G.K. (2023) 'Neither random nor censored: estimating intensity-dependent probabilities for missing values in label-free proteomics', *Bioinformatics*, 39(5), p. btad200.

Li, R., Zou, Z., Wang, W. & Zou, P. (2022) 'Metabolic incorporation of electron-rich ribonucleosides enhances APEX-seq for profiling spatially restricted nascent transcriptome', *Cell Chemical Biology*, 29(7), pp. 1218-1231.e8.

Li, X., Zhou, J., Zhao, W., Wen, Q., Wang, W., Peng, H., Gao, Y., Bouchonville, K.J., Offer, S.M., Chan, K., Wang, Z., Li, N. & Gan, H. (2022) 'Defining Proximity Proteome of Histone Modifications by Antibody-mediated Protein A-APEX2 Labeling', *Genomics, Proteomics & Bioinformatics*, 20(1), pp. 87–100.

Li, Y., Chan, S.C., Brand, L.J., Hwang, T.H., Silverstein, K.A.T. & Dehm, S.M. (2013) 'Androgen receptor splice variants mediate enzalutamide resistance in castration-resistant prostate cancer cell lines', *Cancer research*, 73(2), pp. 483–489.

Li, Y., Donmez, N., Sahinalp, C., Xie, N., Wang, Yuwei, Xue, H., Mo, F., Beltran, H., Gleave, M., Wang, Yuzhuo, Collins, C. & Dong, X. (2017) 'SRRM4 Drives Neuroendocrine Transdifferentiation of Prostate Adenocarcinoma Under Androgen Receptor Pathway Inhibition', *European Urology*, 71(1), pp. 68–78.

Li, Y., Xu, J., Guo, X., Li, Z., Cao, L., Liu, S., Guo, Y., Wang, G., Luo, Y., Zhang, Z., Wei, X., Zhao, Y., Liu, T., Wang, X., Xia, H., Kuang, M., Guo, Q., Li, J., Chen, L., et al. (2023) 'The collateral activity of RfxCas13d can induce lethality in a RfxCas13d knock-in mouse model', *Genome Biology*, 24(1), p. 20.

Li, Z., Adams, R.M., Chourey, K., Hurst, G.B., Hettich, R.L. & Pan, C. (2012) 'Systematic Comparison of Label-Free, Metabolic Labeling, and Isobaric Chemical Labeling for Quantitative Proteomics on LTQ Orbitrap Velos', *Journal of Proteome Research*, 11(3), pp. 1582–1590.

Liang, W.-W. & Cheng, S.-C. (2015) 'A novel mechanism for Prp5 function in prespliceosome formation and proofreading the branch site sequence', *Genes & Development*, 29(1), pp. 81–93.

Liao, R.S., Ma, S., Miao, L., Li, R., Yin, Y. & Raj, G.V. (2013) 'Androgen receptor-mediated non-genomic regulation of prostate cancer cell proliferation', *Translational Andrology and Urology*, 2(3), pp. 187–196.

Liao, Y., Smyth, G.K. & Shi, W. (2014) 'featureCounts: an efficient general purpose program for assigning sequence reads to genomic features', *Bioinformatics (Oxford, England)*, 30(7), pp. 923–930.

Libertini, S.J., Tepper, C.G., Rodriguez, V., Asmuth, D.M., Kung, H.-J. & Mudryj, M. (2007) 'Evidence for Calpain-Mediated Androgen Receptor Cleavage as a Mechanism for Androgen Independence', *Cancer Research*, 67(19), pp. 9001–9005.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P. & Tamayo, P. (2015) 'The Molecular Signatures Database Hallmark Gene Set Collection', *Cell Systems*, 1(6), pp. 417–425.

Lin, C., Jan, Y., Kuo, L., Wang, B., Huo, C., Jiang, S.S., Chen, S., Kuo, Y., Chang, C. & Chuu, C. (2018) 'Elevation of androgen receptor promotes prostate cancer metastasis by induction of epithelial-mesenchymal transition and reduction of KAT5', *Cancer Science*, 109(11), pp. 3564–3574.

Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B., Ohgi, K., Zhang, J., Rose, D.W., Fu, X.-D., Glass, C.K. & Rosenfeld, M.G. (2009) 'Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer', *Cell*, 139(6), pp. 1069–1083.

Lin, T.-T., Chen, Y.-H., Wu, Y.-P., Chen, S.-Z., Li, X.-D., Lin, Y.-Z., Chen, S.-H., Zheng, Q.-S., Wei, Y., Xu, N. & Xue, X.-Y. (2019) 'Risk factors for progression to castration-resistant prostate cancer in metastatic prostate cancer patients', *Journal of Cancer*, 10(22), pp. 5608–5613.

Lin, X., Fonseca, M.A.S., Breunig, J.J., Corona, R.I. & Lawrenson, K. (2021) 'In vivo discovery of RNA proximal proteins via proximity-dependent biotinylation', *RNA biology*, 18(12), pp. 2203–2217.

Lin, X., Tascilar, M., Lee, W.-H., Vles, W.J., Lee, B.H., Veeraswamy, R., Asgari, K., Freije, D., van Rees, B., Gage, W.R., Bova, G.S., Isaacs, W.B., Brooks, J.D., DeWeese, T.L., De Marzo, A.M. & Nelson, W.G. (2001) 'GSTP1 CpG Island Hypermethylation Is Responsible for the Absence of GSTP1 Expression in Human Prostate Cancer Cells', *The American Journal of Pathology*, 159(5), pp. 1815–1826.

Linja, M.J., Savinainen, K.J., Saramäki, O.R., Tammela, T.L.J., Vessella, R.L. & Visakorpi, T. (2001) 'Amplification and Overexpression of Androgen Receptor Gene in Hormone-Refractory Prostate Cancer1', *Cancer Research*, 61(9), pp. 3550–3555.

Liu, A.Y. & True, L.D. (2002) 'Characterization of Prostate Cell Types by CD Cell Surface Molecules', *The American Journal of Pathology*, 160(1), pp. 37–43.

Liu, C., Armstrong, C., Zhu, Y., Lou, W. & Gao, A.C. (2016) 'Niclosamide enhances abiraterone treatment via inhibition of androgen receptor variants in castration resistant prostate cancer', *Oncotarget*, 7(22), pp. 32210–32220.

Liu, F., Zhang, J.Z.H. & Mei, Y. (2016) 'The origin of the cooperativity in the streptavidin-biotin system: A computational investigation through molecular dynamics simulations', *Scientific Reports*, 6(1), p. 27190.

Liu, G., Zhu, M., Zhang, M. & Pan, F. (2023) 'Emerging Role of IGF-1 in Prostate Cancer: A Promising Biomarker and Therapeutic Target', *Cancers*, 15(4), p. 1287.

Liu, H.-L., Zhong, H.-Y., Song, T.-Q. & Li, J.-Z. (2017) 'A Molecular Modeling Study of the Hydroxyflutamide Resistance Mechanism Induced by Androgen Receptor Mutations', *International Journal of Molecular Sciences*, 18(9), p. 1823.

Liu, H., Han, R., Li, J., Liu, Huanxiang & Zheng, L. (2016) 'Molecular mechanism of R-bicalutamide switching from androgen receptor antagonist to agonist induced by amino acid mutations using molecular dynamics simulations and free energy calculation', *Journal of Computer-Aided Molecular Design*, 30(12), pp. 1189–1200.

Liu, L.L., Xie, N., Sun, S., Plymate, S., Mostaghel, E. & Dong, X. (2014) 'Mechanisms of the androgen receptor splicing in prostate cancer cells', *Oncogene*, 33(24), pp. 3140–3150.

Liu, X., Han, W., Gulla, S., Simon, N.I., Gao, Y., Cai, C., Yang, H., Zhang, X., Liu, J., Balk, S.P. & Chen, S. (2015) 'Protein phosphatase 1 suppresses androgen receptor ubiquitylation and degradation', *Oncotarget*, 7(2), pp. 1754–1764.

Liu, Y., Zhou, J. & White, K.P. (2014) 'RNA-seq differential expression studies: more sequence or more replication?', *Bioinformatics*, 30(3), pp. 301–304.

Livak, K.J. & Schmittgen, T.D. (2001) 'Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method', *Methods (San Diego, Calif.)*, 25(4), pp. 402–408.

Locke, J.A., Guns, E.S., Lubik, A.A., Adomat, H.H., Hendy, S.C., Wood, C.A., Ettinger, S.L., Gleave, M.E. & Nelson, C.C. (2008) 'Androgen Levels Increase by Intratumoral De novo Steroidogenesis during Progression of Castration-Resistant Prostate Cancer', *Cancer Research*, 68(15), pp. 6407–6415.

Logothetis, C.J. & Lin, S.-H. (2005) 'Osteoblasts in prostate cancer metastasis to bone', *Nature Reviews Cancer*, 5(1), pp. 21–28.

Long, Y., Sou, W.H., Yung, K.W.Y., Liu, H., Wan, S.W.C., Li, Q., Zeng, C., Law, C.O.K., Chan, G.H.C., Lau, T.C.K. & Ngo, J.C.K. (2019) 'Distinct mechanisms govern the phosphorylation of different SR protein splicing factors', *Journal of Biological Chemistry*, 294(4), pp. 1312–1327.

Lopez, S.M., Agoulnik, A.I., Zhang, M., Peterson, L.E., Suarez, E., Gandarillas, G.A., Frolov, A., Li, R., Rajapakshe, K., Coarfa, C., Ittmann, M.M., Weigel, N.L. & Agoulnik, I.U. (2016) 'Nuclear Receptor Corepressor 1 expression and output declines with prostate cancer progression', *Clinical cancer research : an official journal of the American Association for Cancer Research*, 22(15), pp. 3937–3949.

Lorenzo, P.I. & Saatcioglu, F. (2008) 'Inhibition of Apoptosis in Prostate Cancer Cells by Androgens Is Mediated through Downregulation of c-Jun N-terminal Kinase Activation', *Neoplasia (New York, N.Y.)*, 10(5), pp. 418–428.

Louie, M.C., Yang, H.Q., Ma, A.-H., Xu, W., Zou, J.X., Kung, H.-J. & Chen, H.-W. (2003) 'Androgen-induced recruitment of RNA polymerase II to a nuclear receptor–p160 coactivator complex', *Proceedings of the National Academy of Sciences of the United States of America*, 100(5), pp. 2226–2230.

Love, M.I., Huber, W. & Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), p. 550.

Lu, M.L., Schneider, M.C., Zheng, Y., Zhang, X. & Richie, J.P. (2001) 'Caveolin-1 interacts with androgen receptor. A positive modulator of androgen receptor mediated transactivation', *The Journal of Biological Chemistry*, 276(16), pp. 13442–13451.

Lu, S., Liu, M., Epner, D.E., Tsai, S.Y. & Tsai, M.J. (1999) 'Androgen regulation of the cyclin-dependent kinase inhibitor p21 gene through an androgen response element in the proximal promoter', *Molecular Endocrinology (Baltimore, Md.)*, 13(3), pp. 376–384.

Lu, Z., Williamson, S.R., Carskadon, S., Arachchige, P.D., Dhamdhare, G., Schultz, D.S., Stricker, H., Peabody, J.O., Jeong, W., Chitale, D.A., Bismar, T.A., Rogers, C.G., Menon, M., Gupta, N.S. & Palanisamy, N. (2020) 'Clonal evaluation of early onset prostate cancer by expression profiling of ERG, SPINK1, ETV1, and ETV4 on whole-mount radical prostatectomy tissue', *The Prostate*, 80(1), pp. 38–50.

Lynch, D.C., Revil, T., Schwartzentruber, J., Bhoj, E.J., Innes, A.M., Lamont, R.E., Lemire, E.G., Chodirker, B.N., Taylor, J.P., Zackai, E.H., McLeod, D.R., Kirk, E.P., Hoover-Fong, J., Fleming, L., Savarirayan, R., Majewski, J., Jerome-Majewska, L.A., Parboosingh, J.S. & Bernier, F.P. (2014) 'Disrupted auto-regulation of the spliceosomal gene SNRNPB causes cerebro–costo–mandibular syndrome', *Nature Communications*, 5(1), p. 4483.

Lynch, H.T., Kosoko-Lasaki, O., Leslie, S.W., Rendell, M., Shaw, T., Snyder, C., D'Amico, A.V., Buxbaum, S., Isaacs, W.B., Loeb, S., Moul, J.W. & Powell, I. (2016) 'Screening for familial and hereditary prostate cancer', *International Journal of Cancer*, 138(11), pp. 2579–2591.

Maddison, L.A., Sutherland, B.W., Barrios, R.J. & Greenberg, N.M. (2004) 'Conditional Deletion of Rb Causes Early Stage Prostate Cancer', *Cancer Research*, 64(17), pp. 6018–6025.

Madhani, H.D. & Guthrie, C. (1992) 'A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome', *Cell*, 71(5), pp. 803–817.

Maeder, C., Kutach, A.K. & Guthrie, C. (2009) 'ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C-terminus of Prp8', *Nature structural & molecular biology*, 16(1), pp. 42–48.

Magee, J.A., Abdulkadir, S.A. & Milbrandt, J. (2003) 'Haploinsufficiency at the Nkx3.1 locus. A paradigm for stochastic, dosage-sensitive gene regulation during tumor initiation', *Cancer Cell*, 3(3), pp. 273–283.

Mahas, A., Aman, R. & Mahfouz, M. (2019) 'CRISPR-Cas13d mediates robust RNA virus interference in plants', *Genome Biology*, 20(1), p. 263.

Mahoney, E.M. & Harrison, J.H. (1972) 'Bilateral adrenalectomy for palliative treatment of prostatic cancer', *The Journal of Urology*, 108(6), pp. 936–938.

Maia, T.M., Staes, A., Plasman, K., Pauwels, J., Boucher, K., Argentini, A., Martens, L., Montoye, T., Gevaert, K. & Impens, F. (2020) 'Simple Peptide Quantification Approach for MS-Based Proteomics Quality Control', *ACS Omega*, 5(12), pp. 6754–6762.

Maillet, D., Allioli, N., Péron, J., Plesa, A., Decaussin-Petrucci, M., Tartas, S., Sajous, C., Ruffion, A., Crouzet, S., Freyer, G. & Vlaeminck-Guillem, V. (2021) 'Her2 Expression in Circulating Tumor Cells Is Associated with Poor Outcomes in Patients with Metastatic Castration-Resistant Prostate Cancer', *Cancers*, 13(23), p. 6014.

Mair, A. & Bergmann, D.C. (2022) 'Advances in enzyme-mediated proximity labeling and its potential for plant research', *Plant Physiology*, 188(2), pp. 756–768.

Majumder, P.K. & Kumar, V.L. (1997) 'Androgen receptor mRNA detection in the human foetal prostate', *International Urology and Nephrology*, 29(6), pp. 633–635.

Makkonen, H., Kauhanen, M., Jääskeläinen, T. & Palvimo, J.J. (2011) 'Androgen receptor amplification is reflected in the transcriptional responses of Vertebral-Cancer of the Prostate cells', *Molecular and Cellular Endocrinology*, 331(1), pp. 57–65.

Makowski, K., Vigevani, L., Albericio, F., Valcárcel, J. & Álvarez, M. (2017) 'Sudemycin K: A Synthetic Antitumor Splicing Inhibitor Variant with Improved Activity and Versatile Chemistry', *ACS Chemical Biology*, 12(1), pp. 163–173.

Malca, H., Shomron, N. & Ast, G. (2003) 'The U1 snRNP Base Pairs with the 5' Splice Site within a Penta-snRNP Complex', *Molecular and Cellular Biology*, 23(10), pp. 3442–3455.

Marasco, L.E. & Kornblihtt, A.R. (2023) 'The physiology of alternative splicing', *Nature Reviews Molecular Cell Biology*, 24(4), pp. 242–254.

Martell, J.D., Deerinck, T.J., Sancak, Y., Poulos, T.L., Mootha, V.K., Sosinsky, G.E., Ellisman, M.H. & Ting, A.Y. (2012) 'Engineered ascorbate peroxidase as a genetically-encoded reporter for electron microscopy', *Nature biotechnology*, 30(11), pp. 1143–1148.

Martignano, F., Gurioli, G., Salvi, S., Calistri, D., Costantini, M., Gunelli, R., De Giorgi, U., Foca, F. & Casadio, V. (2016) 'GSTP1 Methylation and Protein Expression in Prostate Cancer: Diagnostic Implications', *Disease Markers*, 2016p. e4358292.

Martin, M. (2011) 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet.journal*, 17(1), pp. 10–12.

Masiello, D., Cheng, S., Bubley, G.J., Lu, M.L. & Balk, S.P. (2002) 'Bicalutamide Functions as an Androgen Receptor Antagonist by Assembly of a Transcriptionally Inactive Receptor \*', *Journal of Biological Chemistry*, 277(29), pp. 26321–26326.

Matias, P.M., Donner, P., Coelho, R., Thomaz, M., Peixoto, C., Macedo, S., Otto, N., Joschko, S., Scholz, P., Wegg, A., Bäsler, S., Schäfer, M., Egner, U. & Carrondo, M.A. (2000) 'Structural Evidence for Ligand Specificity in the Binding Domain of the Human Androgen Receptor: IMPLICATIONS FOR PATHOGENIC GENE MUTATIONS \*', *Journal of Biological Chemistry*, 275(34), pp. 26164–26171.

Maurice-Dror, C., Le Moigne, R., Vaishampayan, U., Montgomery, R.B., Gordon, M.S., Hong, N.H., DiMascio, L., Perabo, F. & Chi, K.N. (2022) 'A phase 1 study to assess the safety, pharmacokinetics, and anti-tumor activity of the androgen receptor n-terminal domain inhibitor epi-506 in patients with metastatic castration-resistant prostate cancer', *Investigational New Drugs*, 40(2), pp. 322–329.

Mavrou, A., Brakspear, K., Hamdollah-Zadeh, M., Damodaran, G., Babaie-Jadidi, R., Oxley, J., Gillatt, D.A., Ladomery, M.R., Harper, S.J., Bates, D.O. & Oltean, S. (2015) 'Serine-arginine protein kinase 1 (SRPK1) inhibition as a potential novel targeted therapeutic strategy in prostate cancer', *Oncogene*, 34(33), pp. 4311–4319.

Mayeda, A., Screamton, G.R., Chandler, S.D., Fu, X.D. & Krainer, A.R. (1999) 'Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements', *Molecular and Cellular Biology*, 19(3), pp. 1853–1863.

McClory, S.P., Lynch, K.W. & Ling, J.P. (2018) 'HnRNP L represses cryptic exons', *RNA (New York, N.Y.)*, 24(6), pp. 761–768.

McEwan, I.J. (2012) 'Intrinsic disorder in the androgen receptor: identification, characterisation and drugability', *Molecular bioSystems*, 8(1), pp. 82–90.

McEwan, I.J. (2004) 'Molecular mechanisms of androgen receptor-mediated gene regulation: structure-function analysis of the AF-1 domain.', *Endocrine-Related Cancer*, 11(2), pp. 281–293.

McEwan, I.J. & Gustafsson, J. (1997) 'Interaction of the human androgen receptor transactivation function with the general transcription factor TFIIF', *Proceedings of the National Academy of Sciences of the United States of America*, 94(16), pp. 8485–8490.

McInerney, E.M., Rose, D.W., Flynn, S.E., Westin, S., Mullen, T.-M., Krones, A., Inostroza, J., Torchia, J., Nolte, R.T., Assa-Munt, N., Milburn, M.V., Glass, C.K. & Rosenfeld, M.G. (1998) 'Determinants of coactivator LXXLL motif specificity in nuclear receptor transcriptional activation', *Genes & Development*, 12(21), pp. 3357–3368.

McKay, R.R., Kwak, L., Crowdus, J.P., Sperger, J.M., Zhao, S.G., Xie, W., Werner, L., Lis, R., Zhang, Z., Wei, X.X., Lang, J.M., Van Allen, E.M., Bhatt, R.S., Yu, E.Y., Nelson, P.S., Bubley, G.J., Montgomery, R.B. & Taplin, M.-E. (2021) 'Phase 2 multicenter study of enzalutamide in metastatic castration resistant prostate cancer to identify mechanisms driving resistance',

*Clinical cancer research : an official journal of the American Association for Cancer Research*, 27(13), pp. 3610–3619.

McNeal, J. (1990) 'Pathology of benign prostatic hyperplasia. Insight into etiology', *The Urologic Clinics of North America*, 17(3), pp. 477–486.

McNeal, J.E. & Bostwick, D.G. (1986) 'Intraductal dysplasia: a premalignant lesion of the prostate', *Human Pathology*, 17(1), pp. 64–71.

Meeks, J.J. & Schaeffer, E.M. (2011) 'Genetic Regulation of Prostate Development', *Journal of Andrology*, 32(3), pp. 210–217.

Mellinghoff, I.K., Vivanco, I., Kwon, A., Tran, C., Wongvipat, J. & Sawyers, C.L. (2004) 'HER2/neu kinase-dependent modulation of androgen receptor function through effects on DNA binding and stability', *Cancer Cell*, 6(5), pp. 517–527.

Melnyk, J.E., Steri, V., Nguyen, H.G., Hann, B., Feng, F.Y. & Shokat, K.M. (2020) 'The splicing modulator sulfonamide indisulam reduces AR-V7 in prostate cancer cells', *Bioorganic & Medicinal Chemistry*, 28(20), p. 115712.

Mende, Y., Jakubik, M., Riessland, M., Schoenen, F., Rossbach, K., Kleinridders, A., Köhler, C., Buch, T. & Wirth, B. (2010) 'Deficiency of the splicing factor Sfrs10 results in early embryonic lethality in mice and has no impact on full-length SMN/Smn splicing', *Human Molecular Genetics*, 19(11), pp. 2154–2167.

Méndez-Mancilla, A., Wessels, H.-H., Legut, M., Kadina, A., Mabuchi, M., Walker, J., Robb, G.B., Holden, K. & Sanjana, N.E. (2022) 'Chemically modified guide RNAs enhance CRISPR-Cas13 knockdown in human cells', *Cell chemical biology*, 29(2), pp. 321-327.e4.

Merkhofer, E.C., Hu, P. & Johnson, T.L. (2014) 'Introduction to Cotranscriptional RNA Splicing', *Methods in molecular biology* (Clifton, N.J.), 1126pp. 83–96.

Metzger, E., Wissmann, M., Yin, N., Müller, J.M., Schneider, R., Peters, A.H.F.M., Günther, T., Buettner, R. & Schüle, R. (2005) 'LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription', *Nature*, 437(7057), pp. 436–439.

Migliaccio, A., Varricchio, L., De Falco, A., Castoria, G., Arra, C., Yamaguchi, H., Ciociola, A., Lombardi, M., Di Stasio, R., Barbieri, A., Baldi, A., Barone, M.V., Appella, E. & Auricchio, F. (2007) 'Inhibition of the SH3 domain-mediated binding of Src to the androgen receptor and its effect on tumor growth', *Oncogene*, 26(46), pp. 6619–6629.

Mili, S. & Steitz, J.A. (2004) 'Evidence for reassociation of RNA-binding proteins after cell lysis: Implications for the interpretation of immunoprecipitation analyses', *RNA*, 10(11), pp. 1692–1694.

Miller, G.J. & Cygan, J.M. (1994) 'Morphology of Prostate Cancer: The Effects of Multifocality on Histological Grade, Tumor Volume and Capsule Penetration', *The Journal of Urology*, 152(5, Part 2), pp. 1709–1713.

Miyake, H., Matsushita, Y., Watanabe, H., Tamura, K., Motoyama, D., Ito, T., Sugiyama, T. & Otsuka, A. (2019) 'Prognostic Significance of Time to Castration Resistance in Patients With Metastatic Castration-sensitive Prostate Cancer', *Anticancer Research*, 39(3), pp. 1391–1396.

Mizui, Y., Sakai, T., Iwata, M., Uenaka, T., Okamoto, K., Shimizu, H., Yamori, T., Yoshimatsu, K. & Asada, M. (2004) 'Pladienolides, new substances from culture of *Streptomyces platensis* Mer-11107. III. In vitro and in vivo antitumor activities', *The Journal of Antibiotics*, 57(3), pp. 188–196.

Mohammed, H., Taylor, C., Brown, G.D., Papachristou, E.K., Carroll, J.S. & D'Santos, C.S. (2016) 'Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes', *Nature Protocols*, 11(2), pp. 316–326.

Montgomery, R.B., Mostaghel, E.A., Vessella, R., Hess, D.L., Kalhorn, T.F., Higano, C.S., True, L.D. & Nelson, P.S. (2008) 'Maintenance of Intratumoral Androgens in Metastatic Prostate Cancer: A Mechanism for Castration-Resistant Tumor Growth', *Cancer research*, 68(11), pp. 4447–4454.

Monti, S., Di Silverio, F., Toscano, V., Martini, C., Lanzara, S., Varasano, P.A. & Sciarra, F. (1998) 'Androgen Concentrations and Their Receptors in the Periurethral Region Are Higher Than Those of the Subcapsular Zone in Benign Prostatic Hyperplasia (BPH)', *Journal of Andrology*, 19(4), pp. 428–433.

Mooradian, A.D., Morley, J.E. & Korenman, S.G. (1987) 'Biological Actions of Androgens', *Endocrine Reviews*, 8(1), pp. 1–28.

Moreira, D.M., Howard, L.E., Sourbeer, K.N., Amarasekara, H.S., Chow, L.C., Cockrell, D.C., Pratson, C.L., Hanyok, B.T., Aronson, W.J., Kane, C.J., Terris, M.K., Amling, C.L., Cooperberg, M.R. & Freedland, S.J. (2017) 'Predicting Time From Metastasis to Overall Survival in Castration-Resistant Prostate Cancer: Results From SEARCH', *Clinical genitourinary cancer*, 15(1), pp. 60–66.e2.

Morote, J., Aguilar, A., Planas, J. & Trilla, E. (2022) 'Definition of Castrate Resistant Prostate Cancer: New Insights', *Biomedicines*, 10(3), p. 689.

Moses, M.A., Kim, Y.S., Rivera-Marquez, G.M., Oshima, N., Watson, M.J., Beebe, K.E., Wells, C., Lee, S., Zuehlke, A.D., Shao, H., Bingman, W.E., Kumar, V., Malhotra, S.V., Weigel, N.L., Gestwicki, J.E., Trepel, J.B. & Neckers, L.M. (2018) Targeting the HSP40/hsp70 chaperone axis as a novel strategy to treat castration-resistant prostate cancer. *Cancer Research*. 78 (14), 4022–4035.

Mostaghel, E.A., Marck, B.T., Plymate, S., Vessella, R., Balk, S., Matsumoto, A.M., Nelson1, P.S. & Montgomery, R.B. (2011) 'Resistance to CYP17A1 inhibition with abiraterone in castration resistant prostate cancer: Induction of steroidogenesis and androgen receptor splice variants', *Clinical cancer research : an official journal of the American Association for Cancer Research*, 17(18), pp. 5913–5925.

Mostaghel, E.A., Zhang, A., Hernandez, S., Mark, B.T., Zhang, X., Tamae, D., Biehl, H.E., Tretiakova, M., Bartlett, J., Burns, J., Dumpit, R., Ang, L., Matsumoto, A.M., Penning, T.M.,

Balk, S.P., Morrissey, C., Corey, E., True, L.D. & Nelson, P.S. (2019) 'Contribution of Adrenal Glands to Intra-tumor Androgens and Growth of Castration Resistant Prostate Cancer.', *Clinical cancer research : an official journal of the American Association for Cancer Research*, 25(1), pp. 426–439.

Murata, K., Mimura, A., Suzuki, H., Mikami, N., Hamada, Y., Kato, K., Iki, N., Ishida, M., Daitoku, Y., Tanimoto, Y., Okiyoneda, T., Fujiyama, T., Dinh, T.T.H., Mizuno, S. & Sugiyama, F. (2021) 'Efficient induction of proximity-dependent labelling by biotin feeding in BMAL1-BioloD knock-in mice', *The Journal of Biochemistry*, 170(4), pp. 453–461.

Myers, S.A., Wright, J., Peckner, R., Kalish, B.T., Zhang, F. & Carr, S.A. (2018) 'Discovery of proteins associated with a predefined genomic locus in living cells via dCAS9-APEX-mediated proximity labeling', *Nature methods*, 15(6), pp. 437–439.

Nadal, M., Prekovic, S., Gallastegui, N., Helsen, C., Abella, M., Zielinska, K., Gay, M., Vilaseca, M., Taulès, M., Houtsmuller, A.B., van Royen, M.E., Claessens, F., Fuentes-Prior, P. & Estébanez-Perpiñá, E. (2017) 'Structure of the homodimeric androgen receptor ligand-binding domain', *Nature Communications*, 8(1), p. 14388.

Nadiminty, N., Tummala, R., Liu, C., Lou, W., Evans, C.P. & Gao, A.C. (2015) 'NF-kappaB2/p52:c-Myc:hnRNPA1 pathway regulates expression of androgen receptor splice variants and enzalutamide sensitivity in prostate cancer', *Molecular cancer therapeutics*, 14(8), pp. 1884–1895.

Naftelberg, S., Schor, I.E., Ast, G. & Kornblihtt, A.R. (2015) 'Regulation of alternative splicing through coupling with transcription and chromatin structure', *Annual Review of Biochemistry*, 84pp. 165–198.

Nagandla, H., Robertson, M.J., Putluri, V., Putluri, N., Coarfa, C. & Weigel, N.L. (2020) 'Isoform-specific Activities of Androgen Receptor and its Splice Variants in Prostate Cancer Cells', *Endocrinology*, 162(3), p. bcaa227.

Nakata, D., Nakao, S., Nakayama, K., Araki, S., Nakayama, Y., Aparicio, S., Hara, T. & Nakanishi, A. (2017) 'The RNA helicase DDX39B and its paralog DDX39A regulate androgen receptor splice variant AR-V7 generation', *Biochemical and Biophysical Research Communications*, 483(1), pp. 271–276.

Nakayama, M., Gonzalgo, M.L., Yegnasubramanian, S., Lin, X., De Marzo, A.M. & Nelson, W.G. (2004) 'GSTP1 CpG island hypermethylation as a molecular biomarker for prostate cancer', *Journal of Cellular Biochemistry*, 91(3), pp. 540–552.

Nakka, M., Agoulnik, I.U. & Weigel, N.L. (2013) 'Targeted disruption of the p160 coactivator interface of androgen receptor (AR) selectively inhibits AR activity in both androgen-dependent and castration-resistant AR-expressing prostate cancer cells', *The international journal of biochemistry & cell biology*, 45(4), pp. 763–772.

Narayan, P., Trachtenberg, J., Lepor, H., Debruyne, F.M., Tewari, A., Stone, N., Das, S., Jimenez-Cruz, J.F., Shearer, R., Klimberg, I., Schellhammer, P.F. & Costello, A.J. (1996) 'A dose-

response study of the effect of flutamide on benign prostatic hyperplasia: results of a multicenter study', *Urology*, 47(4), pp. 497–504.

Nassar, G.N. & Leslie, S.W. (2022) 'Physiology, Testosterone', in *StatPearls*. [Online]. Treasure Island (FL): StatPearls Publishing.

Neeb, A., Figueiredo, I., Bogdan, D., Cato, L., Stober, J., Jimenez-Vacas, J.M., Gourain, V., Lee, I.I., Seeger, R., Muhle-Goll, C., Gurel, B., Welti, J., Rodrigues, D.N., Rekowski, J., Qiu, X., Jiang, Y., Di Micco, P., Mateos, B., Bielskuté, S., et al. (2022) 'Targeting the Bag-1 family of co-chaperones in Lethal prostate cancer', *bioRxiv*.

Neil, E.E. & Bisaccia, E.K. (2019) 'Nusinersen: A Novel Antisense Oligonucleotide for the Treatment of Spinal Muscular Atrophy', *The Journal of Pediatric Pharmacology and Therapeutics : JPPT*, 24(3), pp. 194–203.

Nesvizhskii, A.I. (2010) 'A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics', *Journal of proteomics*, 73(11), pp. 2092–2123.

Newman, A.J. (1997) 'The role of U5 snRNP in pre-mRNA splicing.', *The EMBO Journal*, 16(19), pp. 5797–5800.

Ng, M. & Baradhi, K.M. (2022) 'Benign Prostatic Hyperplasia', in *StatPearls*. [Online]. Treasure Island (FL): StatPearls Publishing.

Nguyen, T.-V.V., Yao, M. & Pike, C.J. (2007) 'Flutamide and cyproterone acetate exert agonist effects: induction of androgen receptor-dependent neuroprotection', *Endocrinology*, 148(6), pp. 2936–2943.

Ni, L., Llewellyn, R., Kesler, C.T., Kelley, J.B., Spencer, A., Snow, C.J., Shank, L. & Paschal, B.M. (2013) 'Androgen Induces a Switch from Cytoplasmic Retention to Nuclear Import of the Androgen Receptor', *Molecular and Cellular Biology*, 33(24), pp. 4766–4778.

NICE (2016) *1 Recommendations / Abiraterone for treating metastatic hormone-relapsed prostate cancer before chemotherapy is indicated / Guidance / NICE*. [Online] [online]. Available from: <https://www.nice.org.uk/guidance/ta387/chapter/1-Recommendations> (Accessed 3 October 2022).

NICE (2022) *Cyproterone acetate / Drugs / BNF content published by NICE*. [Online] [online]. Available from: <https://bnf.nice.org.uk/drugs/cyproterone-acetate/> (Accessed 15 August 2022).

NICE (2006) *Overview / Docetaxel for the treatment of hormone-refractory metastatic prostate cancer / Guidance / NICE*. [Online] [online]. Available from: <https://www.nice.org.uk/guidance/ta101> (Accessed 14 September 2022).

NICE (2021a) *Overview / Enzalutamide for treating hormone-sensitive metastatic prostate cancer / Guidance / NICE*. [Online] [online]. Available from: <https://www.nice.org.uk/guidance/ta712> (Accessed 16 August 2022).

NICE (2021b) *Recommendations / Prostate cancer: diagnosis and management / Guidance / NICE*. [Online] [online]. Available from: <https://www.nice.org.uk/guidance/ng131/chapter/recommendations#people-having-hormone-therapy> (Accessed 12 August 2022).

Nicholson, T.M., Sehgal, P.D., Drew, S.A., Huang, W. & Ricke, W.A. (2013) 'Sex steroid receptor expression and localization in benign prostatic hyperplasia varies with tissue compartment', *Differentiation*, 85(4), pp. 140–149.

Nickols, N.G., Nazarian, R., Zhao, S.G., Tan, V., Uzunangelov, V., Xia, Z., Baertsch, R., Neeman, E., Gao, A.C., Thomas, G.V., Howard, L., De Hoedt, A.M., Stuart, J., Goldstein, T., Chi, K., Gleave, M.E., Graff, J.N., Beer, T.M., Drake, J.M., et al. (2019) 'MEK-ERK signaling is a therapeutic target in metastatic castration resistant prostate cancer', *Prostate Cancer and Prostatic Diseases*, 22(4), pp. 531–538.

Nikolaou, N., Gordon, P.M., Hamid, F., Taylor, R., Lloyd-Jones, J., Makeyev, E.V. & Houart, C. (2022) 'Cytoplasmic pool of U1 spliceosome protein SNRNP70 shapes the axonal transcriptome and regulates motor connectivity', *Current biology: CB*, 32(23), pp. 5099–5115.e8.

Novara, G., Galfano, A., Secco, S., Ficarra, V. & Artibani, W. (2009) 'Impact of surgical and medical castration on serum testosterone level in prostate cancer patients', *Urologia Internationalis*, 82(3), pp. 249–255.

Novoyatleva, T., Heinrich, B., Tang, Y., Benderska, N., Butchbach, M.E.R., Lorson, C.L., Lorson, M.A., Ben-Dov, C., Fehlbaum, P., Bracco, L., Burghes, A.H.M., Bollen, M. & Stamm, S. (2008) 'Protein phosphatase 1 binds to the RNA recognition motif of several splicing factors and regulates alternative pre-mRNA processing', *Human Molecular Genetics*, 17(1), pp. 52–70.

O'Brien, K., Matlin, A.J., Lowell, A.M. & Moore, M.J. (2008) 'The Biflavonoid Isoginkgetin Is a General Inhibitor of Pre-mRNA Splicing', *The Journal of Biological Chemistry*, 283(48), pp. 33147–33154.

Oefelein, M.G., Feng, A., Scolieri, M.J., Ricchiuti, D. & Resnick, M.I. (2000) 'Reassessment of the definition of castrate levels of testosterone: implications for clinical decision making', *Urology*, 56(6), pp. 1021–1024.

Oh, J.J., West, A.R., Fishbein, M.C. & Slamon, D.J. (2002) 'A candidate tumor suppressor gene, H37, from the human lung cancer tumor suppressor locus 3p21.3', *Cancer Research*, 62(11), pp. 3207–3213.

Oh-Hohenhorst, S.J., Tilki, D., Ahlers, A.-K., Suling, A., Hahn, O., Tennstedt, P., Matuszcak, C., Maar, H., Labitzky, V., Hanika, S., Starzonek, S., Baumgart, S., Johnsen, S.A., Kluth, M., Sirma, H., Simon, R., Sauter, G., Huland, H., Schumacher, U., et al. (2022) 'CHD1 loss negatively influences metastasis-free survival in R0-resected prostate cancer patients and promotes spontaneous metastasis in vivo', *Cancer Gene Therapy*, 29(1), pp. 49–61.

Okamoto, T., Hatakeyama, S., Takahashi, M., Narita, S., Ishida, M., Hoshi, S., Kawamura, S., Ishidoya, S., Shimoda, J., Kawaguchi, T., Tochigi, T., Mitsuzuka, K., Tsuchiya, N., Arai, Y.,

Habuchi, T. & Ohyama, C. (2020) 'The impact of time-to-castration resistance on survival in patients with metastatic hormone-naïve prostate cancer: A multicenter retrospective study.', *Journal of Clinical Oncology*, 38(6\_suppl), pp. 213–213.

Olender, J. & Lee, N.H. (2019) 'Role of alternative splicing in prostate cancer aggressiveness and drug resistance in African Americans.', *Advances in experimental medicine and biology*, 1164pp. 119–139.

Olivieri, J.E., Dehghannasiri, R., Wang, P.L., Jang, S., de Morree, A., Tan, S.Y., Ming, J., Ruohao Wu, A., Tabula Sapiens Consortium, Quake, S.R., Krasnow, M.A. & Salzman, J. (2021) 'RNA splicing programs define tissue compartments and cell types at single-cell resolution' Gene W Yeo & Patricia J Wittkopp (eds.), *eLife*, 10p. e70692.

O'Malley, K.J., Dhir, R., Nelson, J.B., Bost, J., Lin, Y. & Wang, Z. (2009) 'The expression of androgen-responsive genes is up-regulated in the epithelia of benign prostatic hyperplasia', *The Prostate*, 69(16), pp. 1716–1723.

Osguthorpe, D.J. & Hagler, A.T. (2011) 'Mechanism of Androgen Receptor Antagonism by Bicalutamide in the Treatment of Prostate Cancer', *Biochemistry*, 50(19), pp. 4105–4113.

Osman, I., Drobnjak, M., Fazzari, M., Ferrara, J., Scher, H.I. & Cordon-Cardo, C. (1999) 'Inactivation of the p53 Pathway in Prostate Cancer: Impact on Tumor Progression1', *Clinical Cancer Research*, 5(8), pp. 2082–2088.

Ouyang, J., Qin, G., Liu, Z., Jian, X., Shi, T. & Xie, L. (2022) 'ToPP: Tumor online prognostic analysis platform for prognostic feature selection and clinical patient subgroup selection', *iScience*, 25(5), p. 104190.

Ozanne, D.M., Brady, M.E., Cook, S., Gaughan, L., Neal, D.E. & Robson, C.N. (2000) 'Androgen Receptor Nuclear Translocation Is Facilitated by the f-Actin Cross-Linking Protein Filamin', *Molecular Endocrinology*, 14(10), pp. 1618–1626.

Padgett, R.A. (2012) 'New connections between splicing and human disease', *Trends in Genetics*, 28(4), pp. 147–154.

Padrón, A., Iwasaki, S. & Ingolia, N.T. (2019) 'Proximity RNA Labeling by APEX-Seq Reveals the Organization of Translation Initiation Complexes and Repressive RNA Granules', *Molecular Cell*, 75(4), pp. 875-887.e5.

Page, S.T., Lin, D.W., Mostaghel, E.A., Hess, D.L., True, L.D., Amory, J.K., Nelson, P.S., Matsumoto, A.M. & Bremner, W.J. (2006) 'Persistent Intraprostatic Androgen Concentrations after Medical Castration in Healthy Men', *The Journal of Clinical Endocrinology & Metabolism*, 91(10), pp. 3850–3856.

Palaz, F., Kalkan, A.K., Can, Ö., Demir, A.N., Tozluyurt, A., Özcan, A. & Ozsoz, M. (2021) 'CRISPR-Cas13 System as a Promising and Versatile Tool for Cancer Diagnosis, Therapy, and Research', *ACS Synthetic Biology*, 10(6), pp. 1245–1267.

Palmer, E. & Freeman, T. (2004) 'Investigation into the use of C- and N-terminal GFP fusion proteins for subcellular localization studies using reverse transfection microarrays', *Comparative and Functional Genomics*, 5(4), pp. 342–353.

Palvimo, J.J., Reinikainen, P., Ikonen, T., Kallio, P.J., Moilanen, A. & Jänne, O.A. (1996) 'Mutual Transcriptional Interference between RelA and Androgen Receptor \*', *Journal of Biological Chemistry*, 271(39), pp. 24151–24156.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J. & Blencowe, B.J. (2008) 'Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing', *Nature Genetics*, 40(12), pp. 1413–1415.

Pandit, S., Zhou, Y., Shiue, L., Coutinho-Mansfield, G., Li, H., Qiu, J., Huang, J., Yeo, G.W., Ares, M. & Fu, X.-D. (2013) 'Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing', *Molecular Cell*, 50(2), pp. 223–235.

Papadopoulos, J.S. & Agarwala, R. (2007) 'COBALT: constraint-based alignment tool for multiple protein sequences', *Bioinformatics (Oxford, England)*, 23(9), pp. 1073–1079.

Papageorgiou, D.N., Demmers, J. & Strouboulis, J. (2013) 'NP-40 reduces contamination by endogenous biotinylated carboxylases during purification of biotin tagged nuclear proteins', *Protein expression and purification*, 89(1), pp. 80–83.

Pappireddi, N., Martin, L. & Wühr, M. (2019) 'A Review on Quantitative Multiplexed Proteomics', *ChemBioChem*, 20(10), pp. 1210–1224.

Parekh, S., Ziegenhain, C., Vieth, B., Enard, W. & Hellmann, I. (2016) 'The impact of amplification on differential expression analyses by RNA-seq', *Scientific Reports*, 6(1), p. 25533.

Parimi, V., Goyal, R., Poropatich, K. & Yang, X.J. (2014) 'Neuroendocrine differentiation of prostate cancer: a review', *American Journal of Clinical and Experimental Urology*, 2(4), pp. 273–285.

Park, K., Dalton, J.T., Narayanan, R., Barbieri, C.E., Hancock, M.L., Bostwick, D.G., Steiner, M.S. & Rubin, M.A. (2014) 'TMPRSS2:ERG Gene Fusion Predicts Subsequent Detection of Prostate Cancer in Patients With High-Grade Prostatic Intraepithelial Neoplasia', *Journal of Clinical Oncology*, 32(3), pp. 206–211.

Parmley, J.L., Chamary, J.V. & Hurst, L.D. (2006) 'Evidence for Purifying Selection Against Synonymous Mutations in Mammalian Exonic Splicing Enhancers', *Molecular Biology and Evolution*, 23(2), pp. 301–309.

Paronetto, M.P., Cappellari, M., Busà, R., Pedrotti, S., Vitali, R., Comstock, C., Hyslop, T., Knudsen, K.E. & Sette, C. (2010) 'Alternative splicing of the cyclin D1 proto-oncogene is regulated by the RNA-binding protein Sam68', *Cancer Research*, 70(1), pp. 229–239.

Paschalis, A., Welti, J., Neeb, A.J., Yuan, W., Figueiredo, I., Pereira, R., Ferreira, A., Riisnaes, R., Nava Rodrigues, D., Jiménez-Vacas, J.M., Kim, S., Uo, T., Di Micco, P., Tumber, A., Islam, S., Moesser, M.A., Abboud, M., Kawamura, A., Gurel, B., et al. (2021) 'JMJD6 is a druggable

oxygenase that regulates AR-V7 expression in prostate cancer', *Cancer research*, 81(4), pp. 1087–1100.

Patton, R.D., Sanjeev, M., Woodward, L.A., Mabin, J.W., Bundschuh, R. & Singh, G. (2020) 'Chemical crosslinking enhances RNA immunoprecipitation for efficient identification of binding sites of proteins that photo-crosslink poorly with RNA', *RNA*, 26(9), pp. 1216–1233.

Paul, R. & Breul, J. (2000) 'Antiandrogen withdrawal syndrome associated with prostate cancer therapies: incidence and clinical significance', *Drug Safety*, 23(5), pp. 381–390.

Pedram, A., Razandi, M., Sainson, R.C.A., Kim, J.K., Hughes, C.C. & Levin, E.R. (2007) 'A Conserved Mechanism for Steroid Receptor Translocation to the Plasma Membrane \*', *Journal of Biological Chemistry*, 282(31), pp. 22278–22288.

Peer, A., Gottfried, M., Sinibaldi, V., Carducci, M.A., Eisenberger, M.A., Sella, A., Leibowitz-Amit, R., Berger, R. & Keizman, D. (2014) 'Comparison of Abiraterone Acetate Versus Ketoconazolein Patients with Metastatic Castration Resistant Prostate Cancer Refractory to Docetaxel', *The Prostate*, 74(4), pp. 433–440.

Peets, E.A., Faye, HENSON, M. & Neri, R. (1974) 'On the Mechanism of the Anti-androgenic Action of Flutamide ( $\alpha$ - $\alpha$ - $\alpha$ -Trifluoro-2-methyl-4'-nitro-m-propionotoluidide) in the Rat', *Endocrinology*, 94(2), pp. 532–540.

Penson, D.F., Armstrong, A.J., Concepcion, R., Agarwal, N., Olsson, C., Karsh, L., Dunshee, C., Wang, F., Wu, K., Krivoshik, A., Phung, D. & Higano, C.S. (2016) 'Enzalutamide Versus Bicalutamide in Castration-Resistant Prostate Cancer: The STRIVE Trial', *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 34(18), pp. 2098–2106.

Penson, D.F., Armstrong, A.J., Concepcion, R.S., Agarwal, N., Olsson, C.A., Karsh, L.I., Dunshee, C.J., Duggan, W., Shen, Q., Sugg, J., Haas, G.P. & Higano, C.S. (2022) 'Enzalutamide versus bicalutamide in patients with nonmetastatic castration-resistant prostate cancer: a prespecified subgroup analysis of the STRIVE trial', *Prostate Cancer and Prostatic Diseases*, 25(2), pp. 363–365.

Perrakis, A. & Sixma, T.K. (2021) 'AI revolutions in biology', *EMBO reports*, 22(11), p. e54046.

Peterson, C.W.H., Deol, K.K., To, M. & Olzmann, J.A. (2021) 'Optimized protocol for the identification of lipid droplet proteomes using proximity labeling proteomics in cultured human cells', *STAR Protocols*, 2(2), p. 100579.

Peterziel, H., Mink, S., Schonert, A., Becker, M., Klocker, H. & Cato, A.C. (1999) 'Rapid signalling by androgen receptor in prostate cancer cells', *Oncogene*, 18(46), pp. 6322–6329.

Petrylak, D.P., Tangen, C.M., Hussain, M.H.A., Lara, P.N., Jones, J.A., Taplin, M.E., Burch, P.A., Berry, D., Moinpour, C., Kohli, M., Benson, M.C., Small, E.J., Raghavan, D. & Crawford, E.D. (2004) 'Docetaxel and Estramustine Compared with Mitoxantrone and Prednisone for Advanced Refractory Prostate Cancer', *New England Journal of Medicine*, 351(15), pp. 1513–1520.

Pezaro, C.J., Omlin, A., Lorente, D., Nava Rodrigues, D., Ferraldeschi, R., Bianchini, D., Mukherji, D., Riisnaes, R., Altavilla, A., Crespo, M., Tunariu, N., de Bono, J.S. & Attard, G. (2014) 'Visceral Disease in Castration-resistant Prostate Cancer', *European Urology*, 65(2), pp. 270–273.

Ponguta, L.A., Gregory, C.W., French, F.S. & Wilson, E.M. (2008) 'Site-specific Androgen Receptor Serine Phosphorylation Linked to Epidermal Growth Factor-dependent Growth of Castration-recurrent Prostate Cancer', *The Journal of Biological Chemistry*, 283(30), pp. 20989–21001.

Poppel, H. van & Nilsson, S. (2008) 'Testosterone Surge: Rationale for Gonadotropin-Releasing Hormone Blockers?', *Urology*, 71(6), pp. 1001–1006.

Potter, G.A., Barrie, S.E., Jarman, M. & Rowlands, M.G. (1995) 'Novel steroidal inhibitors of human cytochrome P45017 alpha (17 alpha-hydroxylase-C17,20-lyase): potential agents for the treatment of prostatic cancer', *Journal of Medicinal Chemistry*, 38(13), pp. 2463–2471.

Poyet, P. & Labrie, F. (1985) 'Comparison of the antiandrogenic/androgenic activities of flutamide, cyproterone acetate and megestrol acetate', *Molecular and Cellular Endocrinology*, 42(3), pp. 283–288.

Pratt, W.B. & Toft, D.O. (2003) 'Regulation of Signaling Protein Function and Trafficking by the hsp90/hsp70-Based Chaperone Machinery', *Experimental Biology and Medicine*, 228(2), pp. 111–133.

Pratt, W.B. & Toft, D.O. (1997) 'Steroid Receptor Interactions with Heat Shock Protein and Immunophilin Chaperones\*', *Endocrine Reviews*, 18(3), pp. 306–360.

Prekovic, S., van Royen, M.E., Voet, A.R.D., Geverts, B., Houtman, R., Melchers, D., Zhang, K.Y.J., Van den Broeck, T., Smeets, E., Spans, L., Houtsmuller, A.B., Joniau, S., Claessens, F. & Helsen, C. (2016) 'The Effect of F877L and T878A Mutations on Androgen Receptor Response to Enzalutamide', *Molecular Cancer Therapeutics*, 15(7), pp. 1702–1712.

Prostate Cancer Trialists Collaborative Group (2000) 'Maximum androgen blockade in advanced prostate cancer: an overview of the randomised trials', *The Lancet*, 355(9214), pp. 1491–1498.

Qin, D., Huang, L., Wlodaver, A., Andrade, J. & Staley, J.P. (2016) 'Sequencing of lariat termini in *S. cerevisiae* reveals 5' splice sites, branch points, and novel splicing events', *RNA*, 22(2), pp. 237–253.

Qin, W., Cho, K.F., Cavanagh, P.E. & Ting, A.Y. (2021) 'Deciphering molecular interactions by proximity labeling', *Nature Methods*, 18(2), pp. 133–143.

Qiu, J., Qu, R., Lin, M., Xu, J., Zhu, Q., Zhang, Z. & Sun, J. (2022) 'Position-dependent effects of hnRNP A1/A2 in SMN1/2 exon7 splicing', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1865(8), p. 194875.

Qiu, W., Xu, Z., Zhang, M., Zhang, D., Fan, H., Li, T., Wang, Q., Liu, P., Zhu, Z., Du, D., Tan, M., Wen, B. & Liu, Y. (2019) 'Determination of local chromatin interactions using a combined CRISPR and peroxidase APEX2 system', *Nucleic Acids Research*, 47(9), p. e52.

Qiu, X., Boufaied, N., Hallal, T., Feit, A., de Polo, A., Luoma, A.M., Alahmadi, W., Larocque, J., Zadra, G., Xie, Y., Gu, S., Tang, Q., Zhang, Y., Syamala, S., Seo, J.-H., Bell, C., O'Connor, E., Liu, Y., Schaeffer, E.M., et al. (2022) 'MYC drives aggressive prostate cancer by disrupting transcriptional pause release at androgen receptor targets', *Nature Communications*, 13(1), p. 2559.

R Core Team (2021) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Radaeva, M., Ban, F., Zhang, F., LeBlanc, E., Lallous, N., Rennie, P.S., Gleave, M.E. & Cherkasov, A. (2021) 'Development of Novel Inhibitors Targeting the D-Box of the DNA Binding Domain of Androgen Receptor', *International Journal of Molecular Sciences*, 22(5), p. 2493.

Ramos, M., Geistlinger, L., Oh, S., Schiffer, L., Azhar, R., Kodali, H., de Bruijn, I., Gao, J., Carey, V.J., Morgan, M. & Waldron, L. (2020) 'Multiomic Integration of Public Oncology Databases in Bioconductor', *JCO Clinical Cancer Informatics*, (4), pp. 958–971.

Ramsey, S., Veenstra, D., Clarke, L., Gandhi, S., Hirsch, M. & Penson, D. (2005) 'Is combined androgen blockade with bicalutamide cost-effective compared with combined androgen blockade with flutamide?', *Urology*, 66(4), pp. 835–839.

Rana, K., Davey, R.A. & Zajac, J.D. (2014) 'Human androgen deficiency: insights gained from androgen receptor knockout mouse models', *Asian Journal of Andrology*, 16(2), pp. 169–177.

Randall, V.A. (1994) 'Role of 5 alpha-reductase in health and disease', *Bailliere's Clinical Endocrinology and Metabolism*, 8(2), pp. 405–431.

Rebello, R.J., Oing, C., Knudsen, K.E., Loeb, S., Johnson, D.C., Reiter, R.E., Gillessen, S., Van der Kwast, T. & Bristow, R.G. (2021) 'Prostate cancer', *Nature Reviews Disease Primers*, 7(1), pp. 1–27.

Reid, J., Kelly, S.M., Watt, K., Price, N.C. & McEwan, I.J. (2002) 'Conformational analysis of the androgen receptor amino-terminal domain involved in transactivation. Influence of structure-stabilizing solutes and protein-protein interactions', *The Journal of Biological Chemistry*, 277(22), pp. 20079–20086.

Reimers, M.A., Yip, S.M., Zhang, L., Cieslik, M., Dhawan, M., Montgomery, B., Wyatt, A.W., Chi, K.N., Small, E.J., Chinnaiyan, A.M., Alva, A.S., Feng, F.Y. & Chou, J. (2020) 'Clinical Outcomes in Cyclin-dependent Kinase 12 Mutant Advanced Prostate Cancer', *European urology*, 77(3), pp. 333–341.

Rennel, E., Waine, E., Guan, H., Schüler, Y., Leenders, W., Woolard, J., Sugiono, M., Gillatt, D., Kleinerman, E., Bates, D. & Harper, S. (2008) 'The endogenous anti-angiogenic VEGF isoform, VEGF165b inhibits human tumour growth in mice', *British Journal of Cancer*, 98(7), pp. 1250–1257.

Rhee, H.-W., Zou, P., Udeshi, N.D., Martell, J.D., Mootha, V.K., Carr, S.A. & Ting, A.Y. (2013) 'Proteomic Mapping of Mitochondria in Living Cells via Spatially-Restricted Enzymatic Tagging', *Science (New York, N.Y.)*, 339(6125), p. 1328.

Riegman, P.H., Vlietstra, R.J., van der Korput, J.A., Brinkmann, A.O. & Trapman, J. (1991) 'The promoter of the prostate-specific antigen gene contains a functional androgen responsive element', *Molecular Endocrinology (Baltimore, Md.)*, 5(12), pp. 1921–1930.

Ris-Stalpers, C., Verleun-Mooijman, M.C., Trapman, J. & Brinkmann, A.O. (1993) 'Threonine on amino acid position 868 in the human androgen receptor is essential for androgen binding specificity and functional activity', *Biochemical and Biophysical Research Communications*, 196(1), pp. 173–180.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. & Smyth, G.K. (2015) 'limma powers differential expression analyses for RNA-sequencing and microarray studies', *Nucleic Acids Research*, 43(7), p. e47.

Roberts, M.J., Teloken, P., Chambers, S.K., Williams, S.G., Yaxley, J., Samaratunga, H., Frydenberg, M. & Gardiner, R.A. ('Frank') (2000) 'Prostate Cancer Detection', in Kenneth R. Feingold, Bradley Anawalt, Alison Boyce, George Chrousos, Wouter W. de Herder, Ketan Dhatariya, Kathleen Dungan, Jerome M. Hershman, Johannes Hofland, Sanjay Kalra, Gregory Kaltsas, Christian Koch, Peter Kopp, Márta Korbonits, Christopher S. Kovacs, Wendy Kuohung, Blandine Laferrère, Miles Levy, Elizabeth A. McGee, et al. (eds.) *Endotext*. [Online]. South Dartmouth (MA): MDText.com, Inc.

Roberts, T.C., Langer, R. & Wood, M.J.A. (2020) 'Advances in oligonucleotide drug delivery', *Nature Reviews Drug Discovery*, 19(10), pp. 673–694.

Robinson, M.D., McCarthy, D.J. & Smyth, G.K. (2010) 'edgeR: a Bioconductor package for differential expression analysis of digital gene expression data', *Bioinformatics (Oxford, England)*, 26(1), pp. 139–140.

Roehrborn, C.G. (2005) 'Benign Prostatic Hyperplasia: An Overview', *Reviews in Urology*, 7(Suppl 9), pp. S3–S14.

Rooke, N., Markovtsov, V., Cagavi, E. & Black, D.L. (2003) 'Roles for SR Proteins and hnRNP A1 in the Regulation of c-src Exon N1', *Molecular and Cellular Biology*, 23(6), pp. 1874–1884.

Roscigno, R.F. & Garcia-Blanco, M.A. (1995) 'SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome.', *RNA*, 1(7), pp. 692–706.

Rothrock, C.R., House, A.E. & Lynch, K.W. (2005) 'HnRNP L represses exon splicing via a regulated exonic splicing silencer', *The EMBO journal*, 24(15), pp. 2792–2802.

Roux, K.J., Kim, D.I., Burke, B. & May, D.G. (2018) 'BioID: A Screen for Protein-Protein Interactions', *Current protocols in protein science*, 91p. 19.23.1-19.23.15.

Roux, K.J., Kim, D.I., Raida, M. & Burke, B. (2012) 'A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells', *The Journal of Cell Biology*, 196(6), pp. 801–810.

Roy, A., Kucukural, A. & Zhang, Y. (2010) 'I-TASSER: a unified platform for automated protein structure and function prediction', *Nature Protocols*, 5(4), pp. 725–738.

van Royen, M.E., van Cappellen, W.A., de Vos, C., Houtsmuller, A.B. & Trapman, J. (2012) 'Stepwise androgen receptor dimerization', *Journal of Cell Science*, 125(8), pp. 1970–1979.

Rozanova, S., Barkovits, K., Nikolov, M., Schmidt, C., Urlaub, H. & Marcus, K. (2021) 'Quantitative Mass Spectrometry-Based Proteomics: An Overview', in Katrin Marcus, Martin Eisenacher, & Barbara Sitek (eds.) *Quantitative Methods in Proteomics*. Methods in Molecular Biology. [Online]. New York, NY: Springer US. pp. 85–116.

RStudio Team (2022) *RStudio: Integrated Development for R*.

Ruan, W., Powell-Braxton, L., Kopchick, J.J. & Kleinberg, D.L. (1999) 'Evidence that insulin-like growth factor I and growth hormone are required for prostate gland development', *Endocrinology*, 140(5), pp. 1984–1989.

Ryan, C.J., Smith, M.R., Fizazi, K., Saad, F., Mulders, P.F.A., Sternberg, C.N., Miller, K., Logothetis, C.J., Shore, N.D., Small, E.J., Carles, J., Flraig, T.W., Taplin, M.-E., Higano, C.S., de Souza, P., de Bono, J.S., Griffin, T.W., De Porre, P., Yu, M.K., et al. (2015) 'Abiraterone acetate plus prednisone versus placebo plus prednisone in chemotherapy-naïve men with metastatic castration-resistant prostate cancer (COU-AA-302): final overall survival analysis of a randomised, double-blind, placebo-controlled phase 3 study', *The Lancet Oncology*, 16(2), pp. 152–160.

Ryan, C.J., Smith, M.R., Fong, L., Rosenberg, J.E., Kantoff, P., Raynaud, F., Martins, V., Lee, G., Kheoh, T., Kim, J., Molina, A. & Small, E.J. (2010) 'Phase I Clinical Trial of the CYP17 Inhibitor Abiraterone Acetate Demonstrating Clinical Activity in Patients With Castration-Resistant Prostate Cancer Who Received Prior Ketoconazole Therapy', *Journal of Clinical Oncology*, 28(9), pp. 1481–1488.

Sack, J.S., Kish, K.F., Wang, C., Attar, R.M., Kiefer, S.E., An, Y., Wu, G.Y., Scheffler, J.E., Salvati, M.E., Krystek, S.R., Weinmann, R. & Einspahr, H.M. (2001) 'Crystallographic structures of the ligand-binding domains of the androgen receptor and its T877A mutant complexed with the natural agonist dihydrotestosterone', *Proceedings of the National Academy of Sciences*, 98(9), pp. 4904–4909.

Sahin, I., Mega, A.E. & Carneiro, B.A. (2018) 'Androgen receptor-independent prostate cancer: an emerging clinical entity', *Cancer Biology & Therapy*, 19(5), pp. 347–348.

Sahu, B., Laakso, M., Ovaska, K., Mirtti, T., Lundin, J., Rannikko, A., Sankila, A., Turunen, J.-P., Lundin, M., Konsti, J., Vesterinen, T., Nordling, S., Kallioniemi, O., Hautaniemi, S. & Jänne, O.A. (2011) 'Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer', *The EMBO Journal*, 30(19), pp. 3962–3976.

Sajjad, Y., Quenby, S., Nickson, P., Lewis-Jones, D.I. & Vince, G. (2004) 'Immunohistochemical localization of androgen receptors in the urogenital tracts of human embryos', *Reproduction*, 128(3), pp. 331–339.

Sakkiah, S., Kusko, R., Pan, B., Guo, W., Ge, W., Tong, W. & Hong, H. (2018) 'Structural Changes Due to Antagonist Binding in Ligand Binding Pocket of Androgen Receptor Elucidated Through Molecular Dynamics Simulations', *Frontiers in Pharmacology*, 9.

Sakkiah, S., Ng, H.W., Tong, W. & Hong, H. (2016) 'Structures of androgen receptor bound with ligands: advancing understanding of biological functions and drug discovery', *Expert Opinion on Therapeutic Targets*, 20(10), pp. 1267–1282.

Samant, M.P., Gulyas, J., Hong, D.J., Croston, G., Rivier, C. & Rivier, J. (2005) 'Iterative approach to the discovery of novel degarelix analogues: substitutions at positions 3, 7, and 8. Part II', *Journal of Medicinal Chemistry*, 48(15), pp. 4851–4860.

Samson, D.J., Seidenfeld, J., Schmitt, B., Hasselblad, V., Albertsen, P.C., Bennett, C.L., Wilt, T.J. & Aronson, N. (2002) 'Systematic review and meta-analysis of monotherapy compared with combined androgen blockade for patients with advanced prostate carcinoma', *Cancer*, 95(2), pp. 361–376.

Santen, R.J., Van den Bossche, H., Symoens, J., Brugmans, J. & DeCoster, R. (1983) 'Site of action of low dose ketoconazole on androgen biosynthesis in men', *The Journal of Clinical Endocrinology and Metabolism*, 57(4), pp. 732–736.

Santos-Barriopedro, I., van Mierlo, G. & Vermeulen, M. (2021) 'Off-the-shelf proximity biotinylation for interaction proteomics', *Nature Communications*, 12(1), p. 5015.

Sarosdy, M.F. (1999) 'Which is the optimal antiandrogen for use in combined androgen blockade of advanced prostate cancer? The transition from a first- to second-generation antiandrogen', *Anti-Cancer Drugs*, 10(9), pp. 791–796.

Sasaki, M., Kaneuchi, M., Sakuragi, N., Fujimoto, S., Carroll, P.R. & Dahiya, R. (2003) 'The polyglycine and polyglutamine repeats in the androgen receptor gene in Japanese and Caucasian populations', *Biochemical and Biophysical Research Communications*, 312(4), pp. 1244–1247.

Schalken, J. & Fitzpatrick, J.M. (2016) 'Enzalutamide: targeting the androgen signalling pathway in metastatic castration-resistant prostate cancer', *BJU international*, 117(2), pp. 215–225.

Schaufele, F., Carbonell, X., Guerbadot, M., Borngraeber, S., Chapman, M.S., Ma, A.A.K., Miner, J.N. & Diamond, M.I. (2005) 'The structural basis of androgen receptor activation: Intramolecular and intermolecular amino–carboxy interactions', *Proceedings of the National Academy of Sciences*, 102(28), pp. 9802–9807.

Schellhammer, P., Sharifi, R., Block, N., Soloway, M., Venner, P., Patterson, A.L., Sarosdy, M., Vogelzang, N., Jones, J. & Kolvenbag, G. (1995) 'A controlled trial of bicalutamide versus flutamide, each in combination with luteinizing hormone-releasing hormone analogue therapy, in patients with advanced prostate cancer. Casodex Combination Study Group', *Urology*, 45(5), pp. 745–752.

Scher, H.I., Lu, D., Schreiber, N.A., Louw, J., Graf, R.P., Vargas, H.A., Johnson, A., Jendrisak, A., Bambury, R., Danila, D., McLaughlin, B., Wahl, J., Greene, S.B., Heller, G., Marrinucci, D., Fleisher, M. & Dittamore, R. (2016) 'Association of AR-V7 on Circulating Tumor Cells as a Treatment-Specific Biomarker With Outcomes and Survival in Castration-Resistant Prostate Cancer', *JAMA Oncology*, 2(11), pp. 1441–1449.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P. & Cardona, A. (2012) 'Fiji: an open-source platform for biological-image analysis', *Nature Methods*, 9(7), pp. 676–682.

Schmitt, B., Bennett, C., Seidenfeld, J., Samson, D. & Wilt, T.J. (1999) 'Maximal androgen blockade for advanced prostate cancer', *Cochrane Database of Systematic Reviews*, (2), .

Schneikert, J., Peterziel, H., Defossez, P.-A., Klocker, H., Launoit, Y. de & Cato, A.C.B. (1996) 'Androgen Receptor-Ets Protein Interaction Is a Novel Mechanism for Steroid Hormone-mediated Down-modulation of Matrix Metalloproteinase Expression \*', *Journal of Biological Chemistry*, 271(39), pp. 23907–23913.

Schoenmakers, E., Verrijdt, G., Peeters, B., Verhoeven, G., Rombauts, W. & Claessens, F. (2000) 'Differences in DNA Binding Characteristics of the Androgen and Glucocorticoid Receptors Can Determine Hormone-specific Responses\*', *Journal of Biological Chemistry*, 275(16), pp. 12290–12297.

Schröder, F.H. (1993) 'Cyproterone acetate—mechanism of action and clinical effectiveness in prostate cancer treatment', *Cancer*, 72(S12), pp. 3810–3815.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. & Selbach, M. (2011) 'Global quantification of mammalian gene expression control', *Nature*, 473(7347), pp. 337–342.

Schweizer, M.T. & Antonarakis, E.S. (2012) 'Abiraterone and other novel androgen-directed strategies for the treatment of prostate cancer: a new era of hormonal therapies is born', *Therapeutic Advances in Urology*, 4(4), pp. 167–178.

Scott, T., Urak, R., Soemardy, C. & Morris, K.V. (2019) 'Improved Cas9 activity by specific modifications of the tracrRNA', *Scientific Reports*, 9(1), p. 16104.

Sears, R.M., May, D.G. & Roux, K.J. (2019) 'BioID as a Tool for Protein-Proximity Labeling in Living Cells', *Methods in molecular biology (Clifton, N.J.)*, 2012pp. 299–313.

Sedelaar, J.M. & Isaacs, J.T. (2009) 'Tissue Culture Media Supplemented with 10% Fetal Calf Serum Contains a Castrate level of Testosterone', *The Prostate*, 69(16), pp. 1724–1729.

Seiler, M., Yoshimi, A., Darman, R., Chan, B., Keaney, G., Thomas, M., Agrawal, A.A., Caleb, B., Csibi, A., Sean, E., Fekkes, P., Karr, C., Klimek, V., Lai, G., Lee, L., Kumar, P., Lee, S.C.-W., Liu, X., Mackenzie, C., et al. (2018) 'H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers', *Nature Medicine*, 24(4), pp. 497–504.

Sellin, M., Mack, R., Rhodes, M.C., Zhang, L., Berg, S., Joshi, K., Liu, S., Wei, W., S. J., P.B., Larsen, P., Taylor, R.E. & Zhang, J. (2022) 'Molecular mechanisms by which splice modulator GEX1A inhibits leukaemia development and progression', *British Journal of Cancer*, 127(2), pp. 223–236.

Seydel, C. (2022) 'Diving deeper into the proteome', *Nature Methods*, 19(9), pp. 1036–1040.

Shaffer, P.L., Jivan, A., Dollins, D.E., Claessens, F. & Gewirth, D.T. (2004) 'Structural basis of androgen receptor binding to selective androgen response elements', *Proceedings of the National Academy of Sciences of the United States of America*, 101(14), pp. 4758–4763.

Shafik, A., Shafik, A.A., El Sibai, O. & Shafik, I.A. (2006) 'Contractile activity of the prostate at ejaculation: An electrophysiologic study', *Urology*, 67(4), pp. 793–796.

Shao, W., Kim, H.-S., Cao, Y., Xu, Y.-Z. & Query, C.C. (2012) 'A U1-U2 snRNP Interaction Network during Intron Definition', *Molecular and Cellular Biology*, 32(2), pp. 470–478.

Sharp, A., Coleman, I., Yuan, W., Sprenger, C., Dolling, D., Rodrigues, D.N., Russo, J.W., Figueiredo, I., Bertan, C., Seed, G., Riisnaes, R., Uo, T., Neeb, A., Welti, J., Morrissey, C., Carreira, S., Luo, J., Nelson, P.S., Balk, S.P., et al. (2019) 'Androgen receptor splice variant-7 expression emerges with castration resistance in prostate cancer', *The Journal of Clinical Investigation*, 129(1), pp. 192–208.

Shenasa, H. & Bentley, D.L. (2023) Pre-mrna splicing and its cotranscriptional connections. *Trends in Genetics*. 39 (9), 672–685.

Shet, M.S., McPhaul, M., Fisher, C.W., Stallings, N.R. & Estabrook, R.W. (1997) 'Metabolism of the antiandrogenic drug (Flutamide) by human CYP1A2', *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, 25(11), pp. 1298–1303.

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I., Severinov, K., Zhang, F. & Koonin, E.V. (2017) 'Diversity and evolution of class 2 CRISPR–Cas systems', *Nature Reviews Microbiology*, 15(3), pp. 169–182.

Shore, N.D. (2013) 'Experience with degarelix in the treatment of prostate cancer', *Therapeutic Advances in Urology*, 5(1), pp. 11–24.

Shore, N.D., Chowdhury, S., Villers, A., Klotz, L., Siemens, D.R., Phung, D., Os, S. van, Hasabou, N., Wang, F., Bhattacharya, S. & Heidenreich, A. (2016) 'Efficacy and safety of enzalutamide versus bicalutamide for patients with metastatic prostate cancer (TERRAIN): a randomised, double-blind, phase 2 study', *The Lancet Oncology*, 17(2), pp. 153–163.

Sickmier, E.A., Frato, K.E., Shen, H., Paranawithana, S.R., Green, M.R. & Kielkopf, C.L. (2006) 'Structural Basis for Polypyrimidine-Tract Recognition by the Essential pre-mRNA Splicing Factor U2AF65', *Molecular cell*, 23(1), pp. 49–59.

Siemens, D.R., Klotz, L., Heidenreich, A., Chowdhury, S., Villers, A., Baron, B., van Os, S., Hasabou, N., Wang, F., Lin, P. & Shore, N.D. (2018) 'Efficacy and Safety of Enzalutamide vs

Bicalutamide in Younger and Older Patients with Metastatic Castration Resistant Prostate Cancer in the TERRAIN Trial', *The Journal of Urology*, 199(1), pp. 147–154.

Siiteri, P.K. & Wilson, J.D. (1974) 'Testosterone formation and metabolism during male sexual differentiation in the human embryo', *The Journal of Clinical Endocrinology and Metabolism*, 38(1), pp. 113–125.

Simanainen, U., Allan, C.M., Lim, P., McPherson, S., Jimenez, M., Zajac, J.D., Davey, R.A. & Handelsman, D.J. (2007) 'Disruption of prostate epithelial androgen receptor impedes prostate lobe-specific growth and function', *Endocrinology*, 148(5), pp. 2264–2272.

Simanainen, U., McNamara, K., Gao, Y.R. & Handelsman, D.J. (2009) 'Androgen sensitivity of prostate epithelium is enhanced by postnatal androgen receptor inactivation', *American Journal of Physiology. Endocrinology and Metabolism*, 296(6), pp. E1335-1343.

Sinawe, H. & Casadesus, D. (2022) 'Ketoconazole', in *StatPearls*. [Online]. Treasure Island (FL): StatPearls Publishing.

Singh, M., Jha, R., Melamed, J., Shapiro, E., Hayward, S.W. & Lee, P. (2014) 'Stromal Androgen Receptor in Prostate Development and Cancer', *The American Journal of Pathology*, 184(10), pp. 2598–2607.

Singh, R.N. & Singh, N.N. (2018) 'Mechanism of Splicing Regulation of Spinal Muscular Atrophy Genes', *Advances in neurobiology*, 20pp. 31–61.

Sirotnak, F. m., She, Y., Khokhar, N.Z., Hayes, P., Gerald, W. & Scher, H.I. (2004) 'Microarray analysis of prostate cancer progression to reduced androgen dependence: Studies in unique models contrasts early and late molecular events', *Molecular Carcinogenesis*, 41(3), pp. 150–163.

Skoog, J., Steineck, G., Cavallin-Ståhl, E., Wilderäng, U., Håkansson, U.K., Johansson, B., Stierner, U., & SWENOTECA (2011) 'Feelings of loss and uneasiness or shame after removal of a testicle by orchidectomy: a population-based long-term follow-up of testicular cancer survivors', *International Journal of Andrology*, 34(2), pp. 183–192.

Small, E.J., Baron, A. & Bok, R. (1997) 'Simultaneous antiandrogen withdrawal and treatment with ketoconazole and hydrocortisone in patients with advanced prostate carcinoma', *Cancer*, 80(9), pp. 1755–1759.

Small, E.J., Baron, A.D., Fippin, L. & Apodaca, D. (1997) 'Ketoconazole retains activity in advanced prostate cancer patients with progression despite flutamide withdrawal', *The Journal of Urology*, 157(4), pp. 1204–1207.

Small, E.J., Halabi, S., Dawson, N.A., Stadler, W.M., Rini, B.I., Picus, J., Gable, P., Torti, F.M., Kaplan, E. & Vogelzang, N.J. (2004) 'Antiandrogen withdrawal alone or in combination with ketoconazole in androgen-independent prostate cancer patients: a phase III trial (CALGB 9583)', *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 22(6), pp. 1025–1033.

Smargon, A.A., Cox, D.B.T., Pyzocha, N.K., Zheng, K., Slaymaker, I.M., Gootenberg, J.S., Abudayyeh, O.A., Essletzbichler, P., Shmakov, S., Makarova, K.S., Koonin, E.V. & Zhang, F. (2017) 'Cas13b is a Type VI-B CRISPR-associated RNA-Guided RNase differentially regulated by accessory proteins Csx27 and Csx28', *Molecular cell*, 65(4), pp. 618-630.e7.

Smittenaar, C.R., Petersen, K.A., Stewart, K. & Moitt, N. (2016) 'Cancer incidence and mortality projections in the UK until 2035', *British Journal of Cancer*, 115(9), pp. 1147–1155.

Snapp, E. (2005) *Design and Use of Fluorescent Fusion Proteins in Cell Biology - PMC*. [Online] [online]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2875081/> (Accessed 27 February 2023).

Song, W. & Khera, M. (2014) 'Physiological normal levels of androgen inhibit proliferation of prostate cancer cells in vitro', *Asian Journal of Andrology*, 16(6), pp. 864–868.

Sowalsky, A.G., Figueiredo, I., Lis, R.T., Coleman, I., Gurel, B., Bogdan, D., Yuan, W., Russo, J.W., Bright, J.R., Whitlock, N.C., Trostel, S.Y., Ku, A.T., Patel, R.A., True, L.D., Welti, J., Jimenez-Vacas, J.M., Rodrigues, D.N., Riisnaes, R., Neeb, A., et al. (2022) 'Assessment of Androgen Receptor splice variant-7 as a biomarker of clinical response in castration-sensitive prostate cancer', *Clinical cancer research : an official journal of the American Association for Cancer Research*, 28(16), pp. 3509–3525.

Springola, M., Grate, L., Haussler, D. & Ares, M. (1999) 'Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*', *RNA (New York, N.Y.)*, 5(2), pp. 221–234.

Stanbrough, M., Bubley, G.J., Ross, K., Golub, T.R., Rubin, M.A., Penning, T.M., Febbo, P.G. & Balk, S.P. (2006) 'Increased Expression of Genes Converting Adrenal Androgens to Testosterone in Androgen-Independent Prostate Cancer', *Cancer Research*, 66(5), pp. 2815–2825.

Stanworth, R.D. & Jones, T.H. (2008) 'Testosterone for the aging male; current evidence and recommended practice', *Clinical Interventions in Aging*, 3(1), pp. 25–44.

Stasevich, T.J., Hayashi-Takanaka, Y., Sato, Y., Maehara, K., Ohkawa, Y., Sakata-Sogawa, K., Tokunaga, M., Nagase, T., Nozaki, N., McNally, J.G. & Kimura, H. (2014) Regulation of RNA polymerase II activation by histone acetylation in single living cells. *Nature*. 516 (7530), 272–275.

Steensma, D.P., Wermke, M., Klimek, V.M., Greenberg, P.L., Font, P., Komrokji, R.S., Yang, J., Brunner, A.M., Carraway, H.E., Ades, L., Al-Kali, A., Alonso Dominguez, J.M., Alonso, A., Coombs, C.C., Deeg, H.J., Donnellan, W.B., Foran, J.M., Garcia-Manero, G., Maris, M.B., et al. (2019) 'Results of a Clinical Trial of H3B-8800, a Splicing Modulator, in Patients with Myelodysplastic Syndromes (MDS), Acute Myeloid Leukemia (AML) or Chronic Myelomonocytic Leukemia (CMML)', *Blood*, 134(Supplement\_1), p. 673.

Steensma, D.P., Wermke, M., Klimek, V.M., Greenberg, P.L., Font, P., Komrokji, R.S., Yang, J., Brunner, A.M., Carraway, H.E., Ades, L., Al-Kali, A., Alonso-Dominguez, J.M., Alfonso-Piérola, A., Coombs, C.C., Deeg, H.J., Flinn, I., Foran, J.M., Garcia-Manero, G., Maris, M.B., et al. (2021)

'Phase I First-in-Human Dose Escalation Study of the oral SF3B1 modulator H3B-8800 in myeloid neoplasms', *Leukemia*, 35(12), pp. 3542–3550.

Steketee, K., Timmerman, L., Ziel-van der Made, A.C.J., Doesburg, P., Brinkmann, A.O. & Trapman, J. (2002) 'Broadened ligand responsiveness of androgen receptor mutants obtained by random amino acid substitution of H874 and mutation hot spot T877 in prostate cancer', *International Journal of Cancer*, 100(3), pp. 309–317.

Stepath, M., Zülch, B., Maghnouj, A., Schork, K., Turewicz, M., Eisenacher, M., Hahn, S., Sitek, B. & Bracht, T. (2020) 'Systematic Comparison of Label-Free, SILAC, and TMT Techniques to Study Early Adaption toward Inhibition of EGFR Signaling in the Colorectal Cancer Cell Line DiFi', *Journal of Proteome Research*, 19(2), pp. 926–937.

Stephens, M. (2017) 'False discovery rates: a new deal', *Biostatistics*, 18(2), pp. 275–294.

Stewart, R.J., Panigrahy, D., Flynn, E. & Folkman, J. (2001) 'Vascular endothelial growth factor expression and tumor angiogenesis are regulated by androgens in hormone responsive human prostate carcinoma: evidence for androgen dependent destabilization of vascular endothelial growth factor transcripts', *The Journal of Urology*, 165(2), pp. 688–693.

Stockley, J., Markert, E., Zhou, Y., Robson, C.N., Elliott, D.J., Lindberg, J., Leung, H.Y. & Rajan, P. (2015) 'The RNA-binding protein Sam68 regulates expression and transcription function of the androgen receptor splice variant AR-V7', *Scientific Reports*, 5(1), p. 13426.

Stoilov, P., Daoud, R., Nayler, O. & Stamm, S. (2004) 'Human tra2-beta1 autoregulates its protein concentration by influencing alternative splicing of its pre-mRNA', *Human Molecular Genetics*, 13(5), pp. 509–524.

Stone, N.N. & Clejan, S.J. (1991) 'Response of prostate volume, prostate-specific antigen, and testosterone to flutamide in men with benign prostatic hyperplasia', *Journal of Andrology*, 12(6), pp. 376–380.

Storbeck, M., Hupperich, K., Gaspar, J.A., Meganathan, K., Martínez Carrera, L., Wirth, R., Sachinidis, A. & Wirth, B. (2014) 'Neuronal-Specific Deficiency of the Splicing Factor Tra2b Causes Apoptosis in Neurogenic Areas of the Developing Mouse Brain', *PLoS ONE*, 9(2), p. e89020.

Stuopelyte, K., Sabaliauskaite, R., Bakavicius, A., Haflidadóttir, B.S., Visakorpi, T., Väänänen, R.-M., Patel, C., Danila, D.C., Lilja, H., Lazutka, J.R., Ulys, A., Jankevicius, F. & Jarmalaite, S. (2020) 'Analysis of AR-FL and AR-V1 in whole blood of patients with castration resistant prostate cancer as a tool for predicting response to abiraterone acetate', *The Journal of urology*, 204(1), pp. 71–78.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. & Mesirov, J.P. (2005) 'Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles', *Proceedings of the National Academy of Sciences of the United States of America*, 102(43), pp. 15545–15550.

Sugimura, Y., Cunha, G.R. & Donjacour, A.A. (1986) 'Morphological and Histological Study of Castration-Induced Degeneration and Androgen-Induced Regeneration in the Mouse Prostate1', *Biology of Reproduction*, 34(5), pp. 973–983.

Sugimoto, M., Winkler, M.H., Okeke, A.A., Benney, M. & Gillatt, D.A. (2005) 'Bicalutamide vs cyproterone acetate in preventing flare with LHRH analogue therapy for prostate cancer--a pilot study', *Prostate Cancer and Prostatic Diseases*, 8(1), pp. 91–94.

Sugiura, M., Sato, H., Okabe, A., Fukuyo, M., Mano, Y., Shinohara, K., Rahmutulla, B., Higuchi, K., Maimaiti, M., Kanesaka, M., Imamura, Y., Furihata, T., Sakamoto, S., Komiya, A., Anzai, N., Kanai, Y., Luo, J., Ichikawa, T. & Kaneda, A. (2020) 'Identification of AR-V7 downstream genes commonly targeted by AR/AR-V7 and specifically targeted by AR-V7 in castration resistant prostate cancer', *Translational Oncology*, 14(1), p. 100915.

Sun, J.S. & Manley, J.L. (1995) 'A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing', *Genes & Development*, 9(7), pp. 843–854.

Sun, R., Wei, T., Ding, D., Zhang, J., Chen, S., He, H.H., Wang, L. & Huang, H. (2022) 'CYCLIN K down-regulation induces androgen receptor gene intronic polyadenylation, variant expression and PARP inhibitor vulnerability in castration-resistant prostate cancer', *Proceedings of the National Academy of Sciences*, 119(39), p. e2205509119.

Sun, S., Sprenger, C.C.T., Vessella, R.L., Haugk, K., Soriano, K., Mostaghel, E.A., Page, S.T., Coleman, I.M., Nguyen, H.M., Sun, H., Nelson, P.S. & Plymate, S.R. (2010) 'Castration resistance in human prostate cancer is conferred by a frequently occurring androgen receptor splice variant', *The Journal of Clinical Investigation*, 120(8), pp. 2715–2730.

Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A. & Bray, F. (2021) 'Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries', *CA: A Cancer Journal for Clinicians*, 71(3), pp. 209–249.

Suzuki, H., Sato, N., Watabe, Y., Masai, M., Seino, S. & Shimazaki, J. (1993) 'Androgen receptor gene mutations in human prostate cancer', *The Journal of Steroid Biochemistry and Molecular Biology*, 46(6), pp. 759–765.

Svensson, J., Lissbrant, I.F., Gauffin, O., Hjälm-Eriksson, M., Kilany, S., Fagerlund, K. & Stattin, P. (2021) 'Time spent in hormone-sensitive and castration-resistant disease states in men with advanced prostate cancer, and its health economic impact: registry-based study in Sweden', *Scandinavian Journal of Urology*, 55(1), pp. 1–8.

Sweeney, C.J., Chen, Y.-H., Carducci, M., Liu, G., Jarrard, D.F., Eisenberger, M., Wong, Y.-N., Hahn, N., Kohli, M., Cooney, M.M., Dreicer, R., Vogelzang, N.J., Picus, J., Shevrin, D., Hussain, M., Garcia, J.A. & DiPaola, R.S. (2015) 'Chemohormonal Therapy in Metastatic Hormone-Sensitive Prostate Cancer', *New England Journal of Medicine*, 373(8), pp. 737–746.

Swerdloff, R.S., Dudley, R.E., Page, S.T., Wang, C. & Salameh, W.A. (2017) 'Dihydrotestosterone: Biochemistry, Physiology, and Clinical Implications of Elevated Blood Levels', *Endocrine Reviews*, 38(3), pp. 220–254.

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A.L., Fang, T., Doncheva, N.T., Pyysalo, S., Bork, P., Jensen, L.J. & von Mering, C. (2022) 'The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest', *Nucleic Acids Research*, 51(D1), pp. D638–D646.

Tacke, R., Tohyama, M., Ogawa, S. & Manley, J.L. (1998) 'Human Tra2 Proteins Are Sequence-Specific Activators of Pre-mRNA Splicing', *Cell*, 93(1), pp. 139–148.

Taggart, A.J., DeSimone, A.M., Shih, J.S., Filloux, M.E. & Fairbrother, W.G. (2012) 'Large-scale mapping of branchpoints in human pre-mRNA transcripts in vivo', *Nature structural & molecular biology*, 19(7), pp. 719–721.

Tai, S., Sun, Y., Squires, J.M., Zhang, H., Oh, W.K., Liang, C.-Z. & Huang, J. (2011) 'PC3 Is a Cell Line Characteristic of Prostatic Small Cell Carcinoma', *The Prostate*, 71(15), p. 1668.

Takata, H., Nishijima, H., Ogura, S., Sakaguchi, T., Bubulya, P.A., Mochizuki, T. & Shibahara, K. (2009) 'Proteome Analysis of Human Nuclear Insoluble Fractions', *Genes to cells : devoted to molecular & cellular mechanisms*, 14(8), pp. 975–990.

Takayama, K., Suzuki, T., Fujimura, T., Yamada, Y., Takahashi, S., Homma, Y., Suzuki, Y. & Inoue, S. (2017) 'Dysregulation of spliceosome gene expression in advanced prostate cancer by RNA-binding protein PSF', *Proceedings of the National Academy of Sciences*, 114(39), pp. 10461–10466.

Tamada, S., Iguchi, T., Kato, M., Asakawa, J., Kita, K., Yasuda, S., Yamasaki, T., Matsuoka, Y., Yamaguchi, K., Matsumura, K., Go, I., Ohmachi, T. & Nakatani, T. (2018) 'Time to progression to castration-resistant prostate cancer after commencing combined androgen blockade for advanced hormone-sensitive prostate cancer', *Oncotarget*, 9(97), pp. 36966–36974.

Tambe, A., East-Seletsky, A., Knott, G.J., Doudna, J.A. & O'Connell, M.R. (2018) 'RNA-binding and HEPN-nuclease activation are decoupled in CRISPR-Cas13a', *Cell reports*, 24(4), pp. 1025–1036.

Tan, B., Peng, S., Yatim, S.M.J.M., Gunaratne, J., Hunziker, W. & Ludwig, A. (2020) 'An Optimized Protocol for Proximity Biotinylation in Confluent Epithelial Cell Cultures Using the Peroxidase APEX2', *STAR Protocols*, 1(2), p. 100074.

Tan, M.E., Li, J., Xu, H.E., Melcher, K. & Yong, E. (2015) 'Androgen receptor: structure, role in prostate cancer and drug discovery', *Acta Pharmacologica Sinica*, 36(1), pp. 3–23.

Tang, L. (2019) 'Exploring class 1 CRISPR systems', *Nature Methods*, 16(11), pp. 1079–1079.

Tang, S., Sethunath, V., Metaferia, N.Y., Nogueira, M.F., Gallant, D.S., Garner, E.R., Lairson, L.A., Penney, C.M., Li, J., Gelbard, M.K., Abou Alaiwi, S., Seo, J.-H., Hwang, J.H., Strathdee, C.A., Baca, S.C., AbuHammad, S., Zhang, X., Doench, J.G., Hahn, W.C., et al. (2022) 'A genome-scale CRISPR screen reveals PRMT1 as a critical regulator of androgen receptor signaling in prostate cancer', *Cell reports*, 38(8), p. 110417.

Tannock, I.F., de Wit, R., Berry, W.R., Horti, J., Pluzanska, A., Chi, K.N., Oudard, S., Théodore, C., James, N.D., Turesson, I., Rosenthal, M.A. & Eisenberger, M.A. (2004) 'Docetaxel plus

Prednisone or Mitoxantrone plus Prednisone for Advanced Prostate Cancer', *New England Journal of Medicine*, 351(15), pp. 1502–1512.

Taplin, M.-E., Bubley, G.J., Ko, Y.-J., Small, E.J., Upton, M., Rajeshkumar, B. & Balk, S.P. (1999) 'Selection for Androgen Receptor Mutations in Prostate Cancers Treated with Androgen Antagonist1', *Cancer Research*, 59(11), pp. 2511–2515.

Taplin, M.-E., Bubley, G.J., Shuster, T.D., Frantz, M.E., Spooner, A.E., Ogata, G.K., Keer, H.N. & Balk, S.P. (1995) 'Mutation of the Androgen-Receptor Gene in Metastatic Androgen-Independent Prostate Cancer', *New England Journal of Medicine*, 332(21), pp. 1393–1398.

Tarn, W.Y. & Steitz, J.A. (1994) 'SR proteins can compensate for the loss of U1 snRNP functions in vitro.', *Genes & Development*, 8(22), pp. 2704–2717.

TCGA (2015) 'The molecular taxonomy of primary prostate cancer', *Cell*, 163(4), pp. 1011–1025.

Tepper, C.G., Boucher, D.L., Ryan, P.E., Ma, A.-H., Xia, L., Lee, L.-F., Pretlow, T.G. & Kung, H.-J. (2002) 'Characterization of a Novel Androgen Receptor Mutation in a Relapsed CWR22 Prostate Cancer Xenograft and Cell Line1', *Cancer Research*, 62(22), pp. 6606–6614.

Teroerde, M., Nientiedt, C., Duensing, A., Hohenfellner, M., Stenzinger, A. & Duensing, S. (2021) 'Revisiting the Role of p53 in Prostate Cancer', in Simon RJ Bott & Keng Lim Ng (eds.) *Prostate Cancer*. [Online]. Brisbane (AU): Exon Publications.

Thangapazham, R., Saenz, F., Katta, S., Mohamed, A.A., Tan, S.-H., Petrovics, G., Srivastava, S. & Dobi, A. (2014) 'Loss of the NKX3.1 tumorsuppressor promotes the TMPRSS2-ERG fusion gene expression in prostate cancer', *BMC Cancer*, 14(1), p. 16.

Thangavel, C., Boopathi, E., Liu, Y., Haber, A., Ertel, A., Bhardwaj, A., Addya, S., Williams, N., Ciment, S.J., Cotzia, P., Dean, J.L., Snook, A., McNair, C., Price, M., Hernandez, J.R., Zhao, S.G., Birbe, R., McCarthy, J.B., Turley, E.A., et al. (2017) 'RB Loss Promotes Prostate Cancer Metastasis', *Cancer Research*, 77(4), pp. 982–995.

Thirumalai, A., Cooper, L.A., Rubinow, K.B., Amory, J.K., Lin, D.W., Wright, J.L., Marck, B.T., Matsumoto, A.M. & Page, S.T. (2016) 'Stable Intraprostatic Dihydrotestosterone in Healthy Medically Castrate Men Treated With Exogenous Testosterone', *The Journal of Clinical Endocrinology and Metabolism*, 101(7), pp. 2937–2944.

Thomas, M.A., Preece, D.M. & Bentel, J.M. (2010) 'Androgen regulation of the prostatic tumour suppressor NKX3.1 is mediated by its 3' untranslated region', *Biochemical Journal*, 425(3), pp. 575–583.

Thomas, P. & Smart, T.G. (2005) 'HEK293 cell line: A vehicle for the expression of recombinant proteins', *Journal of Pharmacological and Toxicological Methods*, 51(3), pp. 187–200.

Thompson, I.M. (2001) 'Flare Associated with LHRH-Agonist Therapy', *Reviews in Urology*, 3(Suppl 3), pp. S10–S14.

Thomson, A. (2005) 'Dutasteride: an evidence-based review of its clinical impact in the treatment of benign prostatic hyperplasia', *Core Evidence*, 1(2), pp. 143–156.

Thurman, M., van Doorn, J., Danzer, B., Webb, T.R. & Stamm, S. (2017) 'Changes in Alternative Splicing as Pharmacodynamic Markers for Sudemycin D6', *Biomarker Insights*, 12p. 1177271917730557.

Tien, J.C.-Y., Liu, Z., Liao, L., Wang, F., Xu, Y., Wu, Y.-L., Zhou, N., Ittmann, M. & Xu, J. (2013) 'The Steroid Receptor Coactivator-3 Is Required for the Development of Castration-resistant Prostate Cancer', *Cancer research*, 73(13), pp. 3997–4008.

Tilley, W.D., Buchanan, G., Hickey, T.E. & Bentel, J.M. (1996) 'Mutations in the androgen receptor gene are associated with progression of human prostate cancer to androgen independence.', *Clinical Cancer Research*, 2(2), pp. 277–285.

Toivanen, R. & Shen, M.M. (2017) 'Prostate organogenesis: tissue induction, hormonal regulation and cell type specification', *Development (Cambridge, England)*, 144(8), pp. 1382–1398.

Tomlins, Scott A, Laxman, B., Varambally, S., Cao, X., Yu, J., Helgeson, B.E., Cao, Q., Prensner, J.R., Rubin, M.A., Shah, R.B., Mehra, R. & Chinnaiyan, A.M. (2008) 'Role of the TMPRSS2-ERG Gene Fusion in Prostate Cancer', *Neoplasia (New York, N.Y.)*, 10(2), pp. 177–188.

Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.-W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., Lee, C., Montie, J.E., Shah, R.B., Pienta, K.J., Rubin, M.A. & Chinnaiyan, A.M. (2005) 'Recurrent Fusion of TMPRSS2 and ETS Transcription Factor Genes in Prostate Cancer', *Science*, 310(5748), pp. 644–648.

Tomlins, Scott A., Rhodes, D.R., Yu, J., Varambally, S., Mehra, R., Perner, S., Demichelis, F., Helgeson, B.E., Laxman, B., Morris, D.S., Cao, Q., Cao, X., Andrén, O., Fall, K., Johnson, L., Wei, J.T., Shah, R.B., Al-Ahmadie, H., Eastham, J.A., et al. (2008) 'The Role of SPINK1 in ETS Rearrangement Negative Prostate Cancers', *Cancer cell*, 13(6), pp. 519–528.

Trachtenberg, J., Halpern, N. & Pont, A. (1983) 'Ketoconazole: A Novel and Rapid Treatment for Advanced Prostatic Cancer', *The Journal of Urology*, 130(1), pp. 152–153.

Trachtenberg, J. & Pont, A. (1984) 'KETOCONAZOLE THERAPY FOR ADVANCED PROSTATE CANCER', *The Lancet*, 324(8400), pp. 433–435.

Tran, C., Ouk, S., Clegg, N.J., Chen, Y., Watson, P.A., Arora, V., Wongvipat, J., Smith-Jones, P.M., Yoo, D., Kwon, A., Wasieleska, T., Welsbie, D., Chen, C., Higano, C.S., Beer, T.M., Hung, D.T., Scher, H.I., Jung, M. & Sawyers, C.L. (2009) 'Development of a Second-Generation Antiandrogen for Treatment of Advanced Prostate Cancer', *Science (New York, N.Y.)*, 324(5928), pp. 787–790.

Travison, T.G., Vesper, H.W., Orwoll, E., Wu, F., Kaufman, J.M., Wang, Y., Lapauw, B., Fiers, T., Matsumoto, A.M. & Bhasin, S. (2017) 'Harmonized Reference Ranges for Circulating Testosterone Levels in Men of Four Cohort Studies in the United States and Europe', *The Journal of Clinical Endocrinology and Metabolism*, 102(4), pp. 1161–1173.

Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., Meyers, R.M., Ali, L., Goodale, A., Lee, Y., Jiang, G., Hsiao, J., Gerath, W.F.J., Howell, S., Merkel, E., et al. (2017) 'Defining a Cancer Dependency Map', *Cell*, 170(3), pp. 564-576.e16.

Tsuda, K., Someya, T., Kuwasako, K., Takahashi, M., He, F., Unzai, S., Inoue, M., Harada, T., Watanabe, S., Terada, T., Kobayashi, N., Shirouzu, M., Kigawa, T., Tanaka, A., Sugano, S., Güntert, P., Yokoyama, S. & Muto, Y. (2011) 'Structural basis for the dual RNA-recognition modes of human Tra2- $\beta$  RRM', *Nucleic Acids Research*, 39(4), pp. 1538-1553.

Tummala, R., Lou, W., Gao, A.C. & Nadiminty, N. (2017) 'Quercetin Targets hnRNPA1 to Overcome Enzalutamide Resistance in Prostate Cancer Cells', *Molecular Cancer Therapeutics*, 16(12), pp. 2770-2779.

Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., Velankar, S., Kleywegt, G.J., Bateman, A., Evans, R., Pritzel, A., Figurnov, M., Ronneberger, O., Bates, R., Kohl, S.A.A., et al. (2021) 'Highly accurate protein structure prediction for the human proteome', *Nature*, 596(7873), pp. 590-596.

Tut, T.G., Ghadessy, F.J., Trifiro, M.A., Pinsky, L. & Yong, E.L. (1997) 'Long polyglutamine tracts in the androgen receptor are associated with reduced trans-activation, impaired sperm production, and male infertility', *The Journal of Clinical Endocrinology and Metabolism*, 82(11), pp. 3777-3782.

Tytgat, H.L.P., Schoofs, G., Driesen, M., Proost, P., Damme, E.J.M.V., Vanderleyden, J. & Lebeer, S. (2015) 'Endogenous biotin-binding proteins: an overlooked factor causing false positives in streptavidin-based protein detection', *Microbial Biotechnology*, 8(1), p. 164.

Ueda, T., Mawji, N.R., Bruchovsky, N. & Sadar, M.D. (2002) 'Ligand-independent activation of the androgen receptor by interleukin-6 and the role of steroid receptor coactivator-1 in prostate cancer cells', *The Journal of Biological Chemistry*, 277(41), pp. 38087-38094.

Urabe, V.K., Stevers, M., Ghosh, A.K. & Jurica, M.S. (2021) 'U2 snRNA structure is influenced by SF3A and SF3B proteins but not by SF3B inhibitors', *PLOS ONE*, 16(10), p. e0258551.

Urlaub, H., Raker, V.A., Kostka, S. & Lührmann, R. (2001) 'Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure', *The EMBO Journal*, 20(1-2), pp. 187-196.

Uttley, L., Whyte, S., Gomersall, T., Ren, S., Wong, R., Chambers, D. & Tappenden, P. (2017) 'Degarelix for Treating Advanced Hormone-Dependent Prostate Cancer: An Evidence Review Group Perspective of a NICE Single Technology Appraisal', *PharmacoEconomics*, 35(7), pp. 717-726.

Uzor, S., Porazinski, S.R., Li, L., Clark, B., Ajiro, M., Iida, K., Hagiwara, M., Alqasem, A.A., Perks, C.M., Wilson, I.D., Oltean, S. & Ladomery, M.R. (2021) 'CDC2-like (CLK) protein kinase inhibition as a novel targeted therapeutic strategy in prostate cancer', *Scientific Reports*, 11(1), p. 7963.

Vaishampayan, U.N., Heilbrun, L.K., Monk, P., III, Tejwani, S., Sonpavde, G., Hwang, C., Smith, D., Jasti, P., Dobson, K., Dickow, B., Heath, E.I., Semaan, L., Cher, M.L., Fontana, J.A. & Chinni, S. (2021) 'Clinical Efficacy of Enzalutamide vs Bicalutamide Combined With Androgen Deprivation Therapy in Men With Metastatic Hormone-Sensitive Prostate Cancer: A Randomized Clinical Trial', *JAMA Network Open*, 4(1), p. e2034633.

Välikangas, T., Suomi, T. & Elo, L.L. (2018) 'A systematic evaluation of normalization methods in quantitative label-free proteomics', *Briefings in Bioinformatics*, 19(1), pp. 1–11.

Van Etten, J.L., Nyquist, M., Li, Y., Yang, R., Ho, Y., Johnson, R., Ondigi, O., Voytas, D.F., Henzler, C. & Dehm, S.M. (2017) 'Targeting a single alternative polyadenylation site coordinately blocks expression of androgen receptor mRNA splice variants in prostate cancer', *Cancer research*, 77(19), pp. 5228–5235.

Vander Griend, D.J., Litvinov, I.V. & Isaacs, J.T. (2014) 'Conversion of Androgen Receptor Signaling From a Growth Suppressor in Normal Prostate Epithelial Cells to an Oncogene in Prostate Cancer Cells Involves a Gain of Function in c-Myc Regulation', *International Journal of Biological Sciences*, 10(6), pp. 627–642.

Varambally, S., Dhanasekaran, S.M., Zhou, M., Barrette, T.R., Kumar-Sinha, C., Sanda, M.G., Ghosh, D., Pienta, K.J., Sewalt, R.G.A.B., Otte, A.P., Rubin, M.A. & Chinnaiyan, A.M. (2002) 'The polycomb group protein EZH2 is involved in progression of prostate cancer', *Nature*, 419(6907), pp. 624–629.

Varia, S., Potabathula, D., Deng, Z., Bubulya, A. & Bubulya, P.A. (2013) 'Btf and TRAP150 have distinct roles in regulating subcellular mRNA distribution', *Nucleus*, 4(3), pp. 229–240.

Vasaitis, T.S., Bruno, R.D. & Njar, V.C.O. (2011) 'CYP17 inhibitors for prostate cancer therapy', *The Journal of steroid biochemistry and molecular biology*, 125(1–2), pp. 23–31.

Veldscholte, J., Berrevoets, C.A., Ris-Stalpers, C., Kuiper, G.G., Jenster, G., Trapman, J., Brinkmann, A.O. & Mulder, E. (1992) 'The androgen receptor in LNCaP cells contains a mutation in the ligand binding domain which affects steroid binding characteristics and response to antiandrogens', *The Journal of Steroid Biochemistry and Molecular Biology*, 41(3–8), pp. 665–669.

Veldscholte, J., Berrevoets, C.A., Zegers, N.D., van der Kwast, T.H., Grootegoed, J.A. & Mulder, E. (1992) 'Hormone-induced dissociation of the androgen receptor-heat-shock protein complex: use of a new monoclonal antibody to distinguish transformed from nontransformed receptors', *Biochemistry*, 31(32), pp. 7422–7430.

Veldscholte, J., Ris-Stalpers, C., Kuiper, G.G.J.M., Jenster, G., Berrevoets, C., Claassen, E., van Rooij, H.C.J., Trapman, J., Brinkmann, A.O. & Mulder, E. (1990) 'A mutation in the ligand binding domain of the androgen receptor of human LNCaP cells affects steroid binding characteristics and response to anti-androgens', *Biochemical and Biophysical Research Communications*, 173(2), pp. 534–540.

Vieira, J.G., Nishida, S.K., Pereira, A.B., Arraes, R.F. & Verreschi, I.T. (1994) 'Serum levels of prostate-specific antigen in normal boys throughout puberty', *The Journal of Clinical Endocrinology and Metabolism*, 78(5), pp. 1185–1187.

Vietri, M.T., D'Elia, G., Caliendo, G., Resse, M., Casamassimi, A., Passariello, L., Albanese, L., Cioffi, M. & Molinari, A.M. (2021) 'Hereditary Prostate Cancer: Genes Related, Target Therapy and Prevention', *International Journal of Molecular Sciences*, 22(7), p. 3753.

Visakorpi, T., Hyytinen, E., Koivisto, P., Tanner, M., Keinänen, R., Palmberg, C., Palotie, A., Tammela, T., Isola, J. & Kallioniemi, O.P. (1995) 'In vivo amplification of the androgen receptor gene and progression of human prostate cancer', *Nature Genetics*, 9(4), pp. 401–406.

Vlasenok, M., Margasyuk, S. & Pervouchine, D.D. (2022) Transcriptome sequencing suggests that pre-mrna splicing counteracts widespread intronic cleavage and polyadenylation. *NAR Genomics and Bioinformatics*. 5 (2), .

Vohhodina, J., Barros, E.M., Savage, A.L., Liberante, F.G., Manti, L., Bankhead, P., Cosgrove, N., Madden, A.F., Harkin, D.P. & Savage, K.I. (2017) 'The RNA processing factors THRAP3 and BCLAF1 promote the DNA damage response through selective mRNA splicing and nuclear export', *Nucleic Acids Research*, 45(22), pp. 12816–12833.

Wadosky, K.M. & Koochekpour, S. (2017) 'Androgen receptor splice variants and prostate cancer: From bench to bedside', *Oncotarget*, 8(11), pp. 18550–18576.

Wakeling, A.E., Furr, B.J., Glen, A.T. & Hughes, L.R. (1981) 'Receptor binding and biological activity of steroid and nonsteroidal antiandrogens', *Journal of Steroid Biochemistry*, 15pp. 355–359.

Waltering, K.K., Helenius, M.A., Sahu, B., Manni, V., Linja, M.J., Jänne, O.A. & Visakorpi, T. (2009) 'Increased Expression of Androgen Receptor Sensitizes Prostate Cancer Cells to Low Levels of Androgens', *Cancer Research*, 69(20), pp. 8141–8149.

Wan, R., Yan, C., Bai, R., Wang, L., Huang, M., Wong, C.C.L. & Shi, Y. (2016) 'The 3.8 Å structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis', *Science*, 351(6272), pp. 466–475.

Wang, B., Lo, U.-G., Wu, K., Kapur, P., Liu, X., Huang, J., Chen, W., Hernandez, E., Santoyo, J., Ma, S.-H., Pong, R.-C., He, D., Cheng, Y.-Q. & Hsieh, J.-T. (2017) 'Developing new targeting strategy for androgen receptor variants in castration resistant prostate cancer', *International journal of cancer*, 141(10), pp. 2121–2130.

Wang, D. & Farhana, A. (2022) 'Biochemistry, RNA Structure', in *StatPearls*. [Online]. Treasure Island (FL): StatPearls Publishing.

Wang, J., Youkharibache, P., Marchler-Bauer, A., Lanczycki, C., Zhang, D., Lu, S., Madej, T., Marchler, G.H., Cheng, T., Chong, L.C., Zhao, S., Yang, K., Lin, J., Cheng, Z., Dunn, R., Malkaram, S.A., Tai, C.-H., Enoma, D., Busby, B., et al. (2022) 'iCn3D: From Web-Based 3D Viewer to Structural Analysis Tool in Batch Mode', *Frontiers in Molecular Biosciences*, 9p. 831740.

Wang, L.-T., Proulx, M.-È., Kim, A.D., Lelarge, V. & McCaffrey, L. (2021) 'A proximity proteomics screen in three-dimensional spheroid cultures identifies novel regulators of lumen formation', *Scientific Reports*, 11(1), p. 22807.

Wang, Q., Li, W., Liu, X.S., Carroll, J.S., Jänne, O.A., Keeton, E.K., Chinnaiyan, A.M., Pienta, K.J. & Brown, M. (2007) 'A Hierarchical Network of Transcription Factors Governs Androgen Receptor-Dependent Prostate Cancer Growth', *Molecular cell*, 27(3), pp. 380–392.

Wang, S., Gao, J., Lei, Q., Rozengurt, N., Pritchard, C., Jiao, J., Thomas, G.V., Li, G., Roy-Burman, P., Nelson, P.S., Liu, X. & Wu, H. (2003) 'Prostate-specific deletion of the murine Pten tumor suppressor gene leads to metastatic prostate cancer', *Cancer Cell*, 4(3), pp. 209–221.

Wang, S., Li, W., Hu, L., Cheng, J., Yang, H. & Liu, Y. (2020) 'NAguideR: performing and prioritizing missing value imputations for consistent bottom-up proteomic analyses', *Nucleic Acids Research*, 48(14), p. e83.

Wang, S., Li, W., Liu, S. & Xu, J. (2016) 'RaptorX-Property: a web server for protein structure property prediction', *Nucleic Acids Research*, 44(Web Server issue), pp. W430–W435.

Wang, S., Lin, C.-W., Carleton, A.E., Cortez, C.L., Johnson, C., Taniguchi, L.E., Sekulovski, N., Townshend, R.F., Basrur, V., Nesvizhskii, A.I., Zou, P., Fu, J., Gumucio, D.L., Duncan, M.C. & Taniguchi, K. (2021) 'Spatially resolved cell polarity proteomics of a human epiblast model', *Science Advances*, 7(17), p. eabd8407.

Wang, Y., Dai, B. & Ye, D.-W. (2017) 'Serum testosterone level predicts the effective time of androgen deprivation therapy in metastatic prostate cancer patients', *Asian Journal of Andrology*, 19(2), pp. 178–183.

Wang, Z., Wang, Y., Zhang, J., Hu, Q., Zhi, F., Zhang, S., Mao, D., Zhang, Y. & Liang, H. (2017) 'Significance of the TMPRSS2:ERG gene fusion in prostate cancer', *Molecular Medicine Reports*, 16(4), pp. 5450–5458.

Wärnmark, A., Treuter, E., Wright, A.P.H. & Gustafsson, J.-Å. (2003) 'Activation Functions 1 and 2 of Nuclear Receptors: Molecular Strategies for Transcriptional Activation', *Molecular Endocrinology*, 17(10), pp. 1901–1909.

Watson, P.A., Chen, Y.F., Balbas, M.D., Wongvipat, J., Soccia, N.D., Viale, A., Kim, K. & Sawyers, C.L. (2010) 'Constitutively active androgen receptor splice variants expressed in castration-resistant prostate cancer require full-length androgen receptor', *Proceedings of the National Academy of Sciences of the United States of America*, 107(39), pp. 16759–16765.

Waxman, J., Man, A., Hendry, W.F., Whitfield, H.N., Besser, G.M., Tiptaft, R.C., Paris, A.M. & Oliver, R.T. (1985) 'Importance of early tumour exacerbation in patients treated with long acting analogues of gonadotrophin releasing hormone for advanced prostatic cancer.', *British Medical Journal (Clinical research ed.)*, 291(6506), pp. 1387–1388.

Webb-Robertson, B.-J.M., Wiberg, H.K., Matzke, M.M., Brown, J.N., Wang, J., McDermott, J.E., Smith, R.D., Rodland, K.D., Metz, T.O., Pounds, J.G. & Waters, K.M. (2015) 'Review, Evaluation,

and Discussion of the Challenges of Missing Value Imputation for Mass Spectrometry-Based Label-Free Global Proteomics', *Journal of Proteome Research*, 14(5), pp. 1993–2001.

Wei, R., Wang, J., Su, M., Jia, E., Chen, S., Chen, T. & Ni, Y. (2018) 'Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data', *Scientific Reports*, 8(1), p. 663.

Wei, W., Riley, N.M., Yang, A.C., Kim, J.T., Terrell, S.M., Li, V.L., Garcia-Contreras, M., Bertozzi, C.R. & Long, J.Z. (2021) 'Cell type-selective secretome profiling in vivo', *Nature chemical biology*, 17(3), pp. 326–334.

Welsh, M., Moffat, L., McNeilly, A., Brownstein, D., Saunders, P.T.K., Sharpe, R.M. & Smith, L.B. (2011) 'Smooth Muscle Cell-Specific Knockout of Androgen Receptor: A New Model for Prostatic Disease', *Endocrinology*, 152(9), pp. 3541–3551.

Welti, J., Rodrigues, D.N., Sharp, A., Sun, S., Lorente, D., Riisnaes, R., Figueiredo, I., Zafeiriou, Z., Rescigno, P., de Bono, J.S. & Plymate, S.R. (2016) 'Analytical Validation and Clinical Qualification of a New Immunohistochemical Assay for Androgen Receptor Splice Variant-7 Protein Expression in Metastatic Castration-resistant Prostate Cancer', *European Urology*, 70(4), pp. 599–608.

Welti, J., Sharp, A., Brooks, N., Yuan, W., McNair, C., Chand, S.N., Pal, A., Figueiredo, I., Riisnaes, R., Gurel, B., Rekowski, J., Bogdan, D., West, W., Young, B., Raja, M., Prosser, A., Lane, J., Thomson, S., Worthington, J., et al. (2021) 'Targeting p300/CBP in lethal prostate cancer', *Cancer discovery*, 11(5), pp. 1118–1137.

Wen, S., Niu, Y. & Huang, H. (2020) 'Posttranslational regulation of androgen dependent and independent androgen receptor activities in prostate cancer', *Asian Journal of Urology*, 7(3), pp. 203–218.

Wenzel, M., Preisser, F., Hoeh, B., Schroeder, M., Würnschimmel, C., Steuber, T., Heinzer, H., Banek, S., Ahrens, M., Becker, A., Karakiewicz, P.I., Chun, F.K.H., Kluth, L.A. & Mandel, P. (2021) 'Impact of Time to Castration Resistance on Survival in Metastatic Hormone Sensitive Prostate Cancer Patients in the Era of Combination Therapies', *Frontiers in Oncology*, 11.

Wessels, H.-H., Méndez-Mancilla, A., Guo, X., Legut, M., Daniloski, Z. & Sanjana, N.E. (2020) 'Massively parallel Cas13 screens reveal principles for guide RNA design', *Nature Biotechnology*, 38(6), pp. 722–727.

Westin, P., Stattin, P., Damber, J.E. & Bergh, A. (1995) 'Castration therapy rapidly induces apoptosis in a minority and decreases cell proliferation in a majority of human prostatic tumors', *The American Journal of Pathology*, 146(6), pp. 1368–1375.

Weyer-Czernilofsky, U., Hofmann, M.H., Friedbichler, K., Baumgartinger, R., Adam, P.J., Solca, F., Kraut, N., Nguyen, H.M., Corey, E., Liu, G., Sprenger, C.C., Plymate, S.R. & Bogenrieder, T. (2020) 'Antitumor Activity of the IGF-1/IGF-2-Neutralizing Antibody Xentuzumab (BI 836845) in Combination with Enzalutamide in Prostate Cancer Models', *Molecular Cancer Therapeutics*, 19(4), pp. 1059–1069.

Whitacre, D.C., Chauhan, S., Davis, T., Gordon, D., Cress, A.E. & Miesfeld, R.L. (2002) 'Androgen induction of in vitro prostate cell differentiation', *Cell Growth & Differentiation: The Molecular Biology Journal of the American Association for Cancer Research*, 13(1), pp. 1–11.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., et al. (2019) 'Welcome to the Tidyverse', *Journal of Open Source Software*, 4(43), p. 1686.

Wilkinson, M.E., Charenton, C. & Nagai, K. (2020) 'RNA Splicing by the Spliceosome', *Annual Review of Biochemistry*, 89(1), pp. 359–388.

Will, C.L. & Lührmann, R. (2011) 'Spliceosome Structure and Function', *Cold Spring Harbor Perspectives in Biology*, 3(7), p. a003707.

Wilson, C., Chen, P.J., Miao, Z. & Liu, D.R. (2020) 'Programmable m6A modification of cellular RNAs with a Cas13-directed methyltransferase', *Nature Biotechnology*, 38(12), pp. 1431–1440.

Wilson, C.M. & McPhaul, M.J. (1996) 'A and B forms of the androgen receptor are expressed in a variety of human tissues', *Molecular and Cellular Endocrinology*, 120(1), pp. 51–57.

Wirth, M.P., Hakenberg, O.W. & Froehner, M. (2007) 'Antiandrogens in the Treatment of Prostate Cancer', *European Urology*, 51(2), pp. 306–314.

Wojtuszkiewicz, A., Assaraf, Y.G., Jansen, G., Koide, K., Bressin, R.K., Basu, U., Sonneveld, E., Giovannetti, E., Kaspers, G. & Cloos, J. (2014) 'Spliceosome Inhibitor Meayamycin B As a Novel Potential Chemotherapeutic Agent in ALL and AML', *Blood*, 124(21), p. 924.

Wong, C., Kelce, W.R., Sar, M. & Wilson, E.M. (1995) 'Androgen Receptor Antagonist versus Agonist Activities of the Fungicide Vinclozolin Relative to Hydroxyflutamide (\*')', *Journal of Biological Chemistry*, 270(34), pp. 19998–20003.

Woo, H., Dam Ha, S., Lee, S.B., Buratowski, S. & Kim, T. (2017) Modulation of gene expression dynamics by co-transcriptional histone methylations. *Experimental & Molecular Medicine*. 49 (4), .

Word, R.A., George, F.W., Wilson, J.D. & Carr, B.R. (1989) 'Testosterone synthesis and adenylate cyclase activity in the early human fetal testis appear to be independent of human chorionic gonadotropin control', *The Journal of Clinical Endocrinology and Metabolism*, 69(1), pp. 204–208.

Wu, C.-T., Altuwaijri, S., Ricke, W.A., Huang, S.-P., Yeh, S., Zhang, C., Niu, Y., Tsai, M.-Y. & Chang, C. (2007) 'Increased prostate cell proliferation and loss of cell differentiation in mice lacking prostate epithelial androgen receptor', *Proceedings of the National Academy of Sciences of the United States of America*, 104(31), pp. 12679–12684.

Wu, D., Jiang, S., Bowler, M.W. & Song, H. (2012) 'Crystal Structures of Lsm3, Lsm4 and Lsm5/6/7 from *Schizosaccharomyces pombe*', *PLoS ONE*, 7(5), p. e36768.

Wu, D., Lim, E., Vaillant, F., Asselin-Labat, M.-L., Visvader, J.E. & Smyth, G.K. (2010) 'ROAST: rotation gene set tests for complex microarray experiments', *Bioinformatics*, 26(17), pp. 2176–2182.

Wu, E., Guo, X., Teng, X., Zhang, R., Li, F., Cui, Y., Zhang, D., Liu, Q., Luo, J., Wang, J. & Chen, R. (2021) 'Discovery of Plasma Membrane-Associated RNAs through APEX-seq', *Cell Biochemistry and Biophysics*, 79(4), pp. 905–917.

Wu, F., Ding, S., Li, X., Wang, H., Liu, S., Wu, H., Bi, D., Ding, K. & Lu, J. (2016) 'Elevated expression of HIF-1 $\alpha$  in actively growing prostate tissues is associated with clinical features of benign prostatic hyperplasia', *Oncotarget*, 7(11), pp. 12053–12062.

Wu, Q.-W. & Kapfhammer, J.P. (2021) 'The Bacterial Enzyme Cas13 Interferes with Neurite Outgrowth from Cultured Cortical Neurons', *Toxins*, 13(4), p. 262.

Wu, X., Scott, H., Carlsson, S.V., Sjoberg, D.D., Cerundolo, L., Lilja, H., Prevo, R., Rieunier, G., Macaulay, V., Higgins, G.S., Verrill, C.L., Lamb, A.D., Cunliffe, V.T., Bountra, C., Hamdy, F.C. & Bryant, R.J. (2019) 'Increased EZH2 expression in prostate cancer is associated with metastatic recurrence following external beam radiotherapy', *The Prostate*, 79(10), pp. 1079–1089.

Wüstmann, N., Seitzer, K., Humberg, V., Vieler, J., Grundmann, N., Steinestel, J., Tiedje, D., Duensing, S., Krabbe, L.-M., Bögemann, M., Schrader, A.J., Bernemann, C. & Schlack, K. (2023) 'Co-expression and clinical utility of AR-FL and AR splice variants AR-V3, AR-V7 and AR-V9 in prostate cancer', *Biomarker Research*, 11(1), p. 37.

Wyatt, A.W., Azad, A.A., Volik, S.V., Annala, M., Beja, K., McConeghy, B., Haegert, A., Warner, E.W., Mo, F., Brahmbhatt, S., Shukin, R., Le Bihan, S., Gleave, M.E., Nykter, M., Collins, C.C. & Chi, K.N. (2016) 'Genomic Alterations in Cell-Free DNA and Enzalutamide Resistance in Castration-Resistant Prostate Cancer', *JAMA Oncology*, 2(12), pp. 1598–1606.

Xia, B.-W., Zhao, S.-C., Chen, Z.-P., Chen, C., Liu, T.-S., Yang, F. & Yan, Y. (2021) 'Relationship between serum total testosterone and prostate volume in aging men', *Scientific Reports*, 11p. 14122.

Xia, L., Han, Q., Duan, X., Zhu, Y., Pan, J., Dong, B., Xia, W., Xue, W. & Sha, J. (2022) 'm6A-induced repression of SIAH1 facilitates alternative splicing of androgen receptor variant 7 by regulating CPSF1', *Molecular Therapy - Nucleic Acids*, 28pp. 219–230.

Xu, D., Zhan, Y., Qi, Y., Cao, B., Bai, S., Xu, W., Gambhir, S.S., Lee, P., Sartor, O., Flemington, E.K., Zhang, H., Hu, C.-D. & Dong, Y. (2015) 'Androgen receptor splice variants dimerize to transactivate target genes', *Cancer research*, 75(17), pp. 3663–3671.

Xu, J. & Zhang, Y. (2010) 'How significant is a protein structure similarity with TM-score = 0.5?', *Bioinformatics*, 26(7), pp. 889–895.

Xu, X.J., Su, J.G., Bizzarri, A.R., Cannistraro, S., Liu, M., Zeng, Y., Chen, W.Z. & Wang, C.X. (2013) 'Detection of persistent organic pollutants binding modes with androgen receptor ligand binding domain by docking and molecular dynamics', *BMC Structural Biology*, 13p. 16.

Xu, Y., Chen, S.-Y., Ross, K.N. & Balk, S.P. (2006) 'Androgens Induce Prostate Cancer Cell Proliferation through Mammalian Target of Rapamycin Activation and Post-transcriptional Increases in Cyclin D Proteins', *Cancer Research*, 66(15), pp. 7783–7792.

Xue, J., Ma, T. & Zhang, X. (2023) 'TRA2: The dominant power of alternative splicing in tumors', *Heliyon*, 9(4), p. e15516.

Yan, W.X., Chong, S., Zhang, H., Makarova, K.S., Koonin, E.V., Cheng, D.R. & Scott, D.A. (2018) 'Cas13d is a compact RNA-targeting type VI CRISPR effector positively modulated by a WYL domain-containing accessory protein', *Molecular cell*, 70(2), pp. 327-339.e5.

Yang, C.-C., Fazli, L., Loguercio, S., Zharkikh, I., Aza-Blanc, P., Gleave, M.E. & Wolf, D.A. (2015) 'Downregulation of c-SRC kinase CSK promotes castration resistant prostate cancer and pinpoints a novel disease subclass', *Oncotarget*, 6(26), pp. 22060–22071.

Yang, L., Xie, S., Jamaluddin, M.S., Altuwaijri, S., Ni, J., Kim, E., Chen, Y.-T., Hu, Y.-C., Wang, L., Chuang, K.-H., Wu, C.-T. & Chang, C. (2005) 'Induction of Androgen Receptor Expression by Phosphatidylinositol 3-Kinase/Akt Downstream Substrate, FOXO3a, and Their Roles in Apoptosis of LNCaP Prostate Cancer Cells \*', *Journal of Biological Chemistry*, 280(39), pp. 33558–33565.

Yang, L.-Z., Wang, Y., Li, S.-Q., Yao, R.-W., Luan, P.-F., Wu, H., Carmichael, G.G. & Chen, L.-L. (2019) 'Dynamic Imaging of RNA in Living Cells by CRISPR-Cas13 Systems', *Molecular Cell*, 76(6), pp. 981-997.e7.

Yang, Q., Fung, K.-M., Day, W.V., Kropp, B.P. & Lin, H.-K. (2005) 'Androgen receptor signaling is required for androgen-sensitive human prostate cancer cell proliferation and survival', *Cancer Cell International*, 5(1), p. 8.

Yang, Y.C., Banuelos, C.A., Mawji, N.R., Wang, J., Kato, M., Haile, S., McEwan, I.J., Plymate, S. & Sadar, M.D. (2016) 'Targeting androgen receptor activation function-1 with EPI to overcome resistance mechanisms in castration-resistant prostate cancer', *Clinical cancer research : an official journal of the American Association for Cancer Research*, 22(17), pp. 4466–4477.

Yang, Z., Wang, D., Johnson, J.K., Pascal, L.E., Takubo, K., Avula, R., Chakka, A.B., Zhou, J., Chen, W., Zhong, M., Song, Q., Ding, H., Wu, Z., Chandran, U.R., Maskrey, T.S., Nelson, J.B., Wipf, P. & Wang, Z. (2020) 'A Novel Small Molecule Targets Androgen Receptor and Its Splice Variants in Castration-Resistant Prostate Cancer', *Molecular Cancer Therapeutics*, 19(1), pp. 75–88.

Yeh, S., Lin, H.K., Kang, H.Y., Thin, T.H., Lin, M.F. & Chang, C. (1999) 'From HER2/Neu signal cascade to androgen receptor and its coactivators: a novel pathway by induction of androgen target genes through MAP kinase in prostate cancer cells', *Proceedings of the National Academy of Sciences of the United States of America*, 96(10), pp. 5458–5463.

Yeung, Y.-G., Nieves, E., Angeletti, R. & Stanley, E.R. (2008) 'Removal of detergents from protein digests for mass spectrometry analysis', *Analytical biochemistry*, 382(2), pp. 135–137.

Yheskel, M., Sidoli, S. & Secombe, J. (2023) 'Proximity labeling reveals a new in vivo network of interactors for the histone demethylase KDM5', *Epigenetics & Chromatin*, 16(1), p. 8.

Yi, W., Li, J., Zhu, X., Wang, X., Fan, L., Sun, W., Liao, L., Zhang, J., Li, X., Ye, J., Chen, F., Taipale, J., Chan, K.M., Zhang, L. & Yan, J. (2020) 'CRISPR-assisted detection of RNA–protein interactions in living cells', *Nature Methods*, 17(7), pp. 685–688.

Yoshida, T., Kinoshita, H., Segawa, T., Nakamura, E., Inoue, T., Shimizu, Y., Kamoto, T. & Ogawa, O. (2005) 'Antiandrogen Bicalutamide Promotes Tumor Growth in a Novel Androgen-Dependent Prostate Cancer Xenograft Model Derived from a Bicalutamide-Treated Patient', *Cancer Research*, 65(21), pp. 9611–9616.

Yoshimoto, R., Chhipi-Shrestha, J.K., Schneider-Poetsch, T., Furuno, M., Burroughs, A.M., Noma, S., Suzuki, H., Hayashizaki, Y., Mayeda, A., Nakagawa, S., Kaida, D., Iwasaki, S. & Yoshida, M. (2021) 'Spliceostatin A interaction with SF3B limits U1 snRNP availability and causes premature cleavage and polyadenylation', *Cell Chemical Biology*, 28(9), pp. 1356–1365.e4.

Yoshimoto, R., Kaida, D., Furuno, M., Burroughs, A.M., Noma, S., Suzuki, H., Kawamura, Y., Hayashizaki, Y., Mayeda, A. & Yoshida, M. (2017) 'Global analysis of pre-mRNA subcellular localization following splicing inhibition by spliceostatin A', *RNA*, 23(1), pp. 47–57.

Yu, S., Yeh, C.-R., Niu, Y., Chang, H.-C., Tsai, Y.-C., Moses, H.L., Shyr, C.-R., Chang, C. & Yeh, S. (2012) 'Altered prostate epithelial development in mice lacking the androgen receptor in stromal fibroblasts', *The Prostate*, 72(4), pp. 437–449.

Yu, S., Zhang, C., Lin, C.-C., Niu, Y., Lai, K.-P., Chang, H., Yeh, S.-D., Chang, C. & Yeh, S. (2011) 'Altered prostate epithelial development and IGF-1 signal in mice lacking the androgen receptor in stromal smooth muscle cells', *The Prostate*, 71(5), pp. 517–524.

Yu, Z., Chen, S., Sowalsky, A.G., Voznesensky, O.S., Mostaghel, E.A., Nelson, P.S., Cai, C. & Balk, S.P. (2014) 'Rapid Induction of Androgen Receptor Splice Variants by Androgen Deprivation in Prostate Cancer', *Clinical Cancer Research*, 20(6), pp. 1590–1600.

Zahler, A.M., Damgaard, C.K., Kjems, J. & Caputi, M. (2004) 'SC35 and heterogeneous nuclear ribonucleoprotein A/B proteins bind to a juxtaposed exonic splicing enhancer/exonic splicing silencer element to regulate HIV-1 tat exon 2 splicing', *The Journal of Biological Chemistry*, 279(11), pp. 10077–10084.

Zamagni, A., Cortesi, M., Zanoni, M. & Tesei, A. (2019) 'Non-nuclear AR Signaling in Prostate Cancer', *Frontiers in Chemistry*, 7p. 651.

Zerbe, L.K., Pino, I., Pio, R., Cosper, P.F., Dwyer-Nield, L.D., Meyer, A.M., Port, J.D., Montuenga, L.M. & Malkinson, A.M. (2004) 'Relative amounts of antagonistic splicing factors, hnRNP A1 and ASF/SF2, change during neoplastic lung growth: implications for pre-mRNA processing', *Molecular Carcinogenesis*, 41(4), pp. 187–196.

Zhang, C., Konermann, S., Brideau, N.J., Lotfy, P., Wu, X., Novick, S.J., Strutzenberg, T., Griffin, P.R., Hsu, P.D. & Lyumkis, D. (2018) 'Structural Basis for the RNA-Guided Ribonuclease Activity of CRISPR-Cas13d', *Cell*, 175(1), pp. 212-223.e17.

Zhang, D., Hu, Q., Liu, X., Ji, Y., Chao, H.-P., Liu, Y., Tracz, A., Kirk, J., Buonamici, S., Zhu, P., Wang, J., Liu, S. & Tang, D.G. (2020) 'Intron retention is a hallmark and spliceosome represents a therapeutic vulnerability in aggressive prostate cancer', *Nature Communications*, 11(1), p. 2089.

Zhang, J.D., Biczok, R. & Ruschhaupt, M. (2022) *ddCt: The ddCt Algorithm for the Analysis of Quantitative Real-Time PCR (qRT-PCR)*.

Zhang, Y., Qian, J., Gu, C. & Yang, Y. (2021) 'Alternative splicing and cancer: a systematic review', *Signal Transduction and Targeted Therapy*, 6(1), pp. 1–14.

Zhang, Y. & Skolnick, J. (2004) 'Scoring function for automated assessment of protein structure template quality', *Proteins*, 57(4), pp. 702–710.

Zhang, Ziheng, Sun, W., Shi, T., Lu, P., Zhuang, M. & Liu, J.-L. (2020) 'Capturing RNA-protein interaction via CRUIS', *Nucleic Acids Research*, 48(9), p. e52.

Zhang, Zeda, Zhou, C., Li, X., Barnes, S.D., Deng, S., Hoover, E., Chen, C.-C., Lee, Y.S., Zhang, Y., Wang, C., Metang, L.A., Wu, C., Tirado, C.R., Johnson, N.A., Wongvipat, J., Navrazhina, K., Cao, Z., Choi, D., Huang, C.-H., et al. (2020) 'Loss of CHD1 Promotes Heterogeneous Mechanisms of Resistance to AR-Targeted Therapy via Chromatin Dysregulation', *Cancer Cell*, 37(4), pp. 584–598.e11.

Zhao, J., Ning, S., Lou, W., Yang, J.C., Armstrong, C.M., Lombard, A.P., D'Abronzo, L.S., Evans, C.P., Gao, A.C. & Liu, C. (2020) 'Cross-resistance among next generation anti-androgen drugs through the AKR1C3/AR-V7 axis in advanced prostate cancer', *Molecular cancer therapeutics*, 19(8), pp. 1708–1718.

Zhao, X.Y., Boyle, B., Krishnan, A.V., Navone, N.M., Peehl, D.M. & Feldman, D. (1999) 'Two mutations identified in the androgen receptor of the new human prostate cancer cell line MDA PCa 2a', *The Journal of Urology*, 162(6), pp. 2192–2199.

Zheng, Y., Li, P., Huang, H., Ye, X., Chen, W., Xu, G. & Zhang, F. (2021) 'Androgen receptor regulates eIF5A2 expression and promotes prostate cancer metastasis via EMT', *Cell Death Discovery*, 7(1), pp. 1–8.

Zhou, H., Su, J., Hu, X., Zhou, C., Li, H., Chen, Z., Xiao, Q., Wang, B., Wu, W., Sun, Y., Zhou, Y., Tang, C., Liu, F., Wang, L., Feng, C., Liu, M., Li, S., Zhang, Y., Xu, H., et al. (2020) 'Glia-to-Neuron Conversion by CRISPR-CasRx Alleviates Symptoms of Neurological Disease in Mice', *Cell*, 181(3), pp. 590–603.e16.

Zhou, X. & Rokas, A. (2014) 'Prevention, diagnosis and treatment of high-throughput sequencing data pathologies', *Molecular Ecology*, 23(7), pp. 1679–1700.

Zhou, X., Wu, W., Li, H., Cheng, Y., Wei, N., Zong, J., Feng, X., Xie, Z., Chen, D., Manley, J.L., Wang, H. & Feng, Y. (2014) 'Transcriptome analysis of alternative splicing events regulated by SRSF10 reveals position-dependent splicing modulation', *Nucleic Acids Research*, 42(6), pp. 4019–4030.

Zhou, Y., Wang, G., Wang, P., Li, Z., Yue, T., Wang, J. & Zou, P. (2019) 'Expanding APEX2 Substrates for Proximity-Dependent Labeling of Nucleic Acids and Proteins in Living Cells', *Angewandte Chemie*, 131(34), pp. 11889–11893.

Zhou, Z., Corden, J.L. & Brown, T.R. (1997) 'Identification and Characterization of a Novel Androgen Response Element Composed of a Direct Repeat\*', *Journal of Biological Chemistry*, 272(13), pp. 8227–8235.

Zhou, Z., Flesken-Nikitin, A., Corney, D.C., Wang, W., Goodrich, D.W., Roy-Burman, P. & Nikitin, A.Yu. (2006) 'Synergy of p53 and Rb Deficiency in a Conditional Mouse Model for Metastatic Prostate Cancer', *Cancer Research*, 66(16), pp. 7889–7898.

Zhu, J., Mayeda, A. & Krainer, A.R. (2001) 'Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins', *Molecular Cell*, 8(6), pp. 1351–1361.

Ziada, A., Barqawi, A., Glode, L.M., Varella-Garcia, M., Crighton, F., Majeski, S., Rosenblum, M., Kane, M., Chen, L. & Crawford, E.D. (2004) 'The use of trastuzumab in the treatment of hormone refractory prostate cancer; phase II trial', *The Prostate*, 60(4), pp. 332–337.

Zou, J.X., Zhong, Z., Shi, X.-B., Tepper, C.G., deVere White, R.W., Kung, H.-J. & Chen, H. (2006) 'ACTR/AIB1/SRC-3 and androgen receptor control prostate cancer cell proliferation and tumor growth through direct control of cell cycle genes', *The Prostate*, 66(14), pp. 1474–1486.

Zuris, J.A., Thompson, D.B., Shu, Y., Guilinger, J.P., Bessen, J.L., Hu, J.H., Maeder, M.L., Joung, J.K., Chen, Z.-Y. & Liu, D.R. (2015) 'Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo', *Nature Biotechnology*, 33(1), pp. 73–80.