

**Generation of CRISPR engineered prostate cancer  
cell line models to study androgen receptor  
signalling in advanced prostate cancer**

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



**Evangelia Eirini Kounatidou**

**Northern Institute for Cancer Research**

**Faculty of Medical Sciences**

**Newcastle University**

**September 2019**



## Abstract

Prostate cancer resistance to AR targeted therapies due to the emergence of AR point mutations and AR splice variants that cannot be targeted by the currently available agents comprise a major clinical challenge.

There is paucity of models that accurately reflect the mechanisms of AR regulation in advanced disease. This highlights the high demand of generating novel disease relevant models. A CRISPR pipeline was developed to generate cell line models which harbour specific point mutations in the LBD of AR as well as stop codons in AR exon 5 which resulted in AR-FL knock-out so that the remaining endogenous AR-Vs could be studied discriminately of interfering AR-FL.

Using a streptavidin-tagged Cas9 in conjugation with a biotinylated donor template resulted in high donor template knock-in efficiencies and yielded (i) an AR<sub>W741L</sub> CWR22Rv1 cell line derivative and (ii) an AR-FL knock-out cell line derivative called CWR22Rv1-AR-EK (Exon Knock-out). CWR22Rv1-AR-EK cells retained all endogenous AR-Vs following AR gene editing. AR-Vs acted unhindered following AR-FL deletion to drive cell growth and expression of androgenic genes. Global transcriptomics demonstrated that AR-Vs drive expression of a cohort of cell cycle and DNA damage response genes and depletion of AR-Vs sensitised cells to ionising radiation.

To date, elimination of AR-Vs by pharmacological inhibition remains challenging. However, disruption of AR pre-mRNA splicing is nowadays a highly attractive option. A CRISPR-based approach, called CRIME (Cas9-directed Rapid Immunoprecipitation Mass Spectrometry of Endogenous proteins) was developed to isolate and identify the AR-V7-specific spliceosome, in association with a nuclease deficient Cas9 at AR cryptic exon 3. Mass spec-derived hits were screened for their ability to alter AR-V7 mRNA levels in CWR22Rv1 cells. SRSF3 was identified as a potential AR-V7 splicer which promotes cryptic exon 3 skipping in normal prostate. In CRPC, SRSF3 is significantly downregulated and hence cryptic exon 3 inclusion in the mature AR-V7 mRNA transcript is permitted leading to generation of AR-V7.

## Acknowledgements

First, I would like to thank my supervisor Dr Luke Gaughan for the opportunity he offered me to work in his group. Nothing would be possible without his constant guidance and support at difficult times. He is a true mentor and working alongside him was an incredible experience. I would also like to thank the members of the STTD group, especially, my co-supervisor Prof. Craig Robson, Dominic Jones, Lewis Chaytor, Laura Wilson and Alex Bainbridge for providing precious advice and help when needed.

Special thanks to Prof. Scott Dehm, University of Minnesota, and Dr Lakis Liloglou, University of Liverpool who have significantly assisted with my PhD project by providing valuable advice and resources to achieve my scientific goals.

Nobody can manage surviving a PhD without true friends. Therefore, I am eternally grateful to my lovely friends Katerina, Eirini, Argiro and Ricky who always had the patience to babysit and encourage me at times I thought I couldn't make it. They all stood by my side despite their busy schedules and I feel extremely lucky to have them in my life.

I cannot thank enough my incredible parents. They have both been by my side during this journey despite the thousands of miles that physically separate us. For this, and much more, I am beyond thankful.

*" No amount of experimentation can ever prove me right; a single experiment can prove me wrong."*

A. Einstein



## Contents

|   |    |
|---|----|
| Abstract.....   | i  |
| Acknowledgements.....   | ii |
| Abbreviations.....  | 1  |
| List of Figures.....  | 6  |
| List of Tables.....   | 9  |
| Chapter 1. Introduction.....  | 10 |
| 1.1. Anatomy and function of the prostate .....                                       | 10 |
| 1.2. Development of the prostate .....  | 10 |
| 1.3. Prostate carcinogenesis .....  | 11 |
| 1.3.1. Risk factors.....  | 11 |
| 1.3.2. Genomic landscape of PC.....   | 11 |
| 1.4. Prostate cancer incidence and mortality .....                                    | 13 |
| 1.5. The androgen receptor .....  | 14 |
| 1.5.1. The N-terminal domain.....   | 15 |
| 1.5.2. The DNA binding domain .....   | 16 |
| 1.5.3. The Hinge region .....   | 17 |
| 1.5.4. The Ligand binding domain.....   | 17 |
| 1.6. AR signalling .....  | 18 |
| 1.7. Clinical management of PC .....  | 19 |
| 1.8. Targeting the AR signalling pathway.....   | 20 |
| 1.8.1. Targeting androgen synthesis via the hypothalamic-pituitary-gonadal axis ..... | 20 |
| 1.8.2. Targeting adrenal androgen synthesis.....                                      | 20 |
| 1.8.3. Anti-androgen therapy.....   | 21 |
| 1.9. Mechanisms of PC Resistance to Current AR-targeted Therapies.....                | 23 |
| 1.9.1. AR Amplification.....  | 25 |
| 1.9.2. Nuclear Receptor Cross-talk .....  | 25 |
| 1.9.3. AR co-regulator dysregulation .....  | 26 |
| 1.9.4. Gain-of-function mutations within the LBD of AR .....                          | 27 |
| 1.10. AR splice variants .....  | 30 |
| 1.10.1. Origin of AR splice variants .....  | 30 |
| 1.10.2. Generation of AR splice variants.....   | 31 |
| 1.11. CRISPR genome editing.....  | 34 |

|                                       |   |    |
|---------------------------------------|---|----|
| 1.11.1.                               | Origin of CRISPR and its application in biomedicine .....     | 34 |
| 1.11.2.                               | Mechanism of CRISPR editing .....                             | 35 |
| 1.12.                                 | Expansion of the CRISPR toolbox .....                         | 37 |
| 1.13.                                 | Mechanism of splicing .....                                   | 39 |
| 1.14.                                 | Alternative splicing.....                                     | 40 |
| 1.15.                                 | Co-transcriptional splicing .....                             | 41 |
| 1.16.                                 | Alternative splicing in cancer .....                          | 43 |
| 1.17.                                 | Splicing in PC .....  | 44 |
| 1.18.                                 | Therapeutic targeting of splicing .....                       | 46 |
| Chapter 2. Aims and Objectives.....   |   | 48 |
| Chapter 3. Materials and Methods..... |   | 49 |
| 3.1.                                  | Mammalian cell culture - Cell maintenance and passaging ..... | 50 |
| 3.2.                                  | Cell lines .....  | 50 |
| 3.3.                                  | Compounds .....   | 51 |
| 3.4.                                  | Western Blotting (WB) .....                                   | 52 |
| 3.5.                                  | Agarose gel DNA electrophoresis .....                         | 54 |
| 3.6.                                  | Immunofluorescence .....                                      | 54 |
| 3.7.                                  | Gene expression analysis .....                                | 55 |
| 3.7.1.                                | RNA extraction .....  | 55 |
| 3.7.2.                                | Reverse transcription (RT).....                               | 55 |
| 3.7.3.                                | Quantitative real-time PCR (qPCR) .....                       | 56 |
| 3.8.                                  | Plasmids .....  | 58 |
| 3.9.                                  | Bacterial transformation.....                                 | 59 |
| 3.10.                                 | Plasmid extraction .....                                      | 59 |
| 3.11.                                 | Plasmid transfections.....                                    | 60 |
| 3.12.                                 | siRNA transfections.....                                      | 60 |
| 3.13.                                 | Immunoprecipitation (IP).....                                 | 61 |
| 3.14.                                 | Chromatin Immunoprecipitation (ChIP) .....                    | 61 |
| 3.14.1.                               | Fixation.....   | 61 |
| 3.14.2.                               | Cell lysis and chromatin fractionation .....                  | 62 |
| 3.14.3.                               | Chromatin Immunoprecipitation .....                           | 62 |
| 3.14.4.                               | Elution and reverse cross-linking .....                       | 62 |
| 3.14.5.                               | Chromatin extraction .....                                    | 63 |

|   |   |     |
|---|---|-----|
| 3.15.   | Cell proliferation assays .....   | 64  |
| 3.15.1.   | Sulforhodamine B (SRB) assay .....  | 64  |
| 3.15.2.   | Colony formation assay.....   | 64  |
| Chapter 4. CRISPR knock-in pipeline for novel PC cell line development..... |   | 64  |
| 4.1.  | Introduction .....  | 65  |
| 4.2.  | Specific Materials & Methods.....   | 70  |
| 4.2.1.  | CRISPR complex components: CRISPR vectors – synthetic gRNAs .....   | 70  |
| 4.2.2.  | Construction of pSpCas9-mSA-2A-Puro.....  | 70  |
| 4.2.3.  | Cell transfection/nucleofection .....   | 71  |
| 4.2.4.  | Mutation detection assay (TIDE and SURVEYOR assay).....   | 72  |
| 4.2.5.  | ssODN design .....  | 73  |
| 4.2.6.  | Flow cytometry and cell sorting.....  | 74  |
| 4.2.7.  | HDR detection .....   | 75  |
| 4.2.8.  | Colony screening – RFLP/Sanger sequencing .....   | 75  |
| 4.2.9.  | CRISPR off-target analysis .....  | 75  |
| 4.2.10.   | Subcellular Fractionation and WB.....   | 75  |
| 4.2.11.   | siRNA knockdown .....   | 76  |
| 4.3.  | Results.....  | 77  |
| 4.3.1.  | The sgRNA_2/Cas9 complex demonstrated high editing efficiency in HEK293 and CWR22Rv1 cells.....   | 78  |
| 4.3.2.  | gRNA_2 efficiently guides Cas9 to target chromatin sites.....   | 83  |
| 4.3.3.  | sgRNA_2/Cas9 and ssODN delivery in CWR22Rv1 cells to generate the clinically relevant AR <sup>W741L</sup> mutation via HDR .....                          | 84  |
| 4.3.4.  | Successful HDR events following CRISPR induced DSBs were observed in CWR22Rv1 cells   | 86  |
| 4.3.5.  | Optimisation steps to enhance genome editing via HDR-mediated ssODN knock-in .....  | 88  |
| 4.3.6.  | Mutation-to-cut-site distance dictates ssODN knock-in fate .....  | 91  |
| 4.3.7.  | Depletion of key components of the NHEJ pathway does not enhance HDR .....  | 93  |
| 4.3.8.  | Reducing the mutation-to-cut-site distance in concert with biotinylation of the donor template and covalent tagging of Cas9 with mSA facilitate HDR ..... | 95  |
| 4.3.9.  | Generation of an in-house mSA-tagged Cas9 expressing vector.....  | 99  |
| 4.3.10.   | Validation of successful Cas9-mSA expression .....  | 103 |
| 4.3.11.   | Conjugation of mSA-tagged Cas9 with a biotinylated donor template yielded elevated template knock-in rates.....   | 104 |
| 4.4.  | Discussion.....   | 109 |

|   |     |
|---|-----|
| Chapter 5. Generation of an AR-FL knock out CWR22Rv1 cell line derivative using CRISPR.....                           | 114 |
| 5.1. Introduction .....   | 114 |
| 5.2. Specific Materials & Methods .....   | 118 |
| 5.2.1. Generation of the CWR22Rv1-AR-EK cell line using CRISPR .....  | 118 |
| 5.2.2. CRISPR off-target analysis .....   | 118 |
| 1.5. RNA sequencing and transcriptomic analysis .....   | 119 |
| 5.2.3. Data Availability .....  | 120 |
| 5.2.4. Cell proliferation and Colony Formation assays .....   | 120 |
| 5.2.5. Statistics .....   | 120 |
| 5.3. Results.....   | 121 |
| 5.3.1. CWR22Rv1 cells were genetically modified to knock-out AR-FL .....  | 121 |
| 5.3.2. Identification of the genomic alterations caused in the AR gene upon editing by Cas9                           | 123 |
| 5.3.3. CRISPR off-target analysis shows no promiscuous activity of Cas9 .....   | 125 |
| 5.3.4. AR-FL is no longer expressed in CWR22Rv1-AR-EK cells upon CRISPR editing .....                                 | 126 |
| 5.3.5. AR-V expression remains intact in AR-EK cells.....   | 128 |
| 5.3.6. AR-Vs maintain expression of AR target genes in the absence of AR-FL.....                                      | 129 |
| 5.3.7. Transcriptomic analysis of CWR22Rv1-AR-EK cells.....   | 133 |
| 5.3.8. AR-Vs are cell cycle and DNA damage response (DDR) regulators.....   | 137 |
| 5.3.9. AR-Vs regulate cell cycle, cell proliferation and cell fat.....  | 138 |
| 5.3.10. AR-Vs are master regulators of DDR.....   | 140 |
| 5.3.11. AR-Vs attenuate sensitisation to ionising radiation.....  | 146 |
| 5.3.12. AR-Vs may act as transcriptional repressors of tumour suppressor genes .....                                  | 149 |
| 5.4. Discussion.....  | 150 |
| Chapter 6. Repurposing CRISPR to identify AR-V7 specific splicing factors.....  | 156 |
| 6.1. Introduction .....   | 154 |
| 6.2. Specific Materials and Methods .....   | 158 |
| 6.2.1. Construction of gRNA expressing plentiCRISPR-iCer vectors.....   | 158 |
| 6.2.2. CRIME, Silver staining and LC-MS .....   | 158 |
| 6.2.3. Flow Cytometry .....   | 159 |
| 6.2.4. siRNA design.....  | 159 |
| 6.3. Results.....   | 161 |
| 6.3.1. CWR22Rv1 cell line suitability to apply CRIME .....  | 161 |
| 6.3.2. Successful RNA-guided recruitment of dCas9 to the AR cryptic exon 3 chromatin region<br>in CWR22Rv1 cells..... | 161 |

|   |  |     |
|---|--|-----|
| 6.3.3.  | The nuclease-deficient Cas9 blocks transcription in a strand-specific manner .....               | 166 |
| 6.3.4.  | Immunoprecipitation of FLAG-dCas9 yielded sufficient material for CRIME analysis ....            | 168 |
| 6.3.5.  | dCas9 was successfully detected by LC-MS/MS .....  | 170 |
| 6.3.6.  | RNA-associated proteins were enriched using CRIME.....   | 171 |
| 6.3.7.  | Splicing factor hit validation.....  | 172 |
| 6.3.8.  | SRSF3 modulates AR-V7 mRNA levels.....   | 173 |
| 6.3.9.  | SRSF3 expression is downregulated in mCRPC and negatively correlates with AR<br>expression ..... | 174 |
| 6.3.10.   | SRSF3 is a potential tumour suppressor in CRPC.....  | 176 |
| 6.4.  | Discussion.....  | 178 |
| Chapter 7. Key conclusions and future work..... |  | 180 |
| Appendices.....                                 |  | 186 |
| References.....                                 |  | 202 |

## Abbreviations

|        |   |
|--------|---|
| ADT    | Androgen deprivation therapy                              |
| AF     | Activation function                                       |
| AR     | Androgen receptor   |
| ARE    | Androgen response element                                 |
| AR-EK  | Androgen receptor exon knock-out                          |
| AR-FL  | Androgen receptor full-length                             |
| AR-V   | Androgen receptor variant                                 |
| ASO    | Antisense short oligonucleotide                           |
| ATCC   | American Type Culture Collection                          |
| bp     | base pair   |
| BPH    | Benign prostate hyperplasia                               |
| Cas9   | CRISPR associated protein 9                               |
| CE     | Cryptic exon  |
| ChIP   | Chromatin immunoprecipitation                             |
| CPA    | Cyproterone acetate                                       |
| CRISPR | Clustered regularly interspaced short palindromic repeats |
| CTC    | Circulating tumour cell                                   |
| CTD    | Carboxy-terminal domain                                   |
| ctDNA  | circulating tumour-derived DNA                            |
| DBD    | DNA binding domain  |
| dCas9  | dead Cas9   |
| DCC    | Dextran-coated charcoal                                   |

|          |                                      |
|----------|--------------------------------------|
| DDR      | DNA damage response                  |
| DHEA     | dehydroepiandrosterone               |
| DHT      | Dihydrotestosterone                  |
| DMSO     | Dimethyl sulfoxide                   |
| DNase    | Deoxyribonuclease                    |
| DNA-PKcs | DNA-protein kinase catalytic subunit |
| dNTP     | Nucleoside triphosphate              |
| DSB      | Double strand break                  |
| dsODN    | double stranded oligo DNA nucleotide |
| EMT      | Epithelial-to-mesenchymal transition |
| ESE      | Exonic splicing enhancer             |
| ESS      | Exonic splicing silencer             |
| FDA      | U.S. Food and Drug Administration    |
| FDR      | False discovery rate                 |
| GFP      | Green fluorescent protein            |
| GnRH     | Gonadotropin-releasing hormone       |
| GR       | Glucocorticoid receptor              |
| gRNA     | guide RNA                            |
| HDR      | Homologous directed repair           |
| HRP      | Horseradish peroxidase               |
| HSP      | Heat shock protein                   |
| IF       | Immunofluorescence                   |
| Indels   | Insertions/Deletions                 |
| IP       | Immunoprecipitation                  |

|       |   |
|-------|---|
| IR    | Ionising radiation                            |
| ISE   | Intronic splicing enhancer                    |
| ISS   | Intronic splicing silencer                    |
| Kb    | Kilobase                                      |
| KDa   | Kilodalton                                    |
| LBD   | Ligand binding domain                         |
| LH    | Leutenising hormone                           |
| LHRH  | Luteinizing hormone-releasing hormone         |
| mCRPC | Metastatic castrate-resistant prostate cancer |
| mM    | millimollar                                   |
| MS    | Mass spectrometry                             |
| mSA   | monoStreptavidin                              |
| ng    | nanogram                                      |
| NGS   | Next-generation sequencing                    |
| NHEJ  | Non-homologous end joining                    |
| NLS   | Nuclear localisation signal                   |
| nM    | nanomollar                                    |
| NMR   | Nuclear magnetic resonance                    |
| NR    | Nuclear receptor                              |
| NTC   | No template control                           |
| NTD   | N-terminal domain                             |
| ORF   | Open reading frame                            |
| PAM   | Protospacer adjacent motif                    |

|            |  |
|------------|--|
| PARP       | Poly (ADP-ribose) polymerase                               |
| PC         | Prostate cancer  |
| PCA        | Principal component analysis                               |
| PDX        | Patient-derived xenografts                                 |
| PR         | Progesterone receptor                                      |
| Pre-mRNA   | Precursor messenger ribonucleic acid                       |
| PSA        | Prostate specific antigen                                  |
| qPCR       | quantitative polymerase chain reaction                     |
| RFP        | Red fluorescent protein                                    |
| RIPA       | Radioimmunoprecipitation assay buffer                      |
| RFLP       | Restriction fragment length polymorphism                   |
| RNA Pol II | RNA polymerase II  |
| RT         | Reverse transcription/ Room temperature/ Radiation therapy |
| SEM        | Standard error of the mean                                 |
| SF         | Splicing factor  |
| siRNA      | Small interfering RNA                                      |
| SNP        | Single nucleotide polymorphism                             |
| snRNP      | Small nuclear ribonucleoprotein                            |
| SSA        | Spliceostatin A  |
| SRB        | Sulforhodamine B   |
| SRSF       | Serine-Arginine splicing factor                            |
| SSO        | Splice switching oligonucleotide                           |
| ssODN      | single stranded oligo DNA nucleotide                       |
| Tau        | Transcription activation unit                              |

|                |                                     |
|----------------|-------------------------------------|
| TFIIF          | Transcription factor II F           |
| TIDE           | Tracking of indels by decomposition |
| T <sub>m</sub> | Melting temperature                 |
| TSS            | Transcription start site            |
| UTR            | Untranslated region                 |
| WB             | Western blot                        |
| μl             | Microlitre                          |

## List of Figures

|   |    |
|---|----|
| Figure 1. Anatomy of the human prostate. ....   | 10 |
| Figure 2. Representative genetic alterations associated with prostate cancer progression. 13  | 13 |
| Figure 3. Prostate cancer mortality rates. ....   | 14 |
| Figure 4. Schematic representation of the structure and organization of AR at DNA, mRNA and protein level. ....   | 15 |
| Figure 5. Schematic illustration of the canonical AR signalling pathway in prostate cells. ....   | 19 |
| Figure 6. Current prostate cancer treatment options. ....   | 23 |
| Figure 7. AR signalling persists in advanced prostate cancer leading to disease progression and death. ....   | 24 |
| Figure 8. AR promiscuity to alternative ligands. ....   | 28 |
| Figure 9. Spectrum of AR splice variants (AR-Vs) emerging in response to ADT. ....  | 34 |
| Figure 10. CRISPR complex formation at target locus. ....   | 36 |
| Figure 11. Repair signalling pathways activated upon DSB induction by Cas9. ....  | 37 |
| Figure 12. Expansion of the CRISPR toolbox. ....  | 38 |
| Figure 13. Schematic of spliceosome assembly. ....  | 40 |
| Figure 14. Modes of alternative splicing. ....  | 41 |
| Figure 15. 'Miller spread' electron micrograph of a <i>D. melanogaster</i> gene and its graphical representation (shown on left and right, respectively). ....                              | 42 |
| Figure 16. Schematic of co-transcriptional splicing. ....   | 43 |
| Figure 17. Map of the all-in-one pLV-U6g-EPCG CRISPR vector used to ectopically express the gRNA and Cas9 nuclease in the host cell lines. 78   | 78 |
| Figure 18. In silico analysis of AR exon 5 and surrounding sequences using the CCTop CRISPR/Cas9 target online predictor. ....  | 78 |
| Figure 19. Outline of the SURVEYOR assay used for the mutational analysis of the CRISPR engineered cell lines. ....   | 79 |
| Figure 20. Chromatograms of sequenced AR exon 5 upon editing by Cas9 in the CWR22Rv1 and HEK293 cell lines. ....  | 81 |
| Figure 21. CRISPR editing efficiency before and after puromycin selection in CWR22Rv1 cells. ....   | 82 |
| Figure 22. Bar charts of the editing efficiencies of gRNA_1/Cas9 and gRNA_2/Cas9 in CWR22Rv1 cells. ....  | 83 |
| Figure 23. Cas9-NLS successfully translocates to the nucleus, where it is observed in foci. ...   | 84 |
| Figure 24. Schematic outline of the strategy designed to CRISPR engineer and generate the CWR22Rv1 AR <sup>W741L</sup> cell line. ....  | 85 |
| Figure 25. Detection of HDR events induced in CWR22Rv1 cells upon Cas9-mediated DSB. .  | 88 |
| Figure 26. Schematic of the alternative strategies applied for the CRISPR-mediated generation of the CWR22Rv1 AR <sup>W741L</sup> cell line. ....   | 90 |
| Figure 27. Assessment of the HDR events and ssODN knock-in efficiencies in the mixed cell populations obtained from all three experimental strategies outlined in Figures 24A and 26A. .... | 91 |
| Figure 28. Spectrum of CRISPR editing events in CWR22Rv1 clonal cell populations. ....  | 92 |
| Figure 29. ....   | 93 |

|   |     |
|---|-----|
| Figure 30. Inhibition of NHEJ does not enhance HDR.....   | 95  |
| Figure 31. Partial sequence of the ssODN_3 designed to improve ARW741L knock-in rates.  | 98  |
| Figure 32. Summary of the cloning strategy applied to generate the in-house px459 Cas9-mSA expressing construct.....                        | 102 |
| Figure 33. Impact of the mSA fusion on Cas9 expression and function in HEK293 cells.....  | 104 |
| Figure 34. Detection of HDR events in CWR22Rv1 cells using a biotinylated donor template and mSA-tagged Cas9. ....                          | 106 |
| Figure 35. HEK293 and CWR22Rv1 mixed cell populations were screened to identify HDR mediated donor template knock-in events. ....           | 107 |
| Figure 36. Identification of ARW741L positive CWR22Rv1 single cell clones. ....   | 108 |
| Figure 37. CRISPR knock-in strategy to introduce a translational stop codon in AR exon 5 to abrogate AR-FL expression. ....                 | 118 |
| Figure 38. Loss of full-length AR expression in C34 and C36 clonal populations. ....  | 122 |
| Figure 39. Diagrammatic representation of the DNA repair events in clone C34 upon Cas9-induced DNA DSB.....                                 | 124 |
| Figure 40. C34 demonstrates similar phenotype to parental CWR22Rv1 cells.....   | 125 |
| Figure 41. Sequence analysis of the top 5 predicted CRISPR off-target genes demonstrate no off-target CRISPR activity.....                  | 126 |
| Figure 42. Validation of AR-FL gene knock-out in CWR22Rv1-AR-EK cells.....  | 128 |
| Figure 43. AR-V levels were maintained in CRISPR edited AR-EK cells.....  | 129 |
| Figure 44. AR-Vs maintain transactivation of canonical AR target genes in CWR22Rv1-AR-EK cells. ....  | 130 |
| Figure 45. Androgenic gene expression was maintained by AR-Vs in CWR22Rv1-AR-EK cells. ....   | 131 |
| Figure 46. Occupancy of <i>cis</i> -regulatory elements of canonical AR target genes by AR-Vs. ...  | 133 |
| Figure 47. Preparation of RNA sequencing samples to define the AR-V transcriptome.....  | 134 |
| Figure 48. Confirming AR-V depletion in siARex1 transfected CWR22Rv1-AR-EK cells. ....  | 135 |
| Figure 49. Deletion of AR-Vs in CWR22Rv1-AR-EK cells impacts global gene expression.....  | 136 |
| Figure 50. Bar chart of the most significantly altered signalling pathways upon AR-V depletion in CWR22Rv1-AR-EK cells. ....                | 137 |
| Figure 51. Bar chart of the most significantly downregulated cell cycle related processes upon AR-V knockdown in CWR22Rv1 AR-EK cells. .... | 138 |
| Figure 52. AR-Vs dictate the proliferation potential of CWR22Rv1 parental and AR-EK cells. ....   | 140 |
| Figure 53. Bar chart of the most significantly downregulated DDR related processes upon AR-V depletion in CWR22Rv1 AR-EK cells.....         | 141 |
| Figure 54. AR-Vs drive expression of a DNA damage response gene signature.....  | 142 |
| Figure 55. Spectrum of DDR signalling pathways activated by AR-Vs in CWR22Rv1 AR-EK cells. ....   | 143 |
| Figure 56. Expression of AR-V regulated DDR genes is elevated in mCRPC and positively correlates with AR-V7 mRNA expression.....            | 146 |
| Figure 57. AR-V depletion sensitises CWR22Rv1-AR-EK cells to IR. ....   | 147 |
| Figure 58. Loss of AR-V expression associates with DNA repair deficiency in CWR22Rv1-AR-EK cells. ....                                      | 148 |

|   |     |
|---|-----|
| Figure 59. Functional analysis of the most up-regulated pathways upon AR-V depletion in CWR22Rv1 cells.....                                 | 150 |
| Figure 60. Schematic of the nuclease deficient Cas9 (dead Cas9, dCas9).....   | 156 |
| Figure 61. CRIME outline. ....  | 156 |
| Figure 62. AR-Vs constitute the most prevalent form of AR in CWR22Rv1 cells. ....   | 161 |
| Figure 63. dCas9 successfully migrates to the nucleus and is present in the chromatin fraction of CWR22Rv1 cells.....                       | 162 |
| Figure 64. dCas9 is successfully RNA-guided to the chromatin region of CE3 in PC3 cells. .  | 163 |
| Figure 65. CWR22Rv1 cells demonstrate relatively low transfection rates using conventional lipofection for dCas9/gRNA vector delivery. .... | 165 |
| Figure 66. dCas9 is successfully RNA-guided to the chromatin region of ce3 in CWR22Rv1 cells. ....  | 166 |
| Figure 67. RNA polymerase II dissociates from chromatin or engages with dCas9 to mediate transcription in a gRNA dependent fashion. ....    | 167 |
| Figure 68. Immunoprecipitation of dCas9 from formaldehyde fixed cells.....  | 169 |
| Figure 69. Validation of dCas9 and gRNA expression and recruitment to AR CE3 prior to MS. ....  | 170 |
| Figure 70. Pathway enrichment analysis of CRIME derived proteins.....   | 171 |
| Figure 71. CRIME hit validation.....  | 173 |
| Figure 72. SRSF3 modulates AR-V7 mRNA levels. ....  | 174 |
| Figure 73. SRSF3 expression is downregulated in CRPC and negatively correlates with AR expression. ....                                     | 175 |
| Figure 74. Schematic of the proposed model of SRSF3's role in AR-V7 pre-mRNA processing. ....   | 177 |

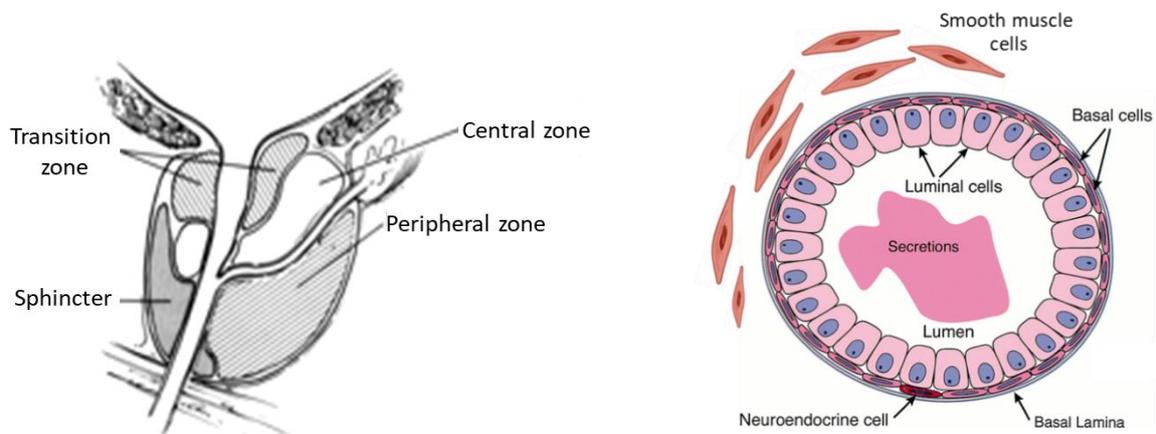
## List of Tables

|   |     |
|---|-----|
| Table 1. Ingredients of a 10% polyacrylamide gel. ....  | 52  |
| Table 2. Incubation conditions, working dilutions, exposure times and supplier information of antibodies used in WB. ....   | 53  |
| Table 3. Reverse transcription cocktail ingredients and volumes used per reaction.....  | 55  |
| Table 4. qPCR mastermix ingredients and volumes used per well/reaction.....   | 56  |
| Table 5. Sequences of primers used in qPCR.....   | 57  |
| Table 6. Control and AR-targeting siRNA sequences. ....   | 60  |
| Table 7. Primers used in qPCR to amplify chromatin regions of interest following ChIP.....  | 63  |
| Table 8. Sequences of gRNAs used to target AR exon 5.....   | 70  |
| Table 9. Primer sequences used for mSA PCR amplification. ....  | 71  |
| Table 10. Reagents used for mutation detection analysis using the SURVEYOR assay.....   | 72  |
| Table 11. Primer sequences used for AR exon 5 amplification. ....   | 73  |
| Table 12. Reagents used per PCR reaction. ....  | 73  |
| Table 13. Primer sequences used for PCR amplification of the knocked-in donor template in AR locus.....   | 75  |
| Table 14. siRNA sequences used to deplete components of the NHEJ pathway for precise genome editing. ....   | 76  |
| Table 15. Primers used in end-point PCR for CRISPR off-target analysis. ....  | 119 |
| Table 16. Primer sequences used to amplify (AR_1) and sequence (AR_2) the genomic region of AR intron 4 to intron 5. ....   | 119 |
| Table 17. Sequences of gRNAs used for CRIME. Part of the BsmBI recognition site is highlighted in red.....  | 158 |
| Table 18. Sequences of siRNAs targeting splicing factor hits identified using CRIME. ....   | 159 |
| Table 19. Sequences of primers used for siRNA knockdown validation. ....  | 160 |
| Table 20. Identification of dCas9 peptides by LC-MS. (I) – sum of raw spectrum intensities, rl – number of peptides detected, log(e) – expectation of detecting the protein stochastically, pI – isoelectric point of the intact gene product, Mr – mass of intact gene product. .... | 171 |

## Chapter 1. Introduction

### 1.1. Anatomy and function of the prostate

The human prostate is a male secondary sex organ which surrounds the urethra, located at the back of the bladder in males. The prostate is structurally made up of three distinct zones: the peripheral, transition and central (Figure 1A). Epithelial cells in each zone produce a glucose rich fluid with alkaline pH which provides an optimal environment for growth of spermatozoa. From a cell type stand-point, the prostate consists of two distinct layers (Figure 1B); the luminal layer which consists of secretory luminal cells, discontinued by neuroendocrine cells present in extremely low frequency and a layer of basal cells which is encapsulated by smooth muscle cells (Maitland, 2013).



**Figure 1. Anatomy of the human prostate. A.** Illustration of the human prostate. Taken from (McNeal, 1969). **B.** Cellular organisation within the prostate duct. Taken and adapted from (Abate-Shen, 2000).

### 1.2. Development of the prostate

Development of the prostate relies on a stream of sex hormones, such as the gonadotropin-releasing hormone (GnRH) and leutenising hormone (LH), which are produced by the hypothalamus and pituitary gland, respectively to stimulate testosterone production and release from the Leydig cells of the testes (Kluth et al., 2014). Upon entering cells of the prostate, testosterone is metabolised to its more potent derivative 5 $\alpha$ -dihydrotestosterone (DHT) and binds to the androgen receptor (AR), a nuclear transcription factor, which drives cell growth, differentiation and prostate homeostasis. Both genders express AR. However,

different levels of androgens between the two genders are responsible for differential female and male sex development. Disruption of this hormone cascade results in failure of prostate development and the affected individuals present a female-like urogenital morphology (Hiort, 2013; Randall, 1994).

### **1.3. Prostate carcinogenesis**

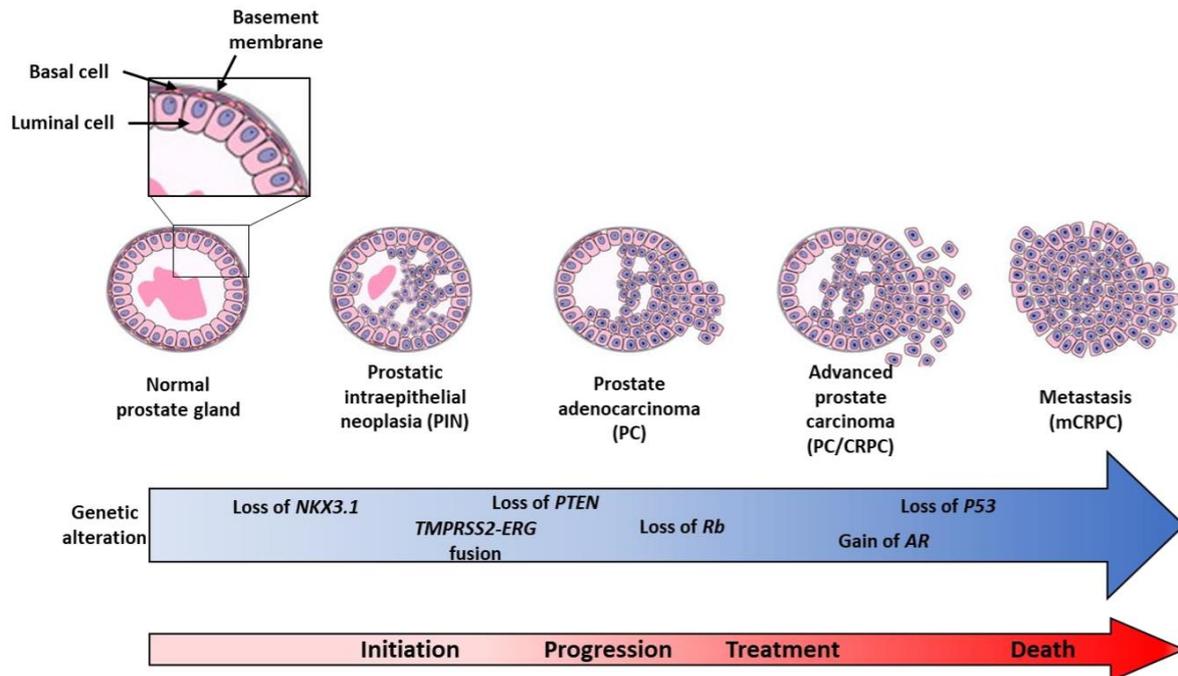
#### **1.3.1. Risk factors**

Prostate cancer emergence is associated with a limited number of risk factors which include age, family history and ethnicity. It is well established that PC incidence increases with age and older men are more likely to develop PC (Cancer Research UK, last updated 20/06/2019). Moreover, the risk of developing PC increases by up to 8 times if a man has a first-degree relative with PC. This is in part attributed to germline mutations observed in affected families. For instance, men under 65 with inherited *BRCA2* mutations demonstrate higher risk of developing PC compared to men with wild-type *BRCA2* due to DNA repair deficiency and increased genomic instability (Kote-Jarai et al., 2011). In addition, low-penetrance polymorphisms present in genes involved in androgen biosynthesis have been identified and these genetic variants are associated with increased susceptibility to PC. In particular, mutations in the *SRD5A2* gene lead to 5 $\alpha$ -reductase deficiency. One variant, Ala49Thr has been reported to enhance the catalytic activity of 5 $\alpha$ -reductase and is associated with an increased risk of advanced PC. This low-penetrance variant appears to increase the risk of PC in African-Americans by 7.2 fold. Finally, epidemiology studies have shown that African-American men are more likely to develop prostate cancer compared to Caucasian men and demonstrate significantly higher mortality rates suggesting differences in the demographic and economic backgrounds of these populations could be attributed to modulating PC risk (Powell and Bollig-Fischer, 2013). However, such populations (black vs white men) were also examined under equal socioeconomic criteria and surprisingly the same trend in PC incidence was observed, indicating that underlying genetic factors might be responsible for the observed discrepancy between the two groups (Ben-Shlomo et al., 2009).

#### **1.3.2. Genomic landscape of PC**

Prostate cancer aetiology is strongly linked to genetic factors such as somatic mutations or aberrations which are acquired rather than inherited in a man's lifetime. Next generation

sequencing (NGS) approaches have facilitated the identification of PC specific genomic alterations that occur at different stages during PC development and contribute to or drive malignant transformation events such as invasion and metastasis (Figure 2) (Baca and Garraway, 2012; Shen and Abate-Shen, 2010). Chromosomal translocations occur in the prostate which result in gene fusions and aberrant end-products and have been linked to early PC development. A typical example is that of the *TMPRSS2:ERG* gene fusion which is derived from the structural fusion of the 5' untranslated region (UTR) of *TMPRSS2* with *ERG*, a member of the ETS family of transcription factors. It is found in >50% of PC patients and critically the fusion product is androgen regulated, due to the presence of an androgen response element (ARE) in *TMPRSS2* which drives *ERG* overexpression, supporting the notion that AR signalling is required during early transformation processes (Cai et al., 2009; Holly et al., 2017). However, this structural rearrangement is not sufficient to promote transformation itself but needs to be accompanied by other oncogenic events such as *PTEN* loss followed by p53 inactivation which is required to overcome cellular senescence induced by *PTEN* loss (Chen et al., 2005). Inactivating deletions and mutations in the tumour suppressor gene *PTEN* occur frequently in PC and invariably lead to aberrant activation of oncogenic PI3K/Akt signalling which potentiate metabolic changes, including increased aerobic glycolysis, a common metabolic switch observed in cancer as described by the Warburg effect (Zhou et al., 2019). These early molecular changes induce transcriptional programmes which fuel malignant cell growth and tumourigenesis. Late events in PC development following these precursor lesions include *P53* mutations observed in 40% of PC patients as well as *RB* loss in ~30% of patients. Critically, aberrations in *AR* (discussed below) occur later in PC, comprise the most frequent molecular abnormalities observed in patients and more importantly they result from selection mechanisms activated in response to AR targeted treatments (Beltran et al., 2013).

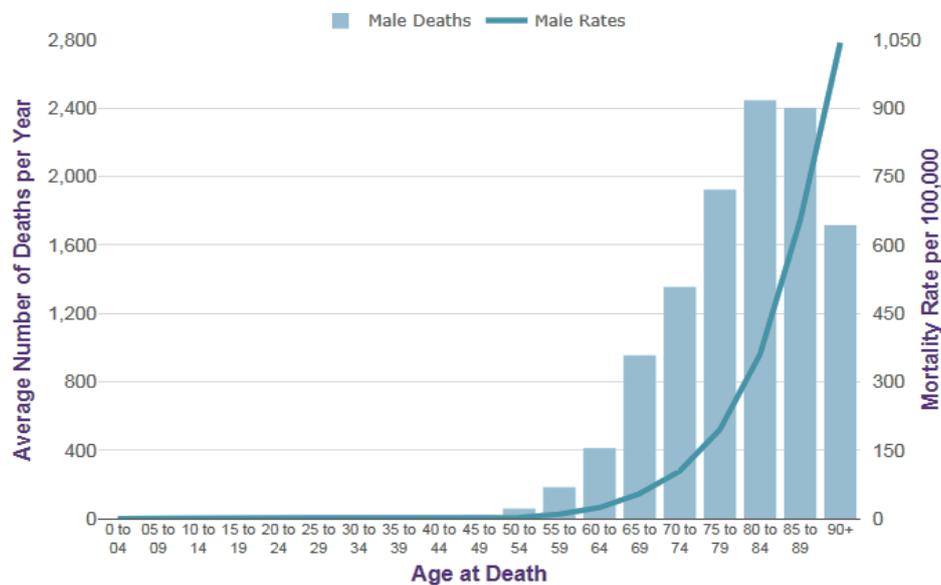


**Figure 2. Representative genetic alterations associated with prostate cancer progression.** Taken and adapted from Shen et al., 2010.

#### 1.4. Prostate cancer incidence and mortality

Prostate cancer (PC) is currently the most commonly diagnosed non-cutaneous cancer affecting UK men. Across the UK, nearly 50,000 new PC cases are recorded and over 11,500 men die of PC every year, whilst it is estimated that 1 in 8 UK men will get PC at some point in their lifetime (Cancer Research UK, 2016). Prognosis is good for men who present with localised disease of low grade. Nevertheless, patients with metastatic tumours show an extremely poor prognosis. To date, metastatic disease remains incurable and mortality rates

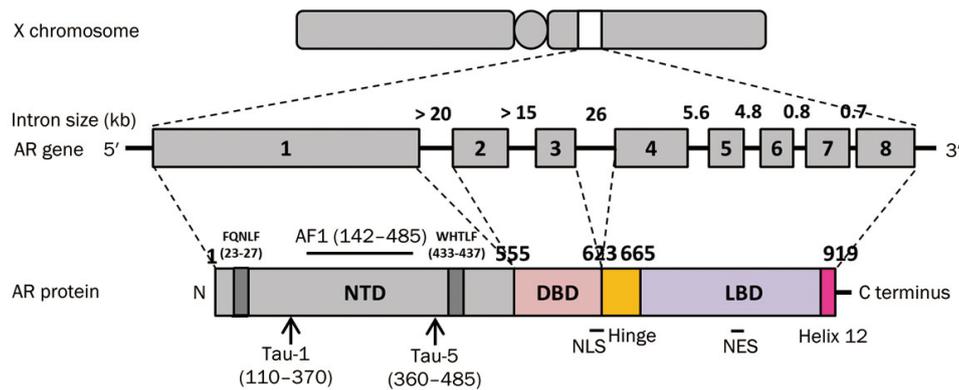
invariably increase each year (Figure 3) despite recent advances in PC therapeutics (Denmeade and Isaacs, 2002).



**Figure 3. Prostate cancer mortality rates.** Age-specific death rates in the UK between 2014 and 2016 (blue line). Average number of deaths recorded annually per age group (blue bars). Taken from Cancer Research UK, 2014-2016.

### 1.5. The androgen receptor

The prostate is dependent on androgens for growth and survival. Androgens bind to and activate the androgen receptor (AR) which in turn plays a vital role in prostate development by promoting transcription of essential genes for prostate function and homeostasis (Huggins and Hodges, 1941). The androgen receptor gene extends over a 90 kb chromatin region on chromosome X at q11-12 and is transcribed into a premature mRNA transcript, which consists of eight exons and encodes a 919 amino acid protein (UniProt: P10275). The AR belongs to the steroid-hormone nuclear receptor superfamily of transcription factors and consists of four functional domains: the N terminal domain (NTD), the DNA binding domain (DBD), a hinge region and the ligand binding domain (LBD) (Figure 4)(Tan et al., 2015).



**Figure 4. Schematic representation of the structure and organization of AR at DNA, mRNA and protein level.** AR consists of eight exons which give rise to four functionally distinct domains. Adapted from Tan et al., 2015.

### **1.5.1. The N-terminal domain**

The NTD of the AR is encoded exclusively by exon 1 and comprises the largest domain of the receptor, occupying 60% of its total length. Two different transcription activation units (Tau) have been mapped within the NTD, Tau1 (aa 101-370) and Tau5 (aa 360-528), which are together encompassed in AF-1 (Activation Function-1). Tau1 and Tau5 are responsible for mediating intramolecular interactions between the NTD and the LBD through AF-2 (Activation Function-2) to form a transcriptionally active receptor (Jenster et al., 1995).

The NTD is structurally disordered making crystallography or nuclear magnetic resonance (NMR) studies extremely difficult; hence our understanding of its structure is extremely limited. This loose three-dimensional conformation allows a multitude of interactions with co-regulators. Circular dichroism experiments coupled with fluorescence spectroscopy revealed that the general transcription factor TFIIF interacts with the AF-1 of AR and facilitates assembly of the transcription initiation complex at target gene promoters while it promotes mRNA transcription elongation by RNA polymerase II (Reid et al., 2002b). It has also been reported that binding of TFIIF to the NTD induced conformational changes which gave rise to an additional surface for further contacts with protein complexes, creating novel interaction networks. Importantly, AR NTD acquired a more stable conformation upon binding to TFIIF due to the formation of an  $\alpha$ -helix. Similarly, partial secondary structure of the NTD has been identified upon exposure to trifluoroethanol and trimethylamine N-oxide. The structure-stabilising properties of these solvents induce conformational changes which are consistent with an increase in  $\alpha$ -helical structure of the NTD (Reid et al., 2002a)

Early studies showed that the AR was able to interact with histone acetyl-transferases (HATs), steroid receptor coactivator-1 (SRC1) and transcriptional intermediary factor-2 (TIF2). These HATs are able to recruit general co-activators, CREB binding protein (CBP), p300 and p300/CBP-associating factor (PCAF), which also have HAT activity and together function to enhance AR transcriptional activity via direct acetylation of target lysine residues in the N-terminal tails of histones H3 and H4. This leads to chromatin relaxation and facilitates the recruitment of basal transcriptional machinery and additional chromatin remodelling complexes such as SWI/SNF (Oñate *et al.*, 1995, Powell *et al.*, 2004). Furthermore, the AR can be post-translationally modified at numerous residues throughout the NTD, with phosphorylation being by far the most well studied post-translational modification. For example, phosphorylation of serine-514 by MAPK has been associated with increased AR activity (Yeh *et al.*, 1999a). The AR NTD is also the target of SUMOylation. Lysine-385, within TAU-5, is an acceptor of small ubiquitin-like modifier (SUMO) peptides from PIAS1/2, the SUMO-1 ligase enzymes which SUMOylates the AR and represses its activity (Callewaert *et al.*, 2004) (Poukka *et al.*, 2000).

Specific sequence motifs which are important for structural rearrangements of the receptor in space also exist within the NTD. In particular, the FxxLF (<sup>23</sup>FQNL<sup>27</sup>) and WxxLF motifs (<sup>433</sup>WHTLF<sup>437</sup>) of the AR NTD associate with the LBD of the receptor following androgen stimulation permitting interaction between the N- and C-terminal domains to facilitate stabilisation of the receptor (Langley *et al.*, 1998). In fact, the FxxLF motif has been shown to make interactions with higher affinity to AF-2 of the LBD compared to the WxxLF motif, thus driving rather than just contributing to interdomain interactions of the AR (He *et al.*, 2000). This direct interaction is unlike other NRs such as the ER, which has an indirect N/C terminal interaction, utilising co-activators as bridging intermediates (Kobayashi *et al.*, 2000).

### **1.5.2. The DNA binding domain**

The DBD is encoded by exons 2 and 3 and consists of two zinc fingers and three  $\alpha$ -helices. The first zinc finger is located in exon 2 and contains the P-box <sup>577</sup>GSCKV<sup>581</sup> motif which forms an  $\alpha$ -helix and mediates direct contact with the DNA groove of specific AR-targeting sequences. This structural organisation is observed in other nuclear receptors (NRs) including the glucocorticoid receptor (GR) and progesterone receptor (PR) (Tsai and O'Malley, 1994). The second zinc finger is located in exon 3 and contains two  $\alpha$ -helices and the D-box <sup>596</sup>ASRND<sup>600</sup>

motif which is essential for receptor dimerisation via direct interaction with the D-box motif of a second AR monomer in a "head-to-head" fashion (Tan et al., 2015) at target binding sites.

The AR binds to specific DNA sequences termed androgen response elements (AREs). AREs consist of two hexameric half sites (5'-AGAACA-3') which are palindromic and separated by a 3 bp spacer (Mangelsdorf et al., 1995). Overlap in binding site recognition has been observed between NRs. The half site of ARE can also be recognised by GR but different binding affinities to AR have been observed. Unlike other NRs, AR can selectively bind to specific, non-conventional AREs throughout the genome which are generated by direct rather than inverted repeats of the consensus ARE and flanking DNA sequences (Haelens et al., 2001). This may be in part explained by the fact that in contrast to AR homodimers which acquire a "head-to-head" configuration, other NRs form "head-to-tail" dimers. Selectivity can finally be attributed to specific and unique residues within the D-box of AR such as a serine residue at position 438 which resides at the dimerization interface, stabilises homodimer formation and increases the binding affinity of the receptor to DNA by mediating electrostatic interactions and additional hydrogen bonds. In all, this allows AR to bind to a broader spectrum of DNA sites with higher affinity compared to other NRs (Shaffer et al., 2004).

### ***1.5.3. The Hinge region***

The hinge region of the AR is encoded by exon 4 and is a short amino acid sequence located between the DBD and the LBD. Despite its small size, the hinge region importantly contributes to translocation of AR to the nucleus by harbouring a bipartite nuclear localisation signal (NLS) (<sup>617</sup>**RKCYEAGMTLGARKLKKL**<sup>634</sup> – NLS in bold) in its N-terminal end (Clinckemalie et al., 2012). The importin- $\alpha$ -importin- $\beta$  complex recognises and binds to the KLKK motif in the hinge region and mediates transportation of AR and release into the nucleus (Cutress et al., 2008)The hinge region is also a hotspot for posttranslational modifications, including acetylation (Gaughan et al., 2002)and methylation (Gaughan et al., 2011), that regulate the transcriptional potency of the receptor.

### ***1.5.4. The Ligand binding domain***

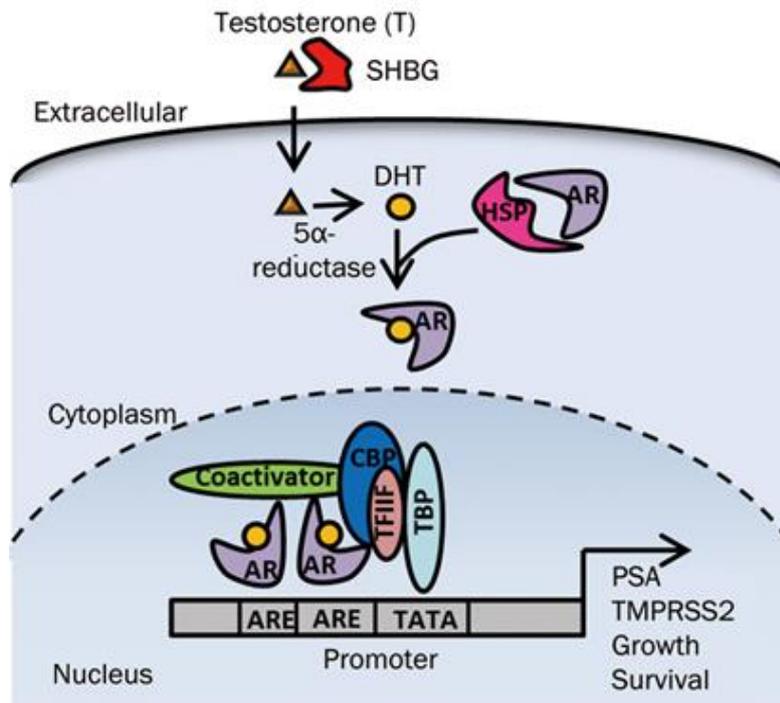
The LBD of AR is encoded by exons 4 to 8. It is structurally conserved between NRs despite differences in the primary amino acid sequence. It consists of 11  $\alpha$ -helices and 1  $\beta$ -sheet, as opposed to the LBD of all the other NRs which contain an additional  $\alpha$ -helix (helix 2). Helices

3, 5, 10 and 11 make up the ligand binding pocket. Remarkable conformational changes are induced upon androgen stimulation to facilitate ligand docking. Helix 12 works as a "lid" which relocates from its inactive to active position and seals the androgen occupied cleft of the receptor. This conformation is stabilised by the presence of the  $\beta$ -sheet (Gao et al., 2005). The LBD is essential for AR activation upon binding of androgens to a hydrophobic cleft formed by the AF-2 and helix 12 which is found at the C-terminal end of the LBD. Importantly, the LBD also provides the interface via its AF-2 for AR interaction with co-regulators that contain LxxLL motifs (termed NR boxes) which, similarly to the FxxLF motif of the AR NTD, form a helical structure to bind AF-2. However, these contacts are much weaker compared to the intradomain interactions between the AF-2 of the LBD and the W/FxxLF motifs of the NTD (He et al., 2002).

The AR LBD comprises a hotspot for point mutations (discussed in section 1.9.4) which alter the structure of the ligand binding pocket and subsequently broaden the ability of the receptor to accommodate novel ligands (Steketee et al., 2002).

### **1.6. AR signalling**

In the prostate, testosterone, the predominant androgen, enters the plasma membrane and is immediately reduced to its more potent derivative DHT by 5- $\alpha$ -reductase in the endoplasmic reticulum. Binding of DHT to AR occurs in the cytoplasm and results in dissociation of repressive heat-shock protein complexes and activation of the receptor by enabling inter-domain interactions between the N- and C-termini. The activated receptor, in turn, translocates into the nucleus, dimerises and binds to androgen response elements (AREs) on *cis*-regulatory sequences of target genes. The receptor then recruits a number of transcriptional machinery components to govern gene transcription (Figure 5) (Brinkmann et al., 1999).



**Figure 5. Schematic illustration of the canonical AR signalling pathway in prostate cells.** Stimulation with testosterone releases AR from heat-shock proteins in the cytosol and promotes translocation of the receptor into the nucleus, where AR regulates transcription of growth- and survival-associated genes. Taken from Tan et al., 2015.

### 1.7. Clinical management of PC

PC is classified as low-, intermediate- and high-risk depending on patient's TNM (Tumour, Node, Metastasis) stage, Gleason score and concentration of the PC biomarker prostate specific antigen (PSA) in the bloodstream.

Low risk PC is subject to active surveillance (AS) avoiding the need for utilising invasive treatment methods. AS showed high efficacy in ~98% of patients who presented a 10-year survival (Klotz et al., 2010). Intermediate-risk PC patients undergo surgical prostatectomy which has been shown to reduce PC-associated mortality. Combination of external beam radiation therapy (EBRT) with short-term androgen deprivation therapy (ADT) is an alternative option for these patients who present lower risk of biochemical relapse compared to ADT alone (Bill-Axelsson et al., 2011). Finally, high-risk PC patients with localised-advanced disease also undergo radical prostatectomy coupled with lymph node dissection for assessment of prognosis and disease stage (Moris et al., 2016). Radiation therapy (RT) in

combination with long-term ADT is strongly recommended for high-risk PC (Bolla et al., 2010) despite the observed side effects of the latter (Alibhai et al., 2006). It has been suggested that irregularly pausing ADT may relieve patients from ADT-associated side effects without limiting treatment efficacy (Klotz et al., 1986). However, more recent evidence indicates that RT synergy with short-term ADT does not benefit high-risk PC patients (Lawton et al., 2007).

## **1.8. Targeting the AR signalling pathway**

### **1.8.1. Targeting androgen synthesis via the hypothalamic-pituitary-gonadal axis**

Development of the prostate relies on androgens which regulate cell growth via the AR. Circulating androgen levels are reduced by targeting the hypothalamic-pituitary-gonadal axis and subsequently disrupting the release of precursor hormone molecules in the bloodstream to eventually reduce downstream testosterone production. Historically, LHRH agonists have been utilised to stimulate increased LHRH production by the anterior pituitary gland (Figure 6). Despite initial increase in LH, the levels of the hormone eventually (within two weeks) decrease due to a decrease in LHRH receptors and lead to reduced testosterone production. LHRH antagonists are also utilised in the clinic to block production of LHRH and disrupt testosterone production. For instance, Degarelix has been used in the clinic with no significant differences in efficacy compared to LH agonists. However, the compound causes a rapid drop in testosterone levels in contrast to LH agonists which stimulate a testosterone burst prior to its reduction, and therefore Degarelix and other LH antagonists are preferred over LH agonists. LHRH blockade has shown significant efficacy in the clinic with reduced tumour growth in over 70% of patients (Denmeade and Isaacs, 2002; Siddiqui and Krauss, 2018).

### **1.8.2. Targeting adrenal androgen synthesis**

Disruption of androgen production via the hypothalamic-pituitary-gonadal axis is not sufficient to eliminate androgen release in the circulation as there are other sources of androgens such as the adrenal glands. Adrenal androgen production originates from lipid biosynthesis. Abiraterone, an active metabolite of abiraterone acetate, prevents androgen synthesis by blocking enzymes involved in the *de novo* androgen biosynthesis cascade by adrenal glands (Figure 6). In particular, abiraterone is a specific CYP17A1 (cytochrome P450 17A1) enzyme inhibitor. In brief, CYP17A1 catalyses production of dehydroepiandrosterone (DHEA) and androstenedione, which are weak androgens but can be converted to

testosterone and DHT supplying the prostate with potent androgens required for cell growth and proliferation. Conceivably, abiraterone administration in patients with redundant adrenal androgen production can eliminate circulating androgens and further starve PC cells leading to tumour regression (Ferraldeschi et al., 2013).

Abiraterone has been successfully used in large clinical trials (COU-AA-301 and COU-AA-302). In both trials, patients with castration resistant PC (CRPC) previously treated with and without taxanes, respectively, received abiraterone and demonstrated significantly prolonged survival with median overall survivals of 15.8 and 34.7 months in the abiraterone treated groups compared to 11.2 and 30.3 months in the placebo control groups, respectively (Fizazi et al., 2012; Ryan et al., 2015). Currently, abiraterone is the preferred treatment option for patients with CRPC who have progressed on first-generation anti-androgens.

### ***1.8.3. Anti-androgen therapy***

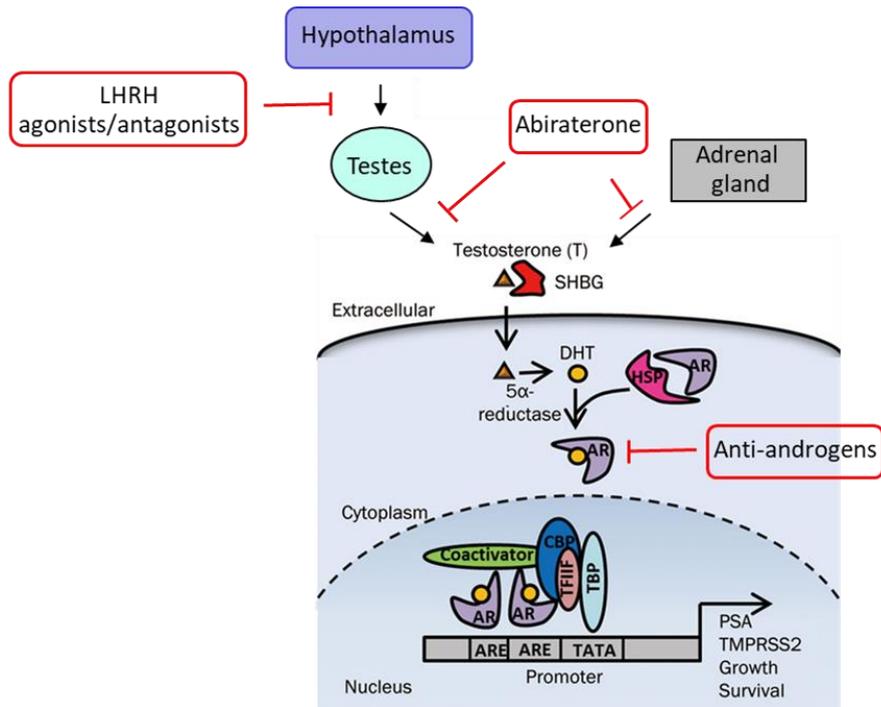
ADT has traditionally been used as the gold standard treatment for high-risk localized and advanced PC. The rationale of ADT relies on the early observation that androgens, such as testosterone are essential for prostate growth and survival (Huggins and Hodges, 1941). Thus, by eliminating testosterone supply to castration levels (~1.7 nmol/L of serum), apoptosis of the AR-positive prostate cancer cells is induced resulting in significant tumour regression (Maitland, 2015; Suzman and Antonarakis, 2015)

Despite the initial response to ADT, patients relapse with more aggressive tumours within 18 months. These tumours are no longer responsive to ADT resulting in castration-resistant prostate cancer (CRPC). In CRPC, AR signalling is inappropriately sustained through a number of different mechanisms (discussed in section 1.9) (Claessens et al., 2014). Researchers around the globe, have been trying to develop novel agents and strategies to target persistent AR in advanced CRPC (Sternberg et al., 2014).

First-generation anti-androgens, such as bicalutamide, have been used as first-line treatment for advanced PC. These agents normally compete with androgens for the ligand-binding site of the AR, ultimately resulting in blockade of the receptor (Figure 6). However, due to significantly lower affinity for AR and persistent tumour growth despite administration of the drugs, their efficacy has been limited, highlighting the need for novel therapies (Clegg et al., 2012).

Next-generation anti-androgens have been developed following the failure of their first-generation counterparts. Promising antitumor activity of enzalutamide (also known as MDV3100) has been reported in clinical trials. Enzalutamide targets AR at multiple levels in the signalling pathway. In particular, the agent initially binds to the LBD of AR with higher affinity compared to previously used anti-androgens, inhibits its translocation to the nucleus, blocks the recruitment of co-activators and the subsequent activation of AR target genes (Tran et al., 2009). In fact, men with advanced CRPC who received the drug demonstrated improved overall survival up to 4.8 months compared to individuals who received the placebo (Scher et al., 2010). At presentation, enzalutamide is the most potent and well-tolerated anti-androgen that has been approved by the FDA to be used as a first-line therapy for treating advanced CRPC (FDA, 2012).

Nonetheless, tumours develop resistance mechanisms to evade antitumour activity of enzalutamide and of other next-generation agents, such as abiraterone, resulting in palliative care being the only choice for patients with metastatic CRPC (Figure 7) (Chandrasekar et al., 2015).



**Figure 6. Current prostate cancer treatment options.** Agents targeting the AR signalling axis are shown in red boxes. Taken and adapted from Tan et al., 2015.

### 1.9. Mechanisms of PC Resistance to Current AR-targeted Therapies

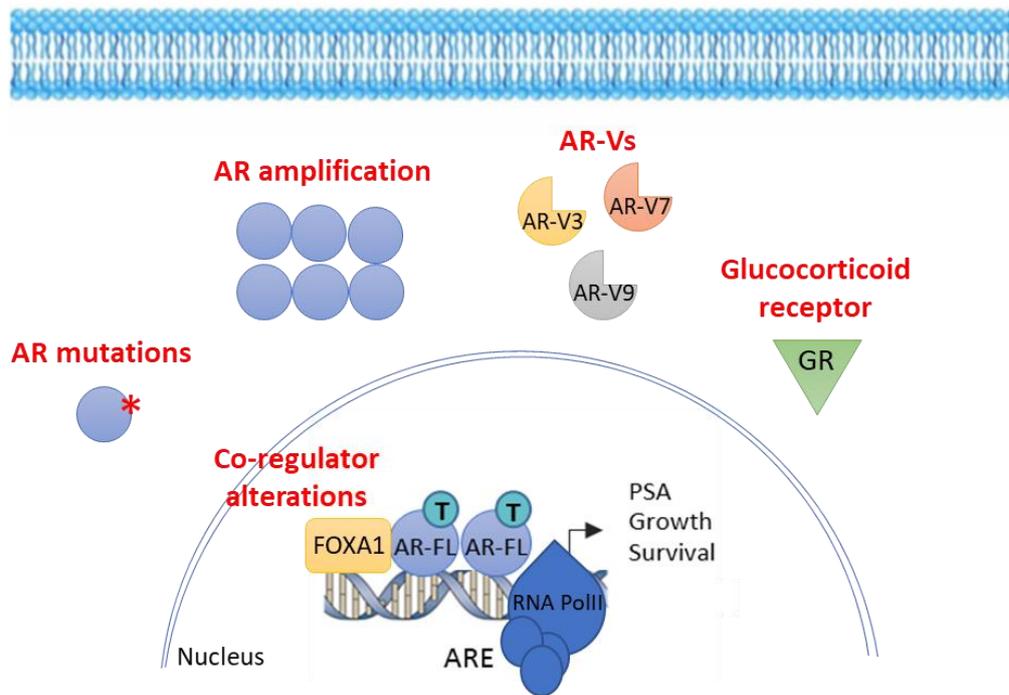
AR remains transcriptionally active in CRPC despite the development of potent agents with significantly improved pharmacological properties, such as enzalutamide. Increased levels of circulating PSA (an AR target gene product) detected in patient's serum upon castration are indicative of the inappropriately active AR in these patients.

Research has focused on deciphering the molecular mechanisms that sustain AR signalling in CRPC and confer resistance to current regimes leading ultimately to patient death (Figure 7B). Understanding the underlying mechanisms of resistance will enable precise and more efficient targeting of the AR, or of other components in the pathway, to provide men affected by PC with hope of more effective treatments.

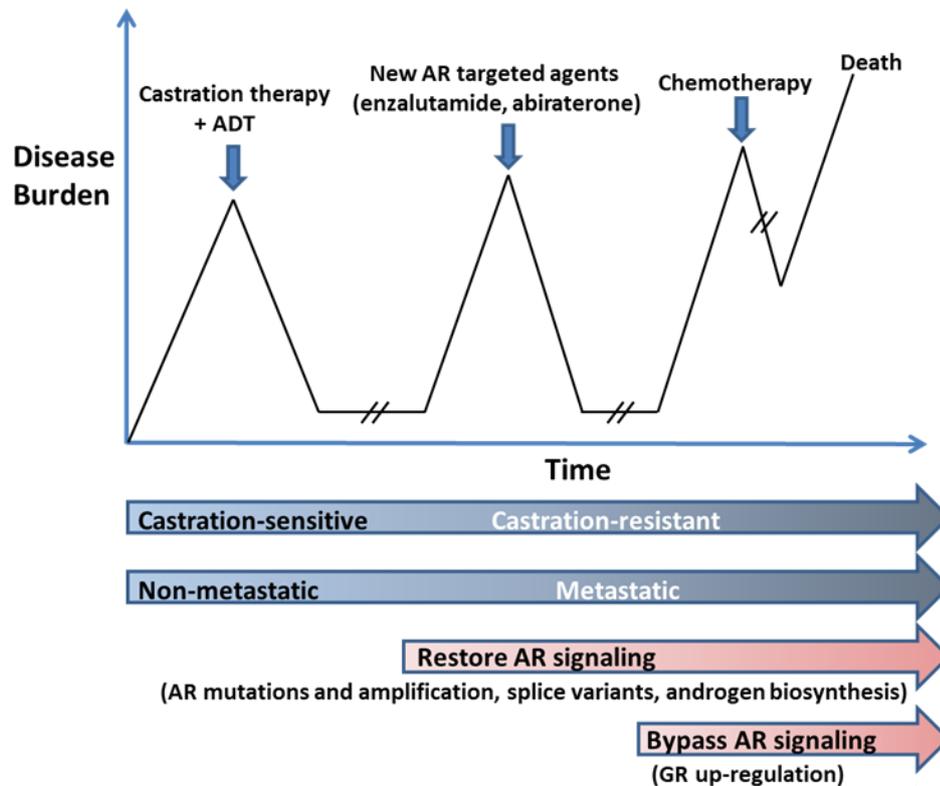
There has been a variety of mechanisms identified that describe maintenance of AR activity in CRPC (summarized in Figure 7A) (Chandrasekar et al., 2015).

**A.**

Cell membrane



**B.**



**Figure 7. AR signalling persists in advanced prostate cancer leading to disease progression and death. A.** Key mechanisms of resistance that emerge in response to AR-targeted therapy. AR mutations are depicted by \*. **B.** Sequence of treatment options for PC patients at different stages during PC progression. Taken from Zenith Epigenetics.

### **1.9.1. AR Amplification**

AR gene amplification has been observed in more than 50% of CRPC patients (Robinson et al., 2015). Increased AR gene copy number results in a 6-fold increase in AR protein levels (Linja et al., 2001). Amplification was not observed in untreated tumours suggesting that this mechanism arose in response to treatment which would hypersensitise cancer cells and compensate for the low levels of androgens (Visakorpi et al., 1995) However, the exact mechanism underlying this phenomenon hasn't been described yet. Initial *in vitro* experiments have demonstrated that hormone-naïve LNCaP cells acquire a CRPC-like phenotype following AR overexpression and hence become capable of proliferation and expansion in castrate conditions (Chen et al., 2004). The clinical relevance of this particular AR genetic aberration is also highlighted by the fact that AR amplification has been detected in circulating tumour cells (CTCs) and circulating tumour-derived DNA (ctDNA) isolated from patients with CRPC and has been associated with treatment resistance to AR targeted therapy including enzalutamide and abiraterone (Conteduca et al., 2017). Finally, a recent study has shown that when co-expressed with AR-V7 or ARv567es, AR-FL could translocate to the nucleus and regulate gene expression in the absence of androgens (Cao et al., 2014), suggesting that this mode of AR gene regulation may also be applied in tumours with simultaneous AR amplification and AR-V expression.

### **1.9.2. Nuclear Receptor Cross-talk**

GR expression was shown to be elevated in CRPC patients who demonstrated resistance to next-generation antiandrogens such as enzalutamide. In fact, clinical assessment of bone metastases prior to and after treatment indicated that GR was significantly elevated in response to enzalutamide treatment, implicating GR function in treatment failure and PC progression. As previously mentioned, AR and GR share the same conventional DNA binding motifs. AR displacement from chromatin, induced by AR inhibition, leads to GR recruitment to AREs and GR driven transcription of AR target genes, thus sustaining an androgenic gene signature. Moreover, GR depletion in the enzalutamide-resistant LNCaP-derived xenograft model LREX supported this finding by showing that the GR cistrome and transcriptome share significant overlaps with those of AR (Arora et al., 2013). Finally, a more recent study has demonstrated that GR blockade by shRNA and pharmacological inhibition by mifepristone

significantly diminished cell growth and spheroid formation of AR positive prostate cancer cell lines including LAPC4 and CWR22Rv1 cells highlighting the dependence of these cells on GR (Puhr et al., 2018). It is therefore conceivable that GR inhibitors may provide clinical benefit in CRPC patients. Importantly, an early phase clinical trial (NCT02012296) is currently ongoing; investigating the treatment of patients with enzalutamide in combination with mifepristone or enzalutamide alone to establish the role of the GR in CRPC. Implications in the clinic involve the use of glucocorticoids in combination with anti-androgens such as abiraterone. Inhibition of CYP17A1 activity by abiraterone leads to an increase in corticosterone production to compensate for the drop in glucocorticoid levels. High corticosterone production results in elevated adrenocorticotrophic hormone (ACTH) levels and secondary mineralocorticoid excess which is associated with severe side effects including high blood pressure (Attard et al., 2012). To tackle this, glucocorticoids such as prednisone are co-administered to patients receiving abiraterone (Auchus et al., 2014). Given the tumour-promoting role of GR in PC, administration of glucocorticoids may facilitate selection of GR positive cell populations to allow tumour progression. Finally, glucocorticoids such as cortisol have been shown to act as agonists to certain AR mutants to promote androgen-independent cell growth and ultimately disease progression (Zhao et al., 2000).

### ***1.9.3. AR co-regulator dysregulation***

As previously mentioned, AR interacts with a plethora of co-regulators which modulate its activity. They display different cellular functions and can be grouped to chromatin remodellers, histone modification enzymes, components of the proteasome and spliceosome as well as components of the DNA repair machinery. Dysregulation of these co-regulators is commonly seen in advanced PC and significantly influences AR activity. Increased expression of AR co-regulators has been documented to lead to enhanced activity of AR even in low androgen conditions (Heemers and Tindall, 2007).

Common examples of AR modulators are the pioneer factors forkhead box A1 (FOXA1) and GATA2 which have been shown to be essential for AR activity at certain genomic loci. Expression levels are significantly increased in advanced disease compared to primary PC resulting in altered AR occupancy across chromatin (Chaytor et al., 2019; Jones et al., 2015). In CRPC, FOXA1 opens up *de novo* chromatin regions, promotes redistribution of AR across the genome and allows it to bind and regulate transcription of a novel set of genes which is

consistent with a CRPC-like gene signature (Robinson et al., 2014). Moreover, GATA2 expression levels are elevated by antiandrogen treatment, which leads to upregulated GATA2-mediated transactivation of AR expression via binding to transcriptional regulatory sites upstream of the AR gene (Wu et al., 2014).

p300, a global transcription co-activator facilitates bridging of transcription factors to the basal transcriptional machinery. It also functions as a histone acetyl-transferase which acetylates AR and enhances its transcriptional activity in an androgen-dependent manner. Interestingly, its levels are significantly increased in advanced PC in response to androgen starvation. Despite the knowledge gap in the mechanism of dysregulation, elevated expression of p300 leads to activation of AR and augments AR signalling in castration conditions (Heemers et al., 2007).

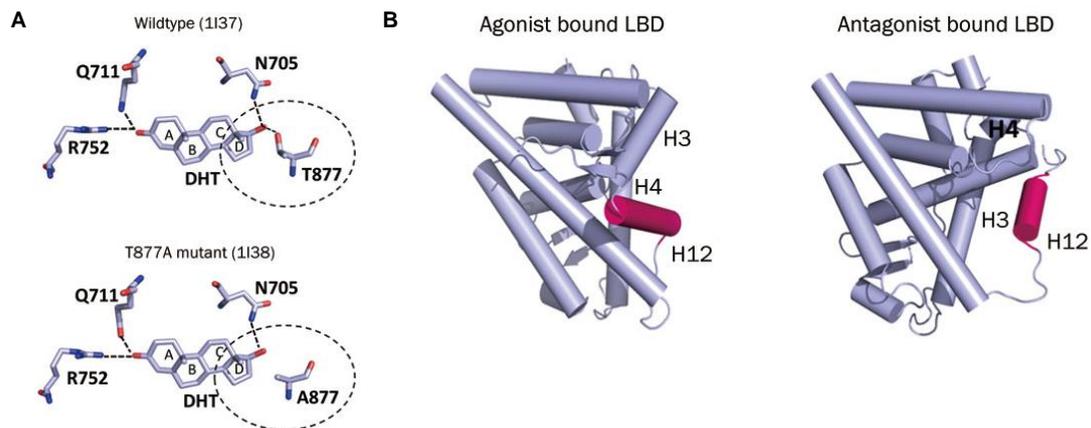
#### ***1.9.4. Gain-of-function mutations within the LBD of AR***

The development of AR mutations is a key molecular event commonly identified in the transition from treatment-sensitive to castrate-resistant PC. Aberrant AR signalling via mutations drives progression of advanced disease that is largely refractory to commonly used anti-AR therapies. The acquisition of AR gain-of-function mutations during ADT is a well characterised mechanism of hormone escape and treatment failure and has been reported to occur in over 60% of castrated patients. Recent advances in next-generation sequencing technologies have highlighted the LBD of AR as a hotspot for accumulation of point mutations (Brooke and Bevan, 2009).

- **The T877A mutation**

In 1990, Veldscholte et al. identified a single point mutation within the LBD of AR in LNCaP cells, which was also later identified in PC specimens derived from patients with bone metastasis (Taplin et al., 1999; Veldscholte et al., 1990). The T877A mutation occurs in exon 8 and alters AR response to ligands. Structural studies of the AR in complex with compounds, such as DHT, highlighted the key events that occur and explained the differential response of the receptor. In particular, substitution of the bulkier threonine to alanine results in a more flexible and open conformation of the ligand binding pocket of the AR (Figure 8A), allowing binding of ligands other than DHT and subsequent activation of the receptor under castrate conditions (Tan et al., 2015). Over the past two decades, a novel mechanism of AR adaptation

has emerged in response to anti-androgen therapy with different research groups attempting to pinpoint the paradoxically enhancing effect of anti-androgens on AR activity. In particular, scientists demonstrated that the T877A mutant was not inhibited by the commonly used anti-androgen flutamide. Conversely, the drug switched to a ligand with an agonistic effect on AR activity, which concurrently remained high, leading eventually to tumour cell growth and progression in humans and *in vitro* models of disease (Sun et al., 2006; Taplin et al., 1999).



**Figure 8. AR promiscuity to alternative ligands.** The wild-type AR (top) and the T877A AR mutant (bottom) in complex with DHT. Substitution of threonine (T) to alanine (A) at position 877 generates a receptor with a larger ligand-binding pocket. **B.** The conformational changes of helix 12 (highlighted in magenta) induced upon binding of pharmacological agents, such as Casodex. H12 is displaced and blocks the interface of helix 3 (H3) which is essential for coactivator binding. Adapted from Tan et al., 2015.

- **The H874Y mutation**

The H874Y mutation occurs within exon 8 and has been found in clinical samples as well as in the castration-resistant CWR22Rv1 cell line. Very similar to T877A, substitution of histidine at position 874 results in a receptor which is responsive to alternative bulkier ligands due to the formation of a binding cleft with larger volume capacity. Histidine, in fact, seems to face away from the LBD and therefore is not involved in direct interaction with ligands. Alongside its wider selectivity, the mutant exhibits enhanced activity during flutamide treatment which is in line with an agonistic switch of the drug. Together, these mechanisms comprise a selective advantage for AR to remain active and sustain tumour growth and ultimately spread in castrate conditions (Duff and McEwan, 2005).

- **The W741L/C mutation**

Bicalutamide (commercially known as Casodex), a structurally bulky anti-androgen, has been shown to switch to a potent agonistic ligand which promotes tumour growth in xenograft and cell line models and more importantly in patients with advanced PC, upon long administration of the drug (Hara et al., 2003; Yoshida et al., 2005). The unexpected outcome highlighted the need of structural-based studies to decipher the interaction between Bicalutamide and AR and subsequently revealed the mechanism that causes the observed differential agonistic action of the drug. Bicalutamide, a first-generation anti-androgen, normally acts by displacing helix 12 of the LBD upon binding to AR. As a result, conformational changes of the receptor are induced, which are no longer compatible with an active state of the receptor (Figure 8B) (Tan et al., 2015). Although sophisticated, inhibition by bicalutamide can be overcome by the acquisition of a gain-of-function mutation in position 741 in the LBD where tryptophan is substituted to either leucine or cysteine which increases the volume of the AR cleft. Thus, AR can now accommodate bicalutamide with helix 12 remaining intact in its initial active position. This subsequently allows activation of the receptor by bicalutamide and finally leads to tumour maintenance and progression (Bohl et al., 2005).

- **The F876L mutation**

Despite the initial excitement about the pharmacological efficacy of enzalutamide in CRPC clinical trials, where remarkable tumour regression and decline in PSA levels initiated a new era in PC treatment, tumours eventually developed mechanisms to escape pharmacological inhibition of AR and sustain their addiction to AR signalling (Karantanos et al., 2015). The AR F876L mutation was identified *in vitro* and in patients with CRPC who have received and become resistant to enzalutamide. Consistent with flutamide and bicalutamide, enzalutamide binds to mutated AR, promotes its translocation to the nucleus and the concurrent transcription of genes, which are necessary for proliferation and progression to more advanced and lethal tumours (Joseph et al., 2013; Korpál et al., 2013). Pharmacological inhibition of the wild-type AR by enzalutamide relies on the disorganization of the regular 3D structure of the receptor. Direct interaction of enzalutamide with phenylalanine 876 allows repositioning of helix 11 of the LBD. This repositioning of helix 11 hinders its adjacent helix 12 from obtaining an active position. It is conceivable that mutation of phenylalanine to leucine

abolishes the formation of this inhibitory complex resulting in restored AR activity in CRPC (Joseph et al., 2013; Tan et al., 2015).

Functional characterisation of AR mutants is crucial to elucidate their yet unknown role *in vivo* and to identify downstream effectors which could provide novel biomarkers and avenues for inactivation in advanced PC.

It is noteworthy that loss-of-function mutations also occur in the AR gene. A relatively recent study examined the activity of a panel of 45 missense mutations detected in PC with metastasis or high Gleason scores. These mutations stretch along the entire length of the AR gene and occur with relatively comparable frequencies. Examples include the L57Q and D221H mutations in the NTD and the L744F and Q902R mutations in the LBD. Interestingly, almost half of the mutations (28 out of 45) led to a loss of function at most or all concentrations of DHT used in a luciferase assay (Hay and McEwan, 2012). However, the implication of AR loss-of-function mutations in PC progression is unknown to date.

#### **1.10. AR splice variants**

Alternative splicing, rearrangements of the AR gene and proteolytic cleavage events result in the presence of a wide range of truncated AR transcripts in response to ADT (Dehm and Tindall, 2011).

##### ***1.10.1. Origin of AR splice variants***

Most commonly, altered splicing events lead to expression of splice variants which are constitutively active and act independent of androgens since they lack the LBD. AR isoforms were originally observed in the CWR22Rv1 cell line (Tepper et al., 2002). CWR22Rv1 cells have been established from an initially androgen-dependent human tumour engrafted in immunodeficient mice (known as CWR22R-2152 xenograft), which nevertheless relapsed within a year after ADT (Sramkoski et al., 1999).

The CWR22Rv1 cell line was established in an attempt to expand the panel of PC cell lines to provide more reliable and robust models and therefore a better understanding of the underlying mechanisms of PC resistance and progression, which would ultimately lead to more efficient PC therapies. The CWR22Rv1 cell line is now indeed the most representative

model to study PC progression since it resembles the molecular and clinical background of patients with advanced disease (Tepper et al., 2002).

A novel AR-FL species which exhibits exon 3 duplication and thus migrates slower than the canonical AR-FL in an SDS-PAGE gel, as well as a shorter AR isoform of ~75 kDa which lacked the AR LBD, were observed in CWR22Rv1 lysates. Expression of that novel truncated AR conferred the advantage of consistent growth of CWR22Rv1 cells in androgen-depleted conditions. These initial observations led to speculation that the novel AR variant may drive PC via promoting tumour androgen independence (Tepper et al., 2002).

Additionally, Libertini et al. confirmed the presence of a truncated AR isoform in the CWR22Rv1 cell line, which was generated by proteolytic cleavage of AR-FL and was transcriptionally more potent than AR-FL. Interestingly, expanding these observations to patient material, the same group observed high expression of a variant with similar size to the CWR22Rv1 AR truncated isoform in cancer specimens compared to specimens with benign prostate hyperproliferative disorder (BPH) (Libertini et al., 2007).

Following these early observations, several more studies have reinforced the concept that AR variants are critical drivers of CRPC and have attempted to characterise the molecular and functional activities of these AR truncated variants in PC (Cao et al., 2014; Lu et al., 2015; Watson et al., 2010).

### ***1.10.2. Generation of AR splice variants***

Alternative splicing has been the most prevalent mechanism reported in the literature for generating truncated AR variants. Inclusion of cryptic exons, which are present in intronic regions in the AR locus, result in the generation of mature atypical transcripts which lack the LBD and harbour unique sequences at their 3' ends (Figure 9) (Dehm and Tindall, 2011).

In 2008, Dehm et al. described two novel AR species of ~75 kDa each, in CWR22Rv1 cells. These variants arise from the inclusion of the cryptic exon 2b in the mature mRNA transcripts, which in turn encode AR isoforms composed of exons 1, 2, 3 and 2b (hereafter called V1/2/3/2b) and exons 1, 2 and 2b (hereafter called V1/2/2b) (Figure 9). In fact, differential siRNA knockdown targeting one exon at a time in the LBD or the NTD indicated differential expression of the AR-FL or the AR-Vs, respectively, meaning that these three AR species arise

from at least two distinct transcripts via alternative splicing (Dehm et al., 2008). This contradicts previous findings which indicated AR isoform generation via enzymatic cleavage of AR-FL (Libertini et al., 2007).

Interestingly, down-regulating expression of AR-Vs significantly impacted on AR target gene transcription in steroid-free conditions, a phenotype which was not observed upon knockdown of AR-FL. Furthermore, depletion of the variants significantly retarded cell proliferation in the absence of androgens, a phenotype which was reversed upon restoration of AR-V expression. Finally, assessment of the expression levels of the variants versus the full-length receptor indicated a significant difference at mRNA level with V1/2/2b and V1/2/3/2b composing the largest portion of the overall AR pool in CWR22Rv1 cells. Taken together, these findings suggest a key role of these AR-Vs in promoting androgen independent growth, which is attributed to loss of the LBD and concurrent escape of androgen starvation (Dehm et al., 2008).

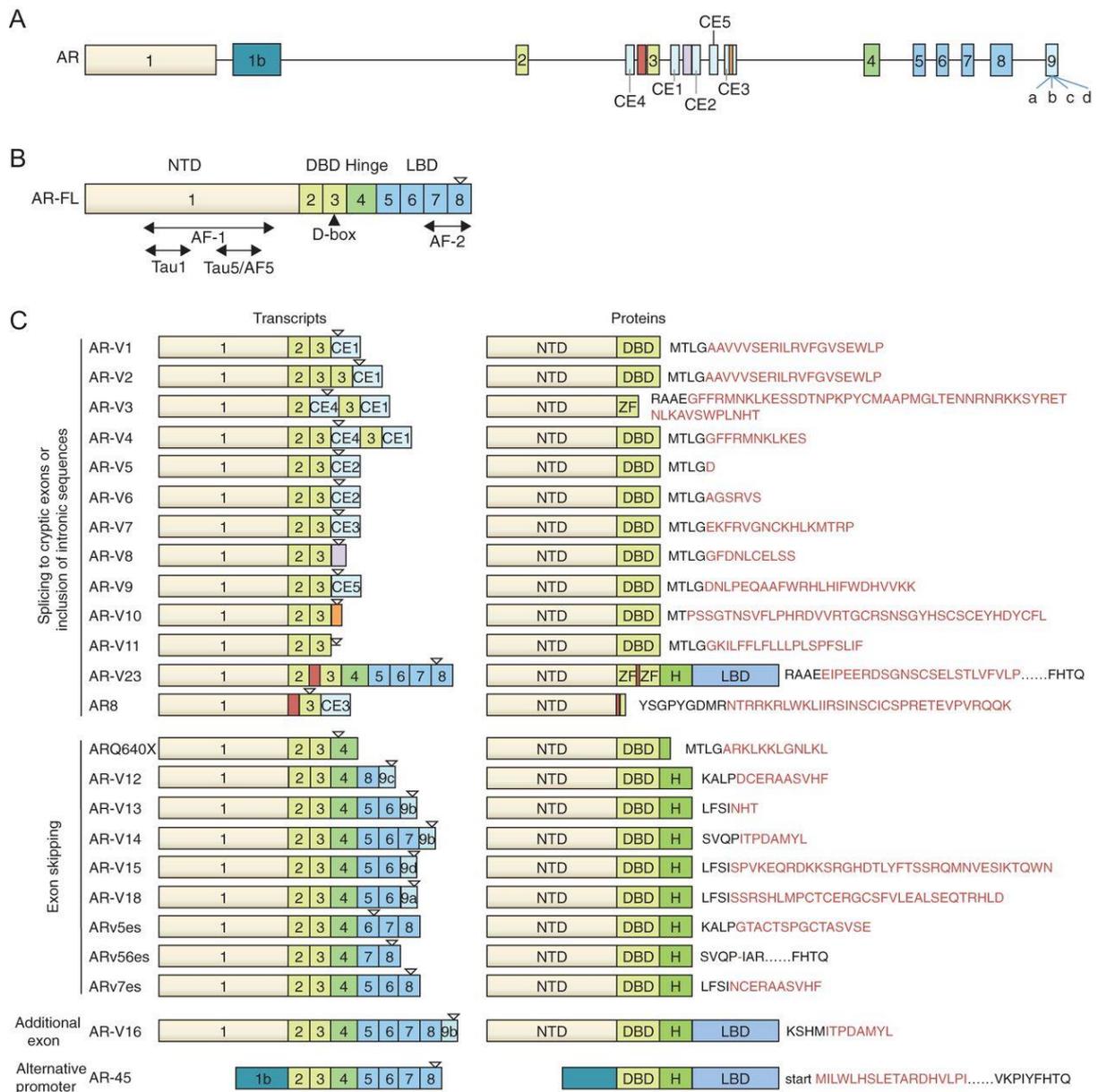
The most extensively studied AR splice variant so far was identified in 2009, when Guo et al. focused on the characterisation of novel AR variants in hormone refractory cell lines. The truncated isoform, initially called AR3 by Guo et al., but now reclassified as AR-V7, consisted of exons 1, 2, 3 and CE3 (cryptic exon 3) and was highly expressed in ADT resistant cell lines, such as the CWR22Rv1 cell line and androgen independent LNCaP derivative cell lines, compared to ADT responsive cell lines. Like V1/2/3/2b, knockdown of AR-V7 was sufficient to induce cell cycle arrest in CWR22Rv1 cells grown in steroid-depleted media, whereas overexpression led to an enhanced proliferative phenotype and elevated PSA levels in parental LNCaP cells grown in castrate conditions. V1/2/3/2b (called AR5 by Guo et al.) was also detected in this study, although AR-V7 was the most abundant and potent AR truncated species (Guo et al., 2009).

The contribution of AR-V7 to tumour progression can be further highlighted by the elevated levels of mesenchymal associated markers observed upon overexpression of the variant in epithelial PC cell lines, such as in LNCaP, suggesting that the variant drives tumour progression to a metastatic state by promoting EMT (epithelial-to-mesenchymal transition) (Cottard et al., 2013).

Important insights were reported by Hu et al. who extensively interrogated the AR intronic regions *in silico* for the presence of cryptic exons and identified seven AR splice variants, some of which were completely novel. All variants were truncated and lacked the LBD (Figure 9). AR-V7 was identified as the predominant AR-V species and a specific antibody was developed against it to map its subcellular localisation in CRPC cell lines and examine its expression profile in a wide collection of clinical specimens. Resulting data indicated that AR-V7 constitutively resided in the nucleus of CWR22Rv1 and VCaP cells in the absence of androgen, thus remaining constitutively active and sustaining a castration-refractory phenotype. More importantly, using immunohistochemistry, the variant demonstrated a significantly higher expression (20-fold) in malignant specimens derived from CRPC patients compared to normal tissue, which was associated with poor clinical outcome indicated by PSA recurrence (Hu et al., 2009).

In summary, AR-Vs play a pivotal role in tumour survival and progression. Agents such as enzalutamide and abiraterone, which target full-length AR, show poor efficacy *in vitro* and in patients due to expression of AR splice variants which are not responsive to these agents; supporting persistent AR transcriptional activity and as a result confer resistance to these agents (Antonarakis et al., 2014; Li et al., 2013).

Conceivably, there is an urgent need to elucidate the molecular mechanisms that regulate these AR isoforms to improve our understanding of how these proteins function *in vivo* and potentially identify regulators and effectors that could become rationale drug targets for new PC therapeutic interventions.



**Figure 9. Spectrum of AR splice variants (AR-Vs) emerging in response to ADT.** **A.** Illustration of the AR gene structure indicating canonical and cryptic exons (CE). **B and C.** The molecular mechanisms responsible for the generation of each AR-V mRNA transcript are indicated on the left-hand side. The novel/unique protein sequences incorporated at the C-terminus of each AR-V are highlighted in red. Arrowheads indicate translation termination. AR-Vs with intact exons 3 and 4 encompass a canonical nuclear localisation signal. Taken from Cao et al., 2016.

### 1.11. CRISPR genome editing

#### 1.11.1. Origin of CRISPR and its application in biomedicine

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) is a technology which originates from bacteria. Surprisingly, short DNA repeats, called spacers which were found in sequenced prokaryotic genomes showed high homology to viral sequences. It was initially

hypothesised and later confirmed that these sequences were indeed derived from viruses that infected bacteria; the latter developing a defence mechanism to destroy foreign DNA and hence prevent recurrent infections. More precisely, bacteria exploited these sequences to target the same or similar DNA fragments of viral origin at subsequent infections and degrade them which resembles an immune response (Sontheimer and Barrangou, 2015). This sophisticated defence mechanism of bacteria against viruses became highly attractive to scientists who attempted to transfer and adopt the novel findings to mammalian cells and animals, such as mice, in order to engineer their genome (Doudna and Charpentier, 2014).

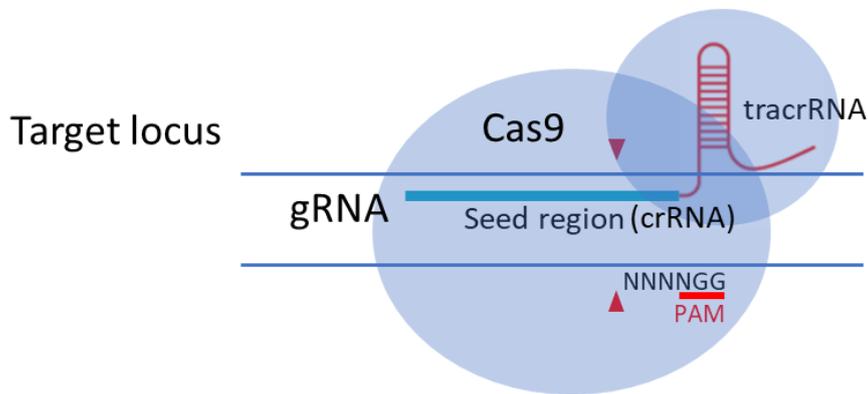
Nowadays, many labs around the world use CRISPR to generate physiologically relevant cell line and mouse models in order to better understand human disease (Inui et al., 2014). In fact, CRISPR has been applied to mammalian cells and more importantly to human primary and stem cells to conduct knock-in experiments in order to edit sequences via introducing precise alterations, such as nucleotide substitutions and tagging endogenous genes in a more clinically-relevant scenario to correct mutations which are responsible for genetic syndromes (Kiskinis et al., 2014; Schwank et al., 2013; Xiang et al., 2019). Additionally, CRISPR has been applied to induce gene knock-outs to study physiological gene function in a preclinical context (Everman et al., 2018).

### ***1.11.2. Mechanism of CRISPR editing***

CRISPR relies on the activity of a DNA nuclease, called CRISPR-associated protein 9 (Cas9) which is recruited to the target DNA sequence by a short RNA fragment known as guide RNA (gRNA). The latter consists of two separate RNA fragments which hybridise to one another to form a single functional gRNA. The first fragment consists of a seed sequence of 20 nucleotides, called CRISPR RNA (crRNA) which is complementary to the target site and associates with Cas9 to form interference complexes, and a scaffold RNA sequence of 89 nucleotides which comprises the trans-activating gRNA (tracrRNA), which resembles a handle necessary for complex stabilisation with Cas9 via intermolecular interactions (Figure 10) (Jinek et al., 2012).

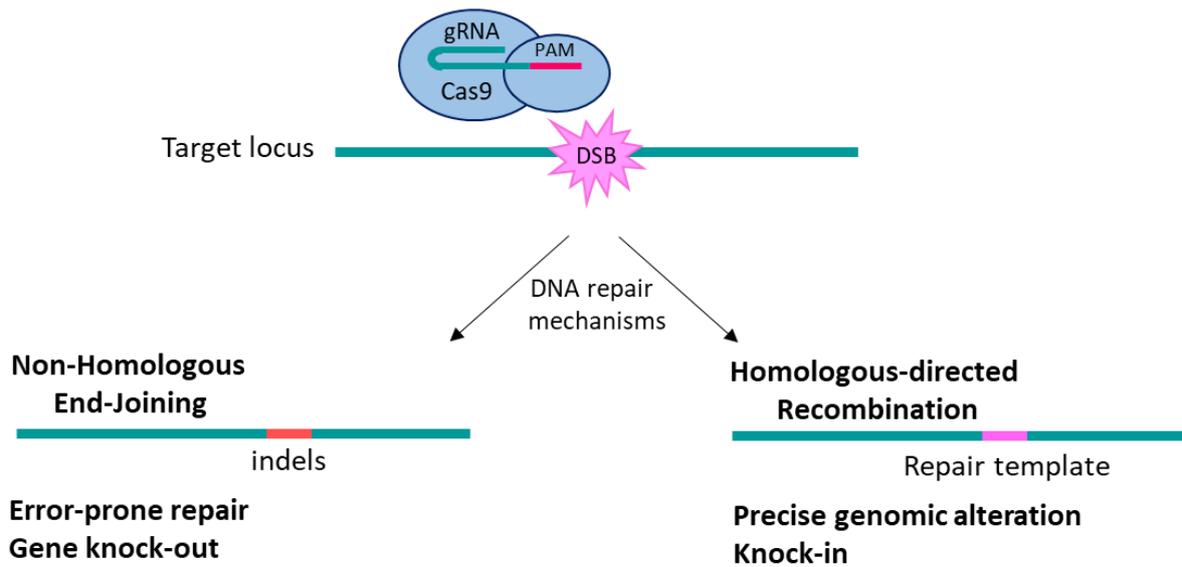
The target DNA sequence must be located adjacent to a PAM (Protospacer Adjacent Motif) sequence which is essential for recognition of the target site by Cas9. Cas9 optimally recognises NGG PAMs which are found quite frequently across the human genome permitting

editing of nearly any sequence of interest (Figure 10). Full complementarity between the seed region of the gRNA and the target region is required for Cas9 to tightly dock and acquire a 3D conformation consistent with an active state. Cas9 will then cleave both strands of DNA 3 base pairs upstream of the PAM leaving blunt DNA ends (Ran et al., 2013).



**Figure 10. CRISPR complex formation at target locus.** Arrowheads depict DNA cleavage site 3 bp upstream of the PAM sequence.

Subsequent repair mechanisms are rapidly triggered to fix the double strand break (DSB) induced by Cas9. The most prevalent signalling cascade activated is the error prone NHEJ (Non-Homologous End Joining) which accounts for frame shifts, missense mutations and incorporation of premature stop codons which invariably result in gene disruption and knock-out. On the contrary, HDR (Homology Directed Repair) mediates precise modifications when a donor DNA template, which shares high degree of homology with the target sequence, is provided (Figure 11). Importantly, HDR events are extremely rare (frequency of <0.1%) and very often outnumbered by NHEJ making knock-in experiments and precise gene editing remarkably challenging (Sander and Joung, 2014).



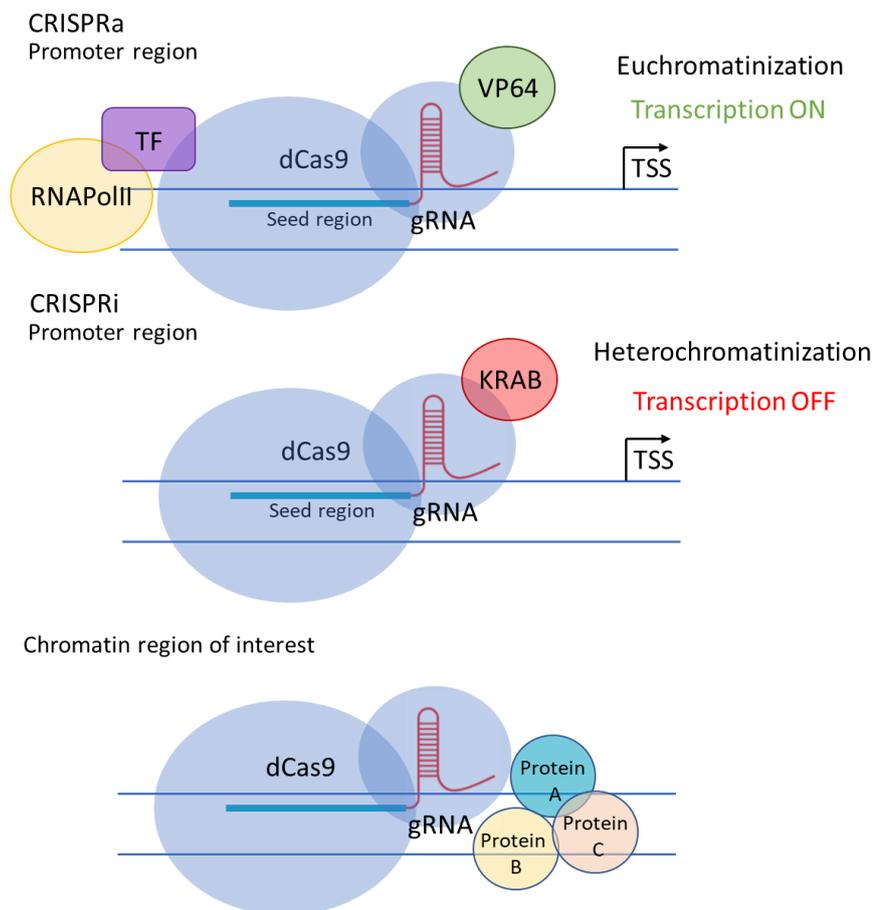
**Figure 11. Repair signalling pathways activated upon DSB induction by Cas9.**

CRISPR is a highly delicate and universal technology which can nowadays be applied with ease to everyday-life sectors such as biomedicine to tackle life-threatening disorder and agriculture to improve crops, for example (Arora and Narula, 2017; Smits, 2019). Its potential remains extremely high and promising with improved versions of the system constantly being developed to enhance features such as the editing efficiency of the CRISPR complex, including chemically modified gRNAs to increase their stability and more accurate cleavage by different species of Cas nucleases with better proof-reading activity to eliminate potential off-target editing (Basila et al., 2017; Hu et al., 2018). Application of CRISPR in human embryos has also been reported and despite the fact that it may comprise a real breakthrough in medicine, it has raised a lot of concerns and ethical issues (Schenkwein and Ylä-Herttuala, 2018). Clinical trials using CRISPR as a tool for gene therapy are also underway or due to commence to treat various types of cancer after considerable controversy associated with its safety (Ghosh et al., 2019). Nevertheless, the technology is far from perfect and more scientific efforts are currently made to improve it for the benefit of patients with severe diseases.

### **1.12. Expansion of the CRISPR toolbox**

CRISPR is not limited to gene editing. It has been repurposed to transiently alter mRNA transcription and gene expression programmes as well. More precisely, a catalytically inactive version of Cas9, named dead Cas9 (dCas9) can bind to DNA just like its wild-type counterpart

but it is not capable of mediating cleavage. It harbours two silencing mutations, D10A and H841A which reside in the RuvC1 and HNH nuclease domains, respectively, and are responsible for its enzymatic incompetence. Consistent with its wild-type counterpart, dCas9 is targeted to selective target loci by gRNAs. Traditionally, dCas9 has been exploited when in conjugation with transcriptional activators (i.e. VP64) or repressors (i.e. KRAB) to activate (CRISPRa) or block (CRISPRi) gene expression, respectively (Kampmann, 2018). More recent approaches implicate RNA-guided dCas9 targeting to specific chromatin regions in order to isolate and identify chromatin associated protein complexes which also associate with dCas9 at these loci and may play regulatory roles in gene expression or chromatin remodelling (Figure 12) (Fujita et al., 2016; Liu et al., 2017). A detailed description of this latter deviation of CRISPR is outlined in Chapter 6.



**Figure 12. Expansion of the CRISPR toolbox.** A nuclease deficient Cas9 (dCas9) can be fused to transcriptional activators (i.e. VP64) or repressors (i.e. KRAB) to potentiate transcription activation (CRISPRa, top panel) or inhibition (CRISPRi, middle panel) at target promoters, respectively. Alternatively, dCas9 can act as a docking platform for recruitment of chromatin regulators at target loci (bottom panel) which can then be isolated and identified by proteomic approaches.

### 1.13. Mechanism of splicing

Splicing was first observed in 1977 (Berget et al., 1977). It refers to mRNA processing whereby introns are removed from precursor mRNA (pre-mRNA) whilst the remaining exons are ligated to form mature mRNA transcripts which are finally translated to proteins (Clancy, 2008).

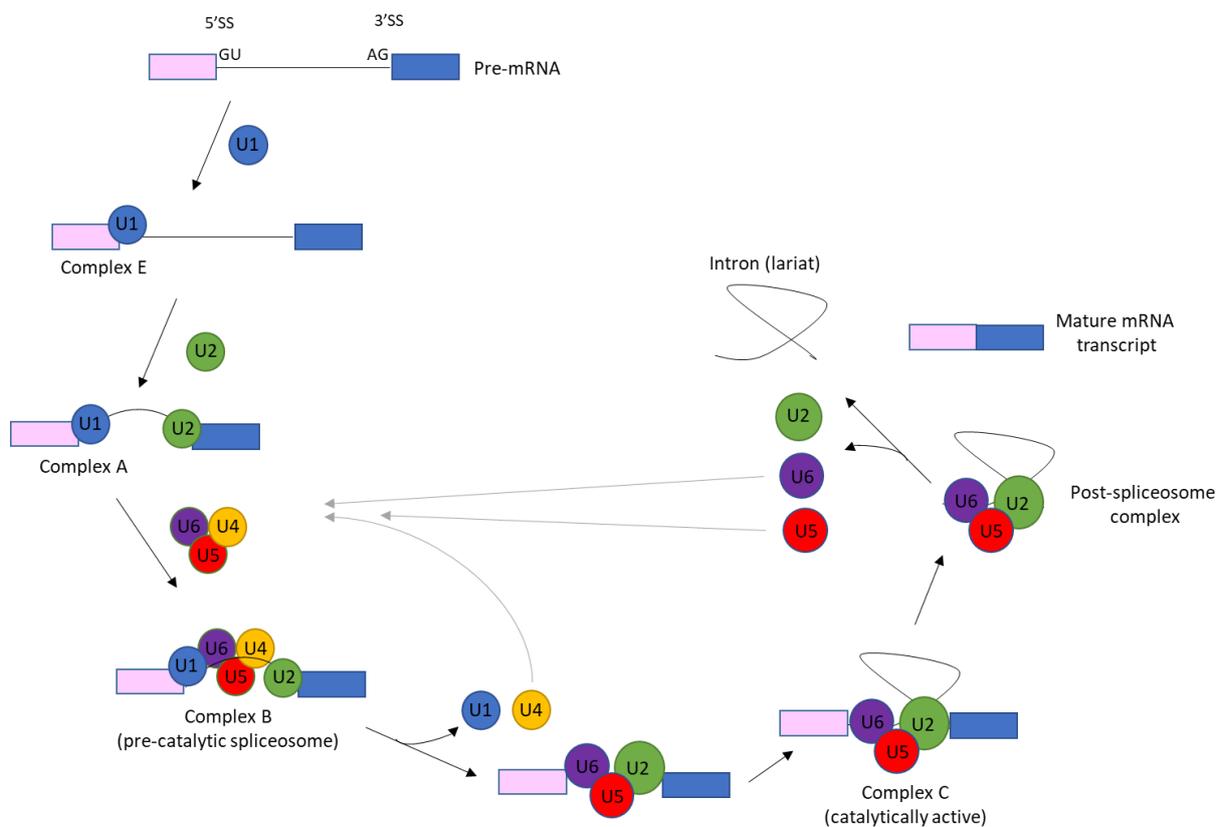
Splicing is catalysed cooperatively by a group of proteins called splicing factors (SFs) which form a macromolecular protein complex known as the spliceosome. The most studied and best characterised SFs belong to the small nuclear ribonucleoprotein (snRNP) and Serine/Arginine (SR)-rich protein families. They bind to intron-exon boundaries by recognising specific sequences on each side of an intron termed 3' and 5' splice sites (SSs) and recruit more SFs to form an active spliceosome (Shen and Green, 2004).

The 5' SS is a GU at the start of an intron while the 3' SS is an AG at the end of an intron. A branch point sequence (BPS) is located between the two SSs, 15-50 nucleotides upstream of the 3' SS. The aforementioned pre-mRNA features comprise essential regulatory sequences for spliceosome assembly. Mutations in SS disrupt splicing (Clancy, 2008). Additional *cis*-regulatory sequences have been identified across a pre-mRNA substrate which function as docks for recruitment of *trans*-acting factors which can either enhance (enhancers) or repress (silencers) intron release (Wang and Burge, 2008).

Spliceosome assembly requires a series of RNA-protein interactions. Components of the spliceosome in direct contact with their mRNA substrate have been identified as well as intermediate components of the spliceosome required for scaffolding.

Splicing is an extremely dynamic process characterised by assembly of concurrent enzymatic complexes (Figure 13). The sequence of recruitment of SFs to their substrates is not stochastic. Five different snRNPs play key roles in the formation of the different RNA-protein complexes during splicing. Initially, the early splicing complex (complex E) is formed when the U1 snRNP binds to the 5' SS of the target intron. Binding of U1 snRNP is followed by recruitment of an SRSF to stabilise the interaction. Part of complex E is the BPS-binding protein SF1 as well as the U2 auxiliary factor (U2AF). The latter consists of the U2AF35 and U2AF65 subunits which bind to the 3' SS of the target intron. Complex A is then formed, whereby SF1 is released from the BPS and replaced by U2 snRNP. Following complex A assembly, complex B is then formed by recruitment of the U4, U6 and U5 snRNPs. Complexes

E, A and B constitute the core spliceosome which is not active at this stage and cannot mediate splicing as internal rearrangements of the snRNPs are required for activation. In particular, disruption of the interaction between the U4 and U6 snRNPs by helicases leads to release of U1 and U4 allowing interaction between U2 and U6 to form an active spliceosome. Recruitment of U5 to the spliceosome brings U2 and U6 in direct contact and all together form complex C which finally catalyses intron removal as a lariat. The spliceosome components can then be dissociated, recycled and used at different sites (Will and Lührmann, 2011).



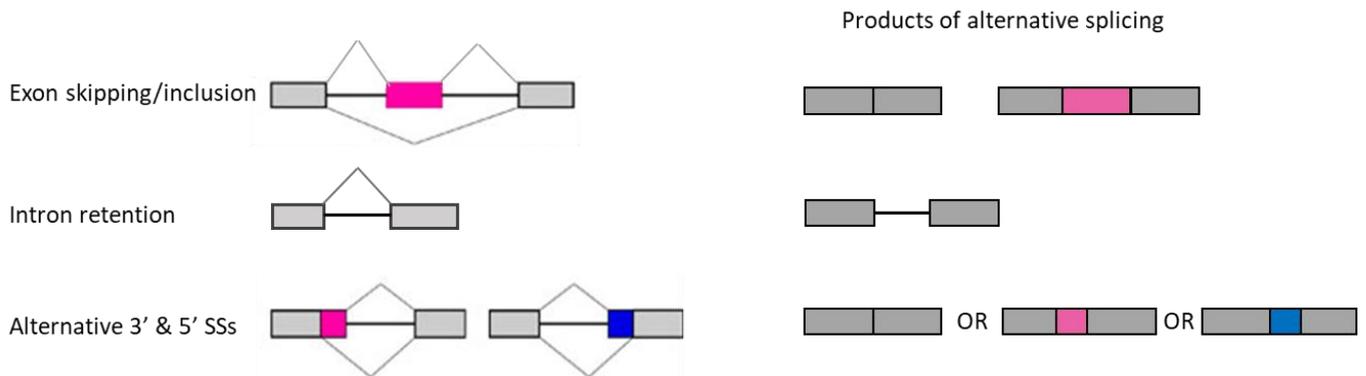
**Figure 13. Schematic of spliceosome assembly.** Sequential recruitment of the 5 snRNPs to their pre-mRNA target during spliceosome assembly. The key spliceosome complexes E, A, B and C are shown.

### 1.14. Alternative splicing

Early observations in the 1980s indicated that more than one immunoglobulin  $\mu$  mRNA transcripts can be produced from a single gene suggesting that alternative mRNA splicing occurs (Early et al., 1980). On average, each human gene can undergo 4 splicing events, remarkably diversifying the human transcriptome (Su et al., 2006). Alternative mRNA isoforms derived from the same gene might show tissue-specificity or might be produced at

certain developmental or differentiation stage or finally might be associated with a certain disease status.

There are different splicing events that occur to produce alternative transcripts. Intron retention, exon skipping/inclusion, alternative 3' and 5' splice site selection are only a few examples of alternative splicing events that are more frequently observed in the human genome (Figure 14) (Black, 2003).



**Figure 14. Modes of alternative splicing.** The most frequent splicing events and their mRNA products are presented.

Competition is observed between different components of the splicing machinery. Essentially, two different classes of splicing mediators exist and compete with each other to dictate the fate of an intron. In fact, SR proteins contain serine (S) and arginine (R) rich domains which mediate binding to RNA and promote splicing.

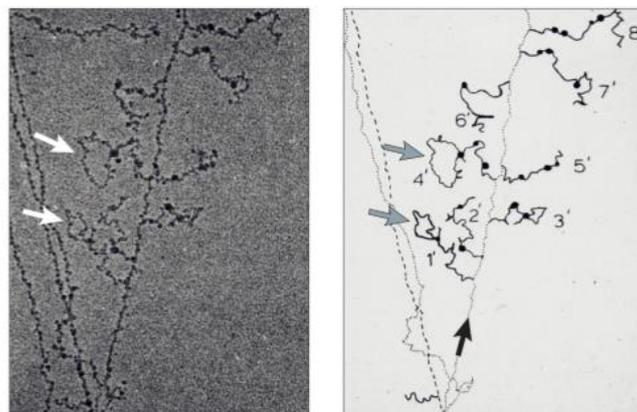
On the contrary, heterogeneous nuclear ribonucleoproteins (hnRNPs) which also bind RNA are negative regulators of splicing. They actively repress splicing via poorly understood mechanisms. Competing roles of these two classes of SFs indicate the presence of a more delicate regulatory network which dictates splicing decisions. More precisely, SRs bind to specific motifs in the pre-mRNA substrate called intronic and exonic splicing enhancers (ISEs and ESE respectively), In a similar manner, hnRNPs bind to intronic and exonic splicing silencers (ISSs and ESSs respectively) (Kornblihtt et al., 2013).

### 1.15. Co-transcriptional splicing

It was initially believed that splicing occurs right after transcription and upon completion of pre-mRNA synthesis by RNA polymerase II (RNA Pol II). Since 1980, a growing body of studies

have indicated that transcription and splicing synchronise and share regulators in space and time, a phenomenon known as spatiotemporal regulation of splicing or alternatively co-transcriptional splicing. Coupling in space refers to co-localization of transcription and splicing factors at the pre-mRNA template while coupling in time refers to coordination of mRNA elongation and maturation within the same time window (Baurén and Wieslander, 1994; Johnson et al., 2011; Tilgner et al., 2012). More specifically, the rate of transcription and mRNA emergence determines the rate of mRNA processing by providing a time window for interactions and complex formation. Finally, co-transcriptional splicing highlights that interactions are not only limited between RNA and protein molecules to form the spliceosome, but chromatin also comprises part of it and may create potential for additional intermolecular contacts and formation of novel complexes (Bentley, 2014).

Coupling of splicing with transcription was first observed in *Drosophila melanogaster*, whereby lariat introns appeared as bubbles showing up from the DNA template while RNA Pol II molecules actively transcribed it (Figure 15) (Beyer and Osheim, 1988).

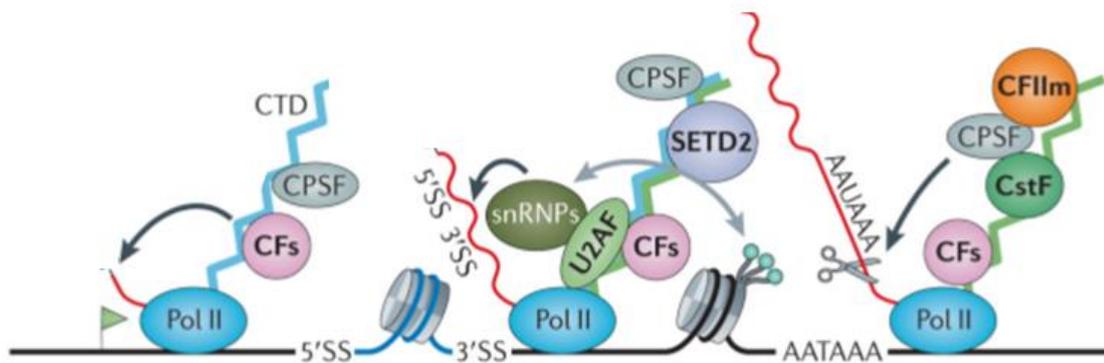


**Figure 15. ‘Miller spread’ electron micrograph of a *D. melanogaster* gene and its graphical representation (shown on left and right, respectively).** The DNA template is shown as a strand which is transcribed by RNA Pol II molecules. Several nascent mRNA transcripts emerge and are associated with spliceosome complexes indicated by black dots. White and gray arrows depict intronic lariats. Taken from Beyer and Osheim, 1988.

The carboxy-terminal domain (CTD) of the large subunit of RNA Pol II plays an essential role in SF recruitment to the pre-mRNA substrate. Mechanistic studies showed that phosphorylation of Serine 2 of the CTD, which is indicative of the elongating form of RNA Pol II stimulates mRNA processing by SFs and subsequent maturation, which is not only restricted

to splicing but also involves mRNA stabilisation processes such as mRNA capping and polyadenylation (McCracken et al., 1997).

The phosphorylated tail of RNA Pol II allows trafficking of SFs. It essentially provides a surface which serves as a dock for SFs to land and interact with the nascent pre-mRNA as soon as it exits the RNA Pol II channel (Figure 16) (Hsin and Manley, 2012). In addition, the CTD allows allosteric contacts of SFs which activate the latter so they can catalyse splicing (Ho and Shuman, 1999). Finally, the essential role of CTD to mRNA maturation is evident by the fact that mutations in the CTD or entire deletion of it abolishes spliceosome formation (McCracken et al., 1997).



**Figure 16. Schematic of co-transcriptional splicing.** Docking of SFs and splicing regulators to the CTD of RNA Pol II is necessary for pre-mRNA (shown in red) maturation. CTD, C-terminal domain; CstF, cleavage stimulation factor; snRNP, small nuclear ribonucleoparticle; U2AF, U2 auxiliary factor; CPSF, cleavage and polyadenylation specificity factor; CF, cleavage factor; SETD2, SET domain-containing 2. Adapted and modified from Bentley, 2014.

### 1.16. Alternative splicing in cancer

Reprogramming of alternative splicing occurs in cancer. Novel mRNA transcripts are produced aberrantly to re-shape the transcriptome of several types of cancer (Dvinge and Bradley, 2015). High throughput screening technologies such as RNA-seq have highlighted differential splicing signatures between cancer tissues and their normal counterparts. Splicing events in cancer either enhance tumour-promoting phenotypes or eliminate tumour-suppression processes.

Examples of genes which undergo aberrant mRNA splicing and result in oncogenic end-products include *BCL2L1*. Anti-apoptotic mRNA isoforms of this gene, such as BCLXL, confer

resistance of cancer cells to apoptosis and hence promote tumour progression. Another example is that of RAC1B, an alternatively spliced isoform of RAC1, which promotes proliferation and invasion of cancer cells (Sveen et al., 2016).

Pan-cancer sequencing studies of cancer patient material have identified critical somatic mutations in genes encoding SFs. SF3B1, in particular, is a cancer-critical SF which is frequently mutated in cancers such as leukaemia and breast. The mutations broaden its target sequence specificity in the pre-mRNA substrate and as a result the mutated SF can recognise novel 3' SS in the genome and potentiate novel splicing events which are associated with tumour progression and poor patient survival (Sveen et al., 2016).

Expression levels of mutation-free SFs alter in cancer. Alterations in SF expression levels are accompanied by a switch in the splicing pattern of target genes with oncogenic or tumour suppressive roles. For instance, SRSF1 has been shown to be upregulated in several cancers and its overexpression in mice caused malignant transformation and tumourigenesis, indicating that SRSF1 is a proto-oncogene in these cancer types and elevated expression of it drives splicing of genes with tumour promoting properties (Anczuków et al., 2012).

Upstream regulators of SFs are sometimes dysregulated in cancer and this leads to alterations in the stability and/or activity of their target SFs. It is well documented that SFs of the SR protein family are frequently phosphorylated by upstream kinases involved in signalling cascades such as AKT. The latter is constitutively active in cancer cells, resulting in continuous activation of its target SFs including SRSF1 via constitutive phosphorylation. This has marked consequences on the target pre-mRNAs and the decisions as to which splice variant will be preferred. For example, caspase 9 pre-mRNA is alternatively spliced by SRSF1. Constitutively active SRSF1 dysregulates caspase 9 mRNA isoform ratio and hence remarkably influences the apoptosis potential of cancer cells (El Marabti and Younis, 2018).

### **1.17. Splicing in PC**

Aberrant splicing events have been observed in genes which are involved in prostate cancer development. Pro-angiogenic and anti-angiogenic VEGF mRNA variants are produced using alternative 3' SS influencing sufficient or insufficient blood supply to tumours and subsequent tumour recession or growth, respectively (Munkley et al., 2017). In addition, CD44 pre-mRNA

isoform switch has been observed. This is in part regulated by the transcriptional co-regulator SND1 which is heavily involved in prostate cancer splicing events, as indicated by a recent CRISPR screen study (Cappellari et al., 2014). Finally, TSC2 is a tumour suppressor gene expressed in the prostate. The full-length protein reduces cell proliferation. However, an alternative promoter is chosen in response to androgens and results in the production of a short mRNA isoform which lacks the N-terminal domain and in contrast to its full-length counterpart, promotes cell growth and prostate cancer expansion (Munckley et al., 2017).

AR splicing is the area of focus of this study. A plethora of alternatively spliced AR mRNA variants emerges in response to anti-androgen therapy. Alternative splicing in this case is exploited by cancer cells to sustain AR signalling despite androgen ablation (Liu et al., 2014).

Over 20 different splice variants have been detected in prostate cancer cell lines and more importantly in prostate cancer clinical specimens including solid and liquid biopsies. However, our understanding of how these truncated receptors are generated remains poor.

Splicing of AR pre-mRNA seems to be more complicated than a single splicing event and this stems from the different splicing modes that take place to produce structurally different variants. Inclusion of different cryptic exons across the AR mRNA is responsible for the emergence of most variants. Selection of alternative polyadenylation sites downstream of these cryptic exons results in their inclusion in the mature AR mRNA transcript. Moreover, skipping of more than one exons has been observed and gives rise to shorter AR mRNA species with clinical relevance (Cao et al., 2016). However, little is known about the splicing mechanisms and components that mediate these processes.

The presence of multiple AR mRNA isoforms raises the question as to whether there is a constitutive spliceosome with certain components that regulate AR mRNA splice variant generation or whether it is more likely and conceivable that distinct spliceosomes specific to individual splice variants are assembled or finally there is possibly some kind of grouping. This is only a hypothesis though which remains elusive.

The number of high-throughput approaches to investigate AR splicing events and identify specific AR-V splicers are limited. HSP90 was identified as a candidate AR-V splicing factor from an siRNA screen in prostate cancer cells and HSP90 inhibitors such as onalespib were

proposed as a therapeutic approach to prevent AR-V7 production in patients (Ferraldeschi et al., 2016).

Candidate-based approaches have been pursued over the past 10 years or so and individual candidates have been identified and investigated for their contribution to AR-V mRNA generation. The majority of research efforts focus on identifying AR-V7 splicers as this particular variant is more prevalent in clinical samples and has been associated with poor patient survival and resistance to certain treatments. Nevertheless, there is now growing interest in other AR splice variants which are also clinically relevant and they have been detected in AR-V7 negative patients.

Sam68 plays a key role in mediating splicing events in prostate cancer. Recent work by Stockley *et al.* indicated that Sam68 is significantly upregulated in aggressive disease and its expression is associated with elevated AR-V7 mRNA expression (Stockley et al., 2015).

An important insight into AR-V7 splicing was provided by Liu et al. and his colleagues who demonstrated that ADT enhances AR transcription rates which in turn provides an optimal time window for interactions between the AR pre-mRNA transcript and certain splicing factors. The same study identified ASF/SF2 and U2AF65 as direct AR-V7 splicers recruited at enhancer sequences at the 3' SS and within CE3 (Liu et al., 2014).

Finally, SF3B2 and SF3B3 have also been shown by our lab and others to mediate AR splicing, not limited to AR-V7, in CRPC relevant cell lines (Chaytor et al., 2019; Kawamura et al., 2019).

### **1.18. Therapeutic targeting of splicing**

Targeting core spliceosome components has been the gold standard approach to disrupt early spliceosome formation. Several compounds with bacterial origin (i.e. FR901464) were developed and used *in vivo* and *in vitro* with decent cytotoxicity. However, their poor stability has limited their performance and use in larger therapeutic schemes. More robust drug design studies were then performed and generated novel compounds with improved pharmacokinetic properties, namely Spliceostatin A (SSA), a more effective compound derivative of FR901464 (Lee and Abdel-Wahab, 2016).

SSA is a small molecule inhibitor developed against SF3B, a key multi-subunit component of the U2 snRNP. It has demonstrated high potency in cancer cells including the HeLa cell line

with remarkable anti-proliferative properties. In fact, SSA treatment of cancer cells *in vitro* caused cell cycle arrest at G2/M at low nM doses by modulating splicing of key cell cycle regulators such as Cyclin A2 and Aurora Kinase A. The exact mechanism of action of SSA involves blockade of spliceosome assembly by retarding complex B formation. In brief, SSA prevents binding of U2 to pre-mRNA and destabilises its interaction with the BPS, resulting in failure in complex A to B transition, complex A accumulation and nuclear retention of semi-processed pre-mRNAs. Importantly, some of these pre-mRNAs can exit the nucleus, be translated to aberrant proteins with anti-tumour activity or high cytotoxicity leading eventually to cell death (Corrionero et al., 2011; Martinez-Montiel et al., 2018). Interestingly, SSA significantly reduced AR-V7 expression in CWR22Rv1 cells highlighting its potential clinical relevance in PC (Kawamura et al., 2019).

Use of antisense short RNA oligonucleotides (ASOs) has comprised an additional approach to target splicing and prevent specific splicing events. Mimicking siRNAs, ASOs are designed against specific regions in the target pre-mRNA. They act by masking pre-mRNA regions in close proximity to splice sites of interest and hence dictate exon inclusion or exclusion into mature mRNA transcripts. Likewise, splice switching oligonucleotides (SSOs) directed to *cis*-regulatory elements of the target pre-mRNA compete with SFs for binding to these sites ultimately abolishing interaction of the latter with their pre-mRNA targets. They promote splicing pattern shifts to favour production of splice variants which drive cellular processes such as apoptosis.

FDA approved antisense oligonucleotides have been used with remarkable success for treatment of muscular dystrophy disease (Martinez-Montiel et al., 2018). Despite application in other disease settings, ASOs/SSOS have not been exploited in cancer therapy yet (Lee and Abdel-Wahab, 2016). Nevertheless, the potential of their use to treat cancer is promising and of keen interest to oncologists. With regards to application of antisense RNA oligonucleotides as an alternative therapeutic strategy to antiandrogens in PC, a recent study of AR splicing in CRPC has suggested the use of a morpholino of 20 nucleotides which binds to a polyadenylation site at the 3'UTR downstream of AR CE3 and prevents recognition of the site by the polyadenylation factor CPSF1, leading to production of the full-length receptor by recognition of the consensus polyA signal at the 3' UTR downstream of terminal exon 8 (Van Etten et al., 2017). Importantly, targeting of ASOs against regulatory regions proximal to and

within CE3 of AR-V7 mRNA significantly suppressed AR-V7 mRNA synthesis and led to cell death *in vitro* (Luna Velez et al., 2019). Finally, SSOs have been used against the prostate-specific membrane antigen (PSMA) pre-mRNA to regulate the formation of specific mRNA isoforms in LNCaP cells (Williams, 2005).

ASO/SSO-pre-mRNA hybrids are prone to degradation by endogenous RNases, limiting their clinical potential. Efforts have focused on improving the stability of ASOs/SSOs to increase resistance to degradation by chemically modifying the oligonucleotides alongside developing methods and appropriate carrier molecules for efficient and safe *in vivo* delivery of ASOs/SSOs to target cells (Martinez-Montiel et al., 2018).

Further insight is required to fully elucidate the mechanisms that regulate splicing in order to develop efficient ways to target them for therapeutic purposes. The field is extremely promising and a vast amount of information is emerging every day to improve our understanding and fill knowledge gaps around pre-mRNA splicing.

## Chapter 2. Aims and Objectives

Resistance to AR-targeted therapies in PC is a major clinical problem and highlights that AR remains a valuable druggable target. Two key mechanisms of treatment resistance include the generation of AR mutants and AR splice variants (AR-Vs) that are refractory to anti-androgens and drive tumour progression, leading invariably to patient death. Our understanding of how AR mutants and AR-Vs function is limited due to difficulties in distinguishing their discriminate activities from full-length AR. This is in part due to the paucity of appropriate preclinical models which would allow us to study each aberrant receptor individually. Therefore, this study aimed to:

- Develop a CRISPR pipeline to engineer prostate cancer cell lines which harbour specific point mutations in the LBD of the AR gene.
- Generate AR-FL knock-out prostate cancer cell lines to study AR-V function in an AR-FL-free cellular background.
- Expand the application of CRISPR in the lab in order to study and elucidate AR-V7 splicing mechanisms, rather than just editing genomic sequences.

More specifically, this study focused on:

- Optimising CRISPR donor template knock-in efficiencies to achieve incorporation of AR point mutations in the genome of CWR22Rv1 cells.
- Generating an AR-FL knock-out CWR22Rv1 cell line by incorporating a STOP codon in AR exon 5 to abolish AR-FL expression and hence distinguish the activity of AR-Vs from AR-FL in an unbiased manner to ultimately identify *bona fide* AR-V regulated pathways.
- Pioneering CRISPR by coupling the use of a nuclease deficient Cas9 to downstream proteomic approaches such as mass spectrometry in order to identify splicing factors that mediate inclusion of CE3 in the mature AR mRNA transcript resulting in AR-V7 mRNA production.

## Chapter 3. Materials and Methods

### 3.1. Mammalian cell culture - Cell maintenance and passaging

HEK293T (ATCC® CRL-3216™), PC-3 (ATCC® CRL-1435™) and CWR22Rv1 (ATCC® CRL-2505™) cells were obtained from ATCC. The CWR22Rv1-AR-EK cell line is a CRISPR-engineered cell line derivative of the parental CWR22Rv1 cell line generated in-house. R1-D567 cells were a kind gift from Scott Dehm. Cells were routinely cultured in RPMI 1640 medium (Sigma) supplemented with 10% (v/v) foetal bovine serum (FBS) and 2 mM L-glutamine (Sigma) (hereafter called full media) and were kept in a humidified incubator (Sanyo) at 37°C and 5% CO<sub>2</sub> routinely.

Cell handling was carried out in a class II biosafety cabinet (BioMAT). Cells were passaged every 72 hours. Culture media were removed from the culture flask and cells were trypsinised in 10% (v/v) trypsin in PBS at 37°C for 3 min following a PBS wash. Cell suspension containing trypsin was neutralised with 3 volumes of full media and was spun at 250 x g for 5 min at room temperature using a Heraeus Multifuge X1 centrifuge (Thermo Fischer). Resultant supernatant was aspirated off and the remaining cell pellet was re-suspended in full media. Finally, ~1x10<sup>6</sup> cells were seeded in 175 cm<sup>2</sup> culture flasks (Corning).

All cell lines were tested routinely (every two months) for mycoplasma contamination using the MycoAlert™ Mycoplasma detection kit (Lonza). CWR22Rv1 and CWR22Rv1-AR-EK cells were STR profiled/authenticated prior to use using the GenePrint® 10 System (Promega) following manufacturer's instructions.

### 3.2. Cell lines

The R1-D567 cell line is a TALEN-engineered cell line derivative of the R1-AD1 monoclonal subline isolated from the CWR-R1 cell line. Genome editing resulted in deletion of AR exons 5, 6 and 7 to express the clinically-relevant ARv567es (exon skipping) splice variant.

The CWR22Rv1 cell line harbours a 35 kb tandem duplication of AR exon 3 which results in the generation of truncated AR isoforms due to the incorporation of cryptic exons downstream of exon 3. Cells also express AR-FL alongside a repertoire of AR-Vs which are more abundant than AR-FL at protein level.

The CWR22Rv1-AR-EK cell line is a CRISPR-edited cell line derivative of CWR22Rv1 cells generated in our lab. Knock-in of a TAA stop codon in AR exon 5 resulted in AR-FL knock-out. All nascent AR-Vs are still expressed in this new cell line derivative at comparable levels to the parental cell line.

The PC-3 cell line was established from a bone metastasis Caucasian prostate cancer patient. It is an easy-to-transfect, AR negative cell line.

The HEK293T cell line is a highly transfectable human embryonic kidney cell line derivative of HEK293 cells, stably transduced to express the SV40 T-antigen.

### **3.3. Compounds**

Enzalutamide is a second-generation anti-androgen currently used as standard therapy in the clinic. Enzalutamide was purchased from Selleckchem and was dissolved in neat dimethyl sulfoxide (DMSO, Sigma) to make a 30 mM stock solution, which was then aliquoted and stored at -80°C for up to six months. Cells were treated with 10  $\mu$ M for at least 24 hours to block AR-FL activity.

5 $\alpha$ -dihydrotestosterone (DHT) is a potent agonistic ligand of AR-FL. DHT was purchased from Sigma and was re-suspended in 20% ethanol to make a 10 mM stock solution. It was then aliquoted and stored at -80°C. Cells were treated with 10 nM DHT for 24 hours to stimulate AR-FL activity.

Scr7 is a potent inhibitor of ligase IV, a downstream component of the Non-Homologous End-Joining repair pathway. Scr7 was purchased from Selleckchem and was re-suspended in DMSO to make a 30 mM stock solution, which was then aliquoted and stored at -80°C for up to six months. Cells were treated with 30  $\mu$ M for 24 hour or as appropriate to block NHEJ.

Puromycin is a protein synthesis inhibitor used as an antibiotic in mammalian cell culture. Puromycin dihydrochloride was purchased from Sigma and was re-suspended in dH<sub>2</sub>O to make a 10 mg/ml stock solution. Aliquots were stored at -80°C. Puromycin selection was performed using a final concentration of 2  $\mu$ g/ml for at least 3 days.

### 3.4. Western Blotting (WB)

Cell lysates were derived from direct lysis of cells seeded in a well of a 6-well plate using 120  $\mu$ l SDS sample buffer (125 mM Tris-HCl pH 6.8, 5% SDS, 10% glycerol, 10%  $\beta$ -mercaptoethanol and 0.01% bromophenol blue), following a wash with pre-warmed PBS. Cell lysates were transferred to 1.5 ml Eppendorf tubes and were boiled for 5 min at 100°C prior to loading on a 10% polyacrylamide gel casted in-house (Table 1). 5  $\mu$ l of Spectra Multicolor Broad Range Protein Ladder (Thermo Fischer Scientific) were loaded per gel for band size estimation. Electrophoresis was carried out in a Mini-PROTEAN® Tetra Cell device/tank (BioRad) filled with sufficient volume of 10% resolving buffer (25 mM Tris, 190 mM glycine, 0.1% SDS). Cell lysates were resolved at 100 Volts using a BioRad power pack until sufficient band separation was achieved. 15% resolving gels were prepared for analysis of low molecular weight proteins such as histones (10 – 15 kDa).

**Table 1.** Ingredients of a 10% polyacrylamide gel.

| Ingredient                                   | Resolving gel (10%) | Stacking gel (5%) |
|--|---------------------|-------------------|
| Bis-acrylamide (37%) - Sigma                 | 3.33 ml             | 1.67 ml           |
| Buffer A (750 mM Tris-HCl, pH 8.8, 0.2% SDS) | 5 ml                | -                 |
| Buffer B (250 mM Tris-HCl, pH 6.8, 0.2% SDS) | -                   | 5 ml              |
| Ammonium persulfate (10%) - Sigma            | 100 $\mu$ l         | 100 $\mu$ l       |
| TEMED - Sigma                                | 18 $\mu$ l          | 12 $\mu$ l        |
| dH <sub>2</sub> O                            | 1.67 ml             | 3.33 ml           |
| Total volume                                 | 10 ml               | 10 ml             |

Following electrophoresis, resolved proteins were transferred onto Hybond ECL nitrocellulose membrane (GE Healthcare) using a Mini Trans-Blot Transfer tank (BioRad). Overnight transfer at 30 Volts or 1 hour transfer at 100 Volts was carried out in 10% transfer buffer (25 mM Tris, pH 8.3, 150 mM glycine, 10% methanol). The membrane was then removed from the transfer apparatus and was blocked in 5% non-fat skimmed milk (Marvel)/TBS solution for 1 hour at room temperature with agitation. Immunoblotting was carried out/the membrane was

probed using the appropriate primary antibody diluted in 5% milk/TBS solution either overnight at 4°C on a roller or for 1 hour at RT. Upon incubation the membrane was washed (3x) with TBS Tween 20 (TBST) to remove excess of primary antibody and was incubated with the appropriate HRP-conjugated secondary antibody (Dako) for 1 hour at room temperature on a roller. Finally, the membrane was washed 3 times with TBST and was incubated in ECL reagent (GE Healthcare) for ~2 min in the dark. Immunoblots were developed using Fuji medical X-ray film (FUJIFILM) and an automated MediPhot 937 developer. Blocking solutions, antibody dilutions, incubation conditions and exposure times were optimised for each antibody used in this study. Antibody information is outlined in Table 2.

**Table 2.** Incubation conditions, working dilutions, exposure times and supplier information of antibodies used in WB.

| <b>Antibody</b>            | <b>Clone</b>              | <b>Supplier</b>                               | <b>Dilution/<br/>Application</b> | <b>Incubation<br/>time</b> | <b>Species<br/>raised<br/>in</b> |
|----------------------------|---------------------------|---|----------------------------------|----------------------------|----------------------------------|
| <b>AR (N-terminal)</b>     | N-20<br>polyclonal        | Santa Cruz<br>Biotechnology<br>(discontinued) | 1:1000/WB                        | Overnight or<br>1h @ RT    | Rabbit                           |
| <b>AR (N-terminal)</b>     | 441<br>monoclonal         | Santa Cruz<br>Biotechnology                   | 1:1000/WB                        | Overnight                  | Mouse                            |
| <b>AR (C-terminal)</b>     | C-19<br>polyclonal        | Santa Cruz<br>Biotechnology<br>(discontinued) | 1:1000/WB                        | Overnight                  | Rabbit                           |
| <b>AR (N-terminal)</b>     | D6F12<br>monoclonal       | Cell Signaling                                | 2 µg/ChIP                        | Overnight                  | Rabbit                           |
| <b>AR (N-terminal)</b>     | G122-434<br>monoclonal    | BD Biosciences                                | 1:1000/WB                        | Overnight                  | Mouse                            |
| <b>AR (N-terminal)</b>     | Proprietary<br>polyclonal | Abcam   | 1:1000/WB                        | Overnight                  | Rabbit                           |
| <b>AR-V7</b>               | EPR15656<br>monoclonal    | Abcam   | 1:1000/WB                        | Overnight                  | Rabbit                           |
| <b>Histone H2B</b>         | ab1790<br>polyclonal      | Abcam   | 1:5000/WB                        | 1h @ RT                    | Rabbit                           |
| <b>Cas9</b>                | 7A9-3A3<br>monoclonal     | Abcam   | 1:1000/WB                        | 1h @ RT                    | Mouse                            |
| <b>RNA pol II (pSer-2)</b> | NA                        | Abcam   | 2 µg/ChIP                        | Overnight                  | Rabbit                           |
| <b>FLAG</b>                | M2<br>monoclonal          | Sigma   | 1:1000/WB                        | Overnight                  | Mouse                            |

|                                    |                       |                |            |                |        |
|------------------------------------|-----------------------|----------------|------------|----------------|--------|
| <b>ATM</b>                         | D2E2<br>monoclonal    | Cell Signaling | 1:1000/WB  | Overnight      | Rabbit |
| <b>pATM (S1981)</b>                | EP1890<br>monoclonal  | Abcam          | 1:2000/WB  | Overnight      | Rabbit |
| <b><math>\beta</math>-actin</b>    | AC-15<br>monoclonal   | Abcam          | 1:10000/WB | 30 min @<br>RT | Mouse  |
| <b><math>\alpha</math>-tubulin</b> | B-5-1-2<br>monoclonal | Sigma          | 1:4000/WB  | 1h @ RT        | Mouse  |

### 3.5. Agarose gel DNA electrophoresis

1-2% agarose gels were prepared in 0.5x TBE buffer (45 mM Tris-borate, 1 mM EDTA pH 8.0) to separate DNA fragments resulting from PCR amplification, restriction enzyme digestion and chromatin sonication. 2  $\mu$ l of a 100 bp and/or a 10 kb Hyperladder (Bioline) were loaded per gel for band size estimation. Tested DNA samples were diluted in 5x loading buffer (Bioline) prior to loading and were finally allowed to run at 70 Volts for ~1 hour or until sufficient separation was observed. Visualisation of bands was carried out using Gel Red nucleic acid stain (Biotium) at 1:10,000 dilution and a G:Box Chemi Gel Doc system (Syngene).

### 3.6. Immunofluorescence

$4 \times 10^3$  cells were seeded in each chamber of a chamber slide (Thermo Fischer Scientific). Cells were allowed to settle/adhere for 24 hours before being forward transfected with plasmid vectors (100-200 ng) or AR targeting siRNAs (25 nM). 48 hours post-transfection, media were removed from each chamber, cells were washed with PBS and fixed with 4% (v/v) paraformaldehyde (Alfa Aesar) for 30 min in the dark. Cells were then washed twice with PBS and were permeabilised in 0.1% Triton X-100/PBS for 15 min. Fresh 4% (w/v) BSA (Sigma) solution in PBS was used to block potential non-specific interactions. Following an 1 hour incubation in BSA solution at RT, cells were incubated with primary antibody at 1:1000 dilution in 4% BSA solution at 4°C overnight. The following day, cells were washed with PBS (3 washes in total) and were incubated with the appropriate AlexaFluor® secondary antibody (Life Technologies) at 1:1000 dilution for 1 hour at RT in the dark. Finally, secondary antibody solution was removed, cells were washed 3 times with PBS and stained with Vectrashield

mounting medium with DAPI (Vector Laboratories) prior to imaging using a Nikon TE2000 fluorescent microscope.

### **3.7. Gene expression analysis**

#### ***3.7.1. RNA extraction***

RNA extraction was carried out using Trizol reagent (Invitrogen). In brief, cells were washed with PBS and directly lysed in 500  $\mu$ l Trizol/well of a 6-well plate. Cell lysates were transferred to 1.5 ml Eppendorf tubes and were vigorously mixed with 100  $\mu$ l chloroform (Sigma). Samples were then allowed to separate for 3 min at RT and were subsequently spun at 12,000 x g at 4°C for 15 min using a standard benchtop centrifuge. Upon centrifugation, the aqueous phase/layer was transferred to a new Eppendorf tube and was mixed with one volume of isopropanol. Samples were then stored at -20°C overnight to allow for efficient RNA precipitation to occur. The following day, samples were spun at 12,000 x g for 10 minutes at 4°C. Resultant RNA pellets were washed twice with 70% ethanol before being air-dried for 15 min and finally re-suspended in 30  $\mu$ l UltraPure™ RNase free dH<sub>2</sub>O (Life Technologies). RNA purity and concentration were measured using NanoDrop 2000 (Thermo Fischer) before RNA samples were stored at -80°C.

#### ***3.7.2. Reverse transcription (RT)***

cDNA synthesis was performed using the M-MLV Reverse Transcription kit (Promega) according to manufacturer's instructions. Briefly, 1  $\mu$ g of RNA was subject to cDNA conversion. The appropriate volume of RNA was diluted in RNase free dH<sub>2</sub>O to a final volume of 12.7  $\mu$ l. RNA samples were incubated at 65°C for 5 min to eliminate potential secondary structures and were immediately placed on ice upon incubation. In a separate tube, an RT mix was prepared as shown in Table 3. 7.3  $\mu$ l of RT mix were added to each RNA sample (final reaction volume of 20  $\mu$ l). cDNA synthesis was carried out at 37°C for 1 hour and was followed by an RTase inactivation step at 100°C for 5 min. Resultant cDNA was diluted 1:20 in RNase free dH<sub>2</sub>O prior to use in qPCR. cDNA samples were stored at -20°C.

**Table 3.** Reverse transcription cocktail ingredients and volumes used per reaction.

| Reagent                             | Volume/reaction ( $\mu$ l) |
|-------------------------------------|----------------------------|
| M-MLV Reaction buffer (5x)          | 4                          |
| M-MLV Reverse Transcriptase (RTase) | 0.3                        |
| dNTPs (4 mM each)                   | 2                          |
| oligodT (100 $\mu$ g/ml)            | 1                          |
| Total volume                        | 20                         |

### 3.7.3. Quantitative real-time PCR (qPCR)

Gene expression analysis was performed by qPCR using the Platinum SYBR Green qPCR SuperMix (Invitrogen) following manufacturer's instructions. A qPCR mastermix was prepared for each target/gene of interest as shown in Table 4. Each reaction was performed in triplicate and was loaded onto a 384-well plate (Thermo Fischer). Plates were then sealed with a MicroAmp optical adhesive film (Thermo Fisher) and were briefly spun using a mini plate spinner (MPS 1000, Labnet) prior to qPCR analysis. The following thermal profile was performed on a QuantStudio 7 Flex Real-Time PCR platform (Applied Biosystems): 50°C for 2 min (UNG activation), 95°C for 10 min (initial denaturation), 95°C for 15 sec and 60°C for 1 min (40 cycles) followed by melt curve analysis to assess non-specific amplification. Raw data was analysed using the QuantStudio Real-Time PCR software (Applied Biosystems). The standard curve quantification method was performed. Standard curve samples were prepared by serial dilutions of the siSCR or DMSO treated cDNA sample depending on the experimental conditions. All quantities were hence calculated relative to the quantity assigned to the standard curve and were normalised to HPRT1. Primer sequences are shown in Table 5.

**Table 4.** qPCR mastermix ingredients and volumes used per well/reaction.

| Ingredient                            | Volume/well ( $\mu$ l) |
|---------------------------------------|------------------------|
| SYBR Green qPCR SuperMix (5x)         | 5                      |
| Primer mix (F/R: 25 ng/ $\mu$ l each) | 0.8                    |

|                    |     |
|--------------------|-----|
| dH <sub>2</sub> O  | 1.2 |
| cDNA sample (1:20) | 3   |
| Total volume       | 10  |

**Table 5.** Sequences of primers used in qPCR.

| <b>Oligo Name</b> | <b>Forward Sequence (5'→3')</b> | <b>Reverse Sequence (5'→3')</b> |
|-------------------|---------------------------------|---------------------------------|
| HPRT1 mRNA        | TTGCTTTCCTTGGTCAGGCA            | AGCTTGCGACCTTGACCATCT           |
| UBE2C mRNA        | TGCCCTGTATGATGTCAGGA            | GGGACTATCAATGTTGGGTTCT          |
| PSA mRNA          | GCAGCATTGAACCAGAGGAG            | AGAACTGGGGAGGCTTGAG             |
| CCNA2 mRNA        | GAAGACGAGACGGGTTGCA             | AGGAGGAACGGTGACATGCT            |
| KLK2 F mRNA       | AGCATCGAACCAGAGGAGTTCT          | TGGAGGCTCACACACCTGAAGA          |
| ATAD2 mRNA        | TGGCACCAGCTGTCATTCAT            | AGCTTCACGAATCACCTGGG            |
| FKBP5 mRNA        | CCCCCTATTTTAATCGGAGTAC          | TTTTGAAGAGCACAGAACACCCT         |
| TMPRSS2 mRNA      | CTGCTGGATTTCCGGGTG              | TTCTGAGGTCTTCCCTTTCTCCT         |
| FL-AR mRNA        | AACAGAAGTACCTGTGCGCC            | TTCAGATTACCAAGTTTCTTCAG         |
| AR exon 3 mRNA    | AACAGAAGTACCTGTGCGCC            | -                               |
| AR-V1 mRNA        | -                               | TGAGACTCCAAACACCCTCA            |
| AR-V3 mRNA        | AGACGAAGCTTCTGGGTGT             | CATGCAGTATGGCTTGGG              |
| AR-V5 mRNA        | -                               | CAAAGAATTGTGGGTAGGAAGC          |
| AR-V7 mRNA        | -                               | TCAGGGTCTGGTCATTTTGA            |
| AR-V9 mRNA        | -                               | GCAAATGTCTCCAAAAAGCAGC          |
| DMC1 mRNA         | AGGTGCCAATGGTTATACCG            | TTGAAGACACCTGGCTCCTC            |
| XRCC2 mRNA        | TCACCTGTGCATGGTGATATT           | TTCCAGGCCACCTTCTGATT            |

|               |                         |                          |
|---------------|-------------------------|--------------------------|
| RMI2 mRNA     | GGCAGGGTAGTGATGGCGGAC   | CCTGAACCACTCCCATCACCAT   |
| BRCA1 mRNA    | CTGAAGACTGCTCAGGGCTATC  | AGGGTAGCTGTTAGAAGGCTGG   |
| RAD51AP1 mRNA | CTTCTGGAAGGCAGTGATGGTG  | AGAGAAGTCTTCGTCATTATCCTC |
| RAD54L mRNA   | CCCTTTCTTCCATCACCTCGCT  | GCCTTAGAGCTGTAACCAGGAG   |
| CHEK1 mRNA    | GTGTCAGAGTCTCCAGTGGAT   | GTTCTGGCTGAGAACTGGAGTAC  |
| EXO1 mRNA     | TCGGATCTCCTAGCTTTTGGCTG | AGCTGTCTGCACATTCTAGCC    |
| NBN1 mRNA     | TCTGTCAGGACGGCAGGAAAGA  | CACCTCAAAGACAACCTGCGGA   |
| RAD54B mRNA   | GGTGTGTCCAAGCTCTTAGCG   | AGCATATCCATGACGCTTACATAC |
| RAD51C mRNA   | GTGAAACCCTCCGAGCTTAGCA  | CCTGCTCAAGAAGTTCCAGTGC   |
| ABCF2 mRNA    | GAGGTTTCACTGGGAGCAAGATC | CTGTAGCGTCTTCTCCTTGCTC   |
| CLSPN mRNA    | AAGGAGCGAATTGAACGAG     | TCTGCAGTGCTTTGGCTG       |
| PCNA mRNA     | GCCATATTGGAGATGCTGT     | TGAGTGTACCGTTGAAGA       |
| BRCA2 mRNA    | GGCTTCAAAAAGCACTCCAGATG | GGATTCTGTATCTCTTGACGTTCC |
| RAD21 mRNA    | TCCCCAGAGGAGCCTCCAA     | AGCAAGAGCTCGCTGGAGACCA   |
| Ku70 mRNA     | TGGCTGTGGTGTTCTATGGT    | TGAGTAGTCAGATCCGTGGC     |
| Ku80 mRNA     | ATCAGAACATCACAGTGCACAG  | AATCACATCCATGCTCACGA     |
| DNA-PKcs mRNA | TGTCCGGAAGTCACTCAACA    | ACTTAATAAGAAGGTCCAGGGCT  |

### 3.8. Plasmids

The Cas9-encoding plasmid vectors px459 V2.0 (plasmid #62988), lentiCRISPR V2.0 (plasmid #52961), 3xFLAG-dCas9/pCMV-7.1 (plasmid #47048) and pCS2+Cas9-mSA (plasmid #103882) were purchased from Addgene. The pLV-U6g-EPCG vector was purchased from Sigma. The lentiCRISPR-iCer vector for gRNA expression was a kind gift from Prof. Olaf Heidenreich. The

px459/Cas9-mSA fusion vector was generated in-house (cloning strategy outlined in section 4.3.9).

### **3.9. Bacterial transformation**

Competent 5-alpha *E. coli* cells (NEB) were thawed on ice. 25 µl of cell suspension were gently mixed with 10-100 ng of plasmid and cells were placed on ice for 30 min prior to a heat-shock step at 42°C for 30 sec on a heat block. Cells were then returned on ice for 2 min before adding 500 µl of pre-warmed SOC outgrowth medium (NEB). Transformed cells were incubated for 1 hour at 37°C at 220 rpm, were then spread onto LB agar plates which contained the appropriate antibiotic (either 100 µg/ml ampicillin or 50 µg/ml kanamycin depending on the vector backbone) and were finally allowed to grow at 37°C overnight (~16h). The following day, single clones were transferred to 5 ml antibiotic-containing LB (starter culture) and were cultured at 37°C at 220 rpm for ~ 5 hours before being transferred to conical flasks containing 200 ml LB supplemented with the appropriate antibiotic. Bacteria were cultured for an extra 16 hours at 37°C at 220 rpm and were finally pelleted by centrifugation at 2000 x g for 5 min prior to plasmid extraction. Lentiviral vectors were introduced into One Shot® Stbl3™ Chemically Competent *E. coli* cells (Invitrogen) as described above. When bacterial stubs were purchased, individual clones were isolated by streaking bacteria from stubs onto LB agar plates. Overnight incubation (up to 16 hours) at 37°C was performed followed by cultivation in LB as described above.

### **3.10. Plasmid extraction**

Highly concentrated plasmids were extracted/prepped/purified from bacterial pellets using the PureLink HiPure Plasmid Maxiprep Kit (Thermo Fischer) according to manufacturer's instructions. In brief, cell pellets were re-suspended in 10 ml R3 buffer with RNase A and subsequently lysed in 10 ml L7 buffer for 5 min at RT. Following lysis, chromosomal DNA and cell debris were precipitated using 10 ml N3 buffer. The cell lysate was then loaded onto a HiPure spin column and was allowed to flow through the column by gravity. Column-captured plasmid DNA was then washed twice with W8 wash buffer and was eluted in 5 ml E4 buffer. Plasmid DNA was then precipitated by adding 3.5 ml isopropanol and subsequently spun at 12,000 x g for 30 min at 4°C. The resulting pellet was washed with 70% ethanol, air-dried for

~10 min and re-suspended in 500 µl DNase free dH<sub>2</sub>O. Plasmid concentration and purity was measured using Nanodrop 2000. Plasmid miniprep was performed using the PureLink™ HiPure Plasmid Miniprep Kit (Thermo Fischer) in a similar manner to the counterpart maxiprep kit. Volumes were adjusted to fit the HiPure miniprep spin column.

### 3.11. Plasmid transfections

Plasmid vectors were transfected into 2.5x10<sup>5</sup> cells seeded into a well of a 6-well plate 24 hours before transfection. Transfections were performed using 1-6 µg of plasmid vector and TransIT®-LT1 transfection reagent (Mirus) at a 1:3 ratio. Transfection mixes were prepared in 100 µl pre-warmed basal RPMI 1640 medium and were incubated for 30 min at RT before being added drop-by-drop to pre-plated cells (forward transfection) in the appropriate well.

### 3.12. siRNA transfections

All siRNA oligos were ordered from Sigma in lyophilised form and were re-suspended in sterile RNase/DNase free dH<sub>2</sub>O to a final concentration of 50 µM before being aliquoted and stored at -80°C. 2.5x10<sup>5</sup> cells grown in phenol red free steroid depleted RPMI 1640 medium were either reverse or forward transfected with 25 nM siRNA using Lipofectamine® RNAiMax transfection reagent (Thermo Fischer) at a 1:2 ratio. Transfection mixes were prepared in 100 µl pre-warmed basal RPMI 1640 medium and were incubated at RT for 30 min before being added drop-by-drop to the appropriate well. See siRNA sequences in Table 6.

**Table 6.** Control and AR-targeting siRNA sequences.

| Oligo     | Sequence (5'→3')    |
|-----------|---------------------|
| siScr     | UUCUCCGAACGUGUCACGU |
| siAR-V    | GUAGUUGUGAGUAUCAUGA |
| siARexon1 | CAAGGGAGGUUACACCAAA |
| siARexon4 | CCAUCUUUCUGAAUGUCCU |
| siARexon7 | GGAACUCGAUCGUAUCAUU |

### **3.13. Immunoprecipitation (IP)**

2x10<sup>6</sup> cells were plated in 90 mm dishes. Cells were harvested/lysed in 1 ml IP lysis buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 0.2 mM Na<sub>3</sub>VO<sub>4</sub>, 0.5% NP-40, 1 mM phenylmethylsulfonyl fluoride, 1 mM dithiothreitol, 1 cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail tablet (Roche) and samples were incubated for 30 min on ice with brief vortexing every 10 min. Lysates were spun at 12,000 x g for 5 min at 4°C to remove cell debris and supernatants were pre-cleared with 20 µl Protein G-Sepharose (PGS) beads (GE Healthcare) for 2 hours at 4°C with rotation to prevent non-specific interactions. PGS beads were pre-washed 3 times in IP lysis buffer prior to use. Following pre-clearing, PGS beads were removed by centrifugation at 12,000 x g for 5 min at 4°C. Resulting supernatants were equally split in two; for antibody and control IgG IPs. 1-2 µg of the appropriate antibody were added to each sample and binding was allowed to occur at 4°C overnight with rotation. The following day, resulting bead conjugates were spun at 12,000 x g for 5 min at 4°C and were washed twice with Wash buffer A (PBS, 0.2% Triton X-100, 350 mM NaCl) and once with Wash buffer B (PBS, 0.2% Triton X-100) before being re-suspended in SDS sample buffer and finally boiled at 100°C for 5 min prior to gel loading for WB analysis as described in section 1.4.

### **3.14. Chromatin Immunoprecipitation (ChIP)**

5x10<sup>6</sup> cells were plated in 150 mm dishes (Corning) in phenol-red free steroid-depleted RPMI 1640 medium and were allowed to grow for 24 hours prior to siRNA or plasmid transfections. The Schmidt *et al.* protocol was followed with the following adjustments/amendments:

#### **3.14.1. Fixation**

48 hours post-transfection cells were fixed with 1% formaldehyde (Sigma) for 7 min at RT with occasional shaking. Formaldehyde was then quenched with 125 mM Glycine (Fischer) and subsequent incubation for 5 min at RT with occasional shaking. Media were aspirated off, cells were washed twice with 10 ml ice cold PBS, scrapped off the dish and finally transferred to pre-chilled 15 ml falcon tubes. Cells were then spun at 2000 x g for 5 min at 4°C. Supernatants were removed, resulting cell pellets were snap frozen in liquid nitrogen and stored at -80°C until being used in downstream steps.

### **3.14.2. Cell lysis and chromatin fractionation**

Cell fractionation was carried out by re-suspending the cell pellets in 10 ml LB1 (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP40, 0.25% Triton X-100) and subsequent incubation on ice for 10 min with shaking. Resulting lysates were spun at 2000 x g for 5 min at 4°C. Supernatant was removed, pellets were re-suspended in 10 ml LB2 (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA) and resulting lysates were incubated on ice for 5 min with shaking. Upon centrifugation at 2000 x g for 5 min at 4°C, resulting nuclei were re-suspended in 500 µl LB3 (100 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-deoxycholate, 0.5% N-lauroylsarcosine) to extract chromatin. Chromatin was sheered down to ~100-200 bp fragments using a Bioruptor® sonication device (Diagenode) and performing 2 x 15 min cycles (30 sec on/off intervals) on High setting. Samples were spun at 12,000 x g for 5 min at 4°C to remove cell debris. Supernatants (chromatin fraction) were transferred to 1.5 ml Eppendorf tubes and chromatin concentration was measured using NanoDrop 2000.

### **3.14.3. Chromatin Immunoprecipitation**

40 µl of Dynabeads Protein A (Invitrogen) per IP were aliquoted into a 1.5 ml Eppendorf tube. Dynabeads were washed twice with syringe filtered ice cold 0.5% (w/v) BSA/PBS solution. 2 µg of the appropriate antibody were then added to Dynabeads previously re-suspended in 700 µl 0.5% BSA/PBS solution and antibody-Dynabeads conjugation was allowed to occur for at least 6 hours at 4°C with rotation. Conjugates were then separated from the solution using a magnetic rack (Life Technologies). 70 µg of sonicated chromatin were diluted in 700 µl LB3 containing 1% Triton X-100. 70 µl of each chromatin sample were aliquoted and stored at -80°C to be used as input samples. Beads were finally re-suspended in 630 µl of the remaining chromatin solution and IP was carried out overnight at 4°C with rotation.

### **3.14.4. Elution and reverse cross-linking**

Following overnight incubation, resulting complexes were thoroughly washed with RIPA buffer (50 mM HEPES-KOH pH 7.5, 500 mM LiCl, 1 mM EDTA, 1% NP40, 0.7% Na-deoxycholate) at 4°C, re-suspended in ice cold TBS and spun at 3000 x g for 5 min at 4°C. TBS was removed using a magnetic rack to capture the beads and the latter were re-suspended in 200 µl ChIP elution buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS). Input samples, prepared in the

previous step were thawed and also diluted to 200 µl in ChIP elution buffer. All samples were incubated at 65°C overnight with mixing by pipetting every 5 min for the first 15 min. Following incubation, beads were removed using a magnetic rack and samples were diluted in 200 µl TE buffer (10 mM Tris, pH 8.0, 0.1 mM EDTA). RNase A (Thermo Fischer) treatment for 30 min at 37°C was followed by Proteinase K (Qiagen) treatment at 65°C for 2 hours.

### **3.14.5. Chromatin extraction**

Eluted chromatin was extracted using the GenElute™ Mammalian Genomic DNA Miniprep Kit (Sigma) following manufacturer's instructions. Briefly, samples were diluted in one volume of Lysis Solution C followed by 200 µl absolute ethanol, vortexed to homogenize, loaded on a GenElute spin column, which was previously equilibrated with 500 µl Column preparation Solution and spun at 8,000 x g for 1 min. Column-captured/bound chromatin was then washed twice with 500 µl Wash Solution and centrifuged at full speed for 5 min to remove remaining ethanol. Chromatin was finally eluted in 130 µl RNase/DNase free dH<sub>2</sub>O and was used in qPCR to assess recruitment of immunoprecipitated proteins to specific chromatin regions. ChIP primer sequences are outlined in Table 7.

**Table 7.** Primers used in qPCR to amplify chromatin regions of interest following ChIP.

| <b>Oligo Name</b>  | <b>Forward Sequence (5'→3')</b> | <b>Reverse Sequence (5'→3')</b> |
|--------------------|---------------------------------|---------------------------------|
| PSA Enh (ChIP)     | TGGGACAACCTGCAAACCTG            | CCAGAGTAGGTCTGTTTTCAATCCA       |
| PSA Prom (ChIP)    | CCTAGATGAAGTCTCCATTGAGCTACA     | GGGAGGGAGAGCTAGCACTTG           |
| CCNA2 (ChIP)       | TTAGTGAGCTGTCCAGTGA CTCAAT      | CCCATGTATTAAAGTAGCTTCTGTAAACA   |
| KLK2 Prom (ChIP)   | ACCCCTGTTGCTGTTCATCCTG          | CCGCCCTTGCCCTGTTGG              |
| TMPRSS2 Enh (ChIP) | TGGTCCTGGATGATAAAAAAAGTT        | GACATACGCCCCACAACAGA            |
| UBE2C Enh (ChIP)   | TGCCTCTGAGTAGGAACAGGTAAGT       | TGCTTTTTCCATCATGGCAG            |

### **3.15. Cell proliferation assays**

#### ***3.15.1. Sulforhodamine B (SRB) assay***

4x10<sup>3</sup> cells per well were seeded into 96-well plates (Corning). Cells were allowed to settle/adhere for 24 hours before being transfected with 25 nM of the appropriate siRNA. Two rounds of siRNA transfection were carried out post-seeding; first round at t=24 hours - second round at t=72 hours. At end-point, cells were fixed with ice cold trichloroacetic acid (TCA) to a final concentration of 10% (v/v) for 1 hour at 4°C. Upon fixation, medium was aspirated off and cells were washed with tap H<sub>2</sub>O and left to air-dry. Cells were then stained with 0.4% (v/v) SRB solution made up in 1% (v/v) glacial acetic acid for 30 min. Once stained, excess SRB was washed off by submerging the plates 5-6 times in 1% glacial acetic acid and were left to air-dry. Finally, bound SRB was dissolved in 100 µl of 10 mM Tris pH 10.8 for 15 min with shaking. Absorbance was measured at 570 nm using a plate reader (BioRad).

#### ***3.15.2. Colony formation assay***

Cells were transiently transfected with either control or AR-targeting siRNAs for 48 hour (as described in section X) prior to re-seeding at densities of 500 and 1000 cells/well in 6-well plates (Corning) for two weeks. At end-point, media were aspirated off, colonies were washed with PBS, fixed with 10% neutral buffered formalin solution (Sigma) for 30 min at RT, washed again with PBS and finally stained using 0.01% (w/v) crystal violet for 30 min before being counted manually.

## Chapter 4. CRISPR knock-in pipeline for novel PC cell line development

### 4.1. Introduction

CRISPR is a rapidly evolving technology which is currently being utilised by a significantly increasing number of research groups around the globe to generate appropriate models to study inherited human diseases. The generation of preclinical mouse or cell line models which harbour disease-specific CRISPR-induced genomic alterations is currently of high interest and public demand. It offers researchers the opportunity to study the background of a certain human disease in more biologically relevant models to eventually improve understanding of human disease and allow the development of more efficient gene therapy strategies (Rossidis et al., 2018; Villiger et al., 2018). In addition, the technology is easy to design and apply compared to previous editing techniques, such as TALEN and zinc-finger approaches; shows high accuracy and is extremely cost-effective, making it accessible to many research groups (Sander and Joung, 2014).

Gene knock-out studies allow us to interrogate the function of the impacted genes and subsequently explore interaction networks and signalling pathways which were previously unknown or we had limited knowledge about (Ihry et al., 2019). In particular, CRISPR genome-wide screens are currently extremely popular across a host of different disease settings which primarily aim to identify and characterise disease drivers and aetiology (Fei et al., 2017).

A number of different genetic disorders are caused by a single base substitution in the DNA sequence of an individual. Correction of these aberrant nucleotide switches using CRISPR editing could result in the reversal of the pathologic phenotype back to wild-type, restore gene function and potentially allow cure of the genetic disease leading to a real breakthrough in the field of biomedicine. Gene therapy is hence much more appealing now and accumulating efforts are being made to correct genes responsible for life-threatening disorders (Reinhardt et al., 2013). For instance, sickle cell anaemia is a typical example of a disorder caused by a point mutation (A to T) in the gene which encodes  $\beta$ -globin in adults. Correction of the point mutation back to the wild-type counterpart in the stem cells of the patient could permanently "fix" the mutated gene and effect a long-term cure of such a common haematological disorder (LiuHong Cai et al., 2018).

Duchenne muscular dystrophy is one more example of a life-threatening genetic disease which could be tackled by CRISPR editing. A deletion caused by aberrant exon skipping leads to loss of dystrophin, the protein responsible for muscle integrity and function in vertebrates. Scientists have attempted to correct the structural aberration of the gene, which encodes dystrophin, by introducing the exons which have been skipped and restore expression of dystrophin in mouse models, stem cells and cell line models. Preclinical models show promising results so far and potentiate the clinical use of CRISPR in humans (Long et al., 2018). Such examples highlight the massive potential of CRISPR engineering in life sciences while paving the way for more pioneer applications to treat human disease.

CRISPR knock-in strategies have been performed in the above examples. CRISPR knock-in relies on the supply of an exogenous DNA sequence, known as donor template which encompasses the desired wild-type nucleotide for correcting point mutations or DNA fragment for correcting larger DNA lesions. Upon DNA cleavage by Cas9, repair mechanisms are triggered to fix the Cas9 induced DSB. In the presence of a donor template which shares homology with the targeted gene sequence, HDR mediates incorporation of the homologous DNA fragment in the host genome, allowing permanent and inherited genetic changes. This process is technically challenging as it occurs infrequently depending on factors such as cell type, prevalence of the alternative DNA repair mechanism NHEJ and transfection efficiency of the host cell line (Paquet et al., 2016).

The technology has evolved rapidly since its emergence and first application. Multiple different approaches have been developed to improve CRISPR knock-in efficiencies to ultimately introduce and aid study or permanently correct pathogenic mutations. In particular, efforts to increase the rate of HDR to enable insertion of the desired DNA template is of major importance as this remains a major limitation to the editing process; with steady-state HDR efficiencies ranging from 1-4% (Liu et al., 2019). Normally a couple of thousand clones (!) would be screened manually in order to find a few which will contain the desired modification (Merkle et al., 2015). NHEJ is the most prominent repair mechanism which is activated upon DSB. It is error-prone and generates imprecise edits. It finally competes and normally occurs a lot more frequently than HDR as it requires less energy, hence limiting knock-in rates. NHEJ often occurs simultaneously with HDR and generates gene knock-outs by introducing indels, frame shifts, premature stop codons to disrupt gene function which is

the basis of CRISPR knockout screening. (Merkle et al., 2015; Paquet et al., 2016). Various approaches have been suggested to prevent NHEJ mediated DNA repair. Pharmacological inhibition of NHEJ has been attempted by targeting different components of the pathway, such as DNA-PK and ligase IV, and an overall increase in HDR rate is claimed by the groups which conducted the studies (Maruyama et al., 2015; Robert et al., 2015). On the contrary, overexpression of key components of the HDR machinery, such as BRCA1, may shift the equilibrium of DNA repair towards the precise repair pathway of homologous recombination at the expense of NHEJ (Pinder et al., 2015). Furthermore, the rate of HDR varies significantly depending on the phase of the cell cycle; it is more prominent in S and G2 phases. This observation was exploited by Lin *et al.* who utilised a cocktail of cell cycle modulators to synchronise the cells at G2 phase and seemed to dramatically increase the HDR efficiency by almost 50% (Lin et al., 2014).

A plethora of different designs have been attempted to date and have focused on improving donor template properties. Essentially, symmetry and polarity of the template seem to affect knock-in rate (Paix et al., 2017). More specifically, it has been shown that sense ssODNs burying the desired mutation in the centre of the sequence (symmetric ssODN) are incorporated more efficiently in the host genome than their antisense, asymmetric counterparts. Additionally, it was believed that integration rates of the repair template are very much dependent on size. Studies have shown that shorter donor templates are integrated easier than longer ones (Paix et al., 2017). In contrast to this notion, knock-in experiments have been performed with success when either longer (up to kb) or shorter DNA fragments were used as donor templates (Ruan et al., 2015). However, most groups now utilise single stranded oligos (ssODNs) of around 150 bp with homology arms of up to 60 bp, since use of ssODNs with longer homology arms (80-90 bp) was associated with a drop in knock-in rates (Okamoto et al., 2019).

It has also been suggested that the editing complex, consisting of Cas9 and gRNA, improves overall editing efficiencies when it is provided to the cells as a purified ribonucleoprotein (RNP) complex (Kim et al., 2014; Liang et al., 2015). Importantly, direct comparison between Cas9 plasmid-derived expression and Cas9 RNP indicated superiority of the latter in knock-in rates for precise gene editing (Okamoto et al., 2019). However, this approach does not allow researchers to select the potentially edited cells since it does not encompass any antibiotic or

fluorescent selection marker which are frequently contained in Cas9-expressing mammalian plasmid vectors, limiting their use in cases that cell transfection efficiency is low.

One of the limiting factors when one performs knock-in experiments is the availability of the ssODN at the site of the Cas9 induced DSB. A recent approach relies on the high affinity between streptavidin and biotin. More specifically, covalent tagging of Cas9 with monomeric Streptavidin (mSA) and biotinylation of the ssODN enables strong coupling of Streptavidin with biotin bringing the ssODN in close proximity to the DSB to increase the concentration of the ssODN at the site of the DSB (Ma et al., 2017). This mechanism enhances the rate of ssODN knock-in to the target region by up to 90% and is gaining more attention by the scientific community as notably high knock-in efficiencies have been reported when the approach was applied in mice and mammalian cancer cells (Roche et al., 2018)

Despite its overwhelming use and numerous advantages it may show, CRISPR editing has its drawbacks too. Off-target effects are one of the major issues to consider when a CRISPR project is designed (Zhang et al., 2015). Rational design of the gRNA sequences is required to minimise and/or eliminate potential Cas9/gRNA off-targeting to partially homologous genomic regions (Doench et al., 2016). Numerous bioinformatics tools have been developed to *in silico* predict off-target sites and facilitate on-target gRNA design. However, not all off-target sites are predicted. There is a significant number of experimental studies which focus on interrogating the specificity of Cas9 and the spectrum of its off-target activity in different experimental set-ups, and highlight the factors that dictate specificity (outlined below).

Imperfect base complementarity can be tolerated by Cas9. The number of mismatches and the position of them across the 20mer gRNA sequence are critical factors which determine specificity. Genome-wide studies have consistently shown that single base mismatches are being bypassed by Cas9 leading to off-targeting. Two mismatches are still well-tolerated but reduce cleavage efficiency as opposed to three mismatches (and beyond) which eliminate off-target cleavage in most cases. (Multiple) mismatches - either dispersed or attached to one another - have a more dramatic impact on complementarity when they are located in the PAM proximal region (12 nucleotide region upstream of the PAM) (Anderson et al., 2015; Hsu et al., 2013). Cas9 still binds to off-target sites, however it does not exert its function since extensive pairing is required for Cas9 to acquire the appropriate 3D conformation to dock onto target DNA and cleave it (Sternberg et al., 2015).

As mentioned above, CRISPR technology has been used for the generation of cell line models which carry disease-associated mutations, not just CRISPR corrected genes for gene therapy studies. In CRPC, AR remains active as a result of the acquired mutations in its LBD following ADT. AR mutants are not responsive to next-generation anti-androgens and therefore contribute to tumour progression to more advanced and ultimately lethal forms (Brooke and Bevan, 2009, Lu *et al.*, 2015). In particular, AR inhibition by bicalutamide can be overcome by the acquisition of a gain-of-function mutation in position 741 in the LBD of AR where tryptophan is substituted to either leucine or cysteine.

To date, analysis of CRPC-relevant AR mutants has been largely restricted to overexpression studies in AR null cell lines (Brooke and Bevan, 2009) and thus our knowledge of the physiological function of these proteins is limited. Previously, the host lab have developed a replacement strategy that enables stable, ectopic expression of specific AR mutants in the background of LNCaP cells depleted of endogenous AR<sub>T877A</sub> to examine the activity of AR mutants in a more physiologically-relevant model, as LNCaP cells are dependent upon the AR signalling cascade for growth (O'Neill *et al.*, 2015). Although this approach has been useful in demonstrating distinct functionality of the studied AR mutants, such as the W741L mutant, it presented two major drawbacks: (i) efficiency of endogenous AR<sub>T877A</sub> depletion by siRNA knockdown was variable between experiments; and (ii) reduction of cellular AR<sub>T877A</sub> was not 100% efficient and hence interference by endogenous receptor was likely to impact on the overarching findings of the system. Therefore, a CRISPR-based genome editing approach was seen as the most appropriate strategy to revert the T877A mutation to wild-type and generate the clinically relevant AR<sub>W741L</sub> mutant in LNCaP cells. This genome editing approach will enable unbiased means for assessing global AR mutant regulation in order to identify genes that show discriminate upregulation in response to anti-androgens, such as bicalutamide and abiraterone, respectively.

The CWR22Rv1 cell line will also be engineered to generate and study the AR W741L mutant. Of note, CWR22Rv1 cells have a single copy of the AR gene which harbours a duplication of exon 3 whilst the H874Y mutant is expressed endogenously. The contribution of these mutations to CRPC will be independently studied, upon reversal of the endogenous H874Y to wild-type, and in concert with the AR splice variants that are endogenously expressed in the CWR22Rv1 cells. Interestingly, the AR-V7 splice variant, alongside the H874Y mutation, have

already been isolated from a patient's circulating PC cells (Steinestel et al., 2015) suggesting that AR splice variants and mutants may synergistically regulate transforming mechanisms of AR activity in CRPC. Identification of shared effectors may provide the opportunity for designing pharmacological agents against them to potentially benefit patients who harbour both AR splice variants and AR mutations.

## 4.2. Specific Materials & Methods

### 4.2.1. CRISPR complex components: CRISPR vectors – synthetic gRNAs

The all-in-one pLenti CRISPR Cas9/gRNA vectors (pLV-U6g-EPCG) were custom-designed by Sigma-Aldrich. Two distinct single-guide RNAs (sgRNAs) were designed against AR exon 5 (Table 8). One shot Stbl3 chemically competent *E.coli* cells (Invitrogen) were transformed with the two CRISPR vectors, and constructs were maxiprepmed using the PureLink HiPure Plasmid Maxiprep Kit (Invitrogen) as described in Chapter X. An additional gRNA/crRNA targeting AR exon 5 was synthesised (Sigma) and cloned into the pSpCas9(BB)-2A-Puro (px459) vector (Addgene) following the "LentiCRISPRv2 and lentiGuide-Puro: lentiviral CRISPR/Cas9 and single guide RNA" protocol as described by Ran *et al.* (Ran et al., 2013).

**Table 8.** Sequences of gRNAs used to target AR exon 5.

| gRNA           | Sequence (5'→3')     | Cloned in                                       | PAM |
|----------------|----------------------|---|-----|
| AR_049/gRNA_1  | GGCTTCCGCAACTTACACG  | pLV-U6g-EPCG                                    | TGG |
| AR_061/gRNA_2  | TTACACGTGGACGACCAGA  | pLV-U6g-EPCG                                    | TGG |
| crRNA_1/gRNA_3 | TGGCTGTCATTCACTACTCC | pSpCas9(BB)-2A-Puro<br>&<br>pSpCas9-mSA-2A-Puro | TGG |

### 4.2.2. Construction of pSpCas9-mSA-2A-Puro

The monomeric Streptavidin (mSA) fragment was PCR amplified from the pCS2+Cas9-mSA vector (Addgene) using the Platinum SuperFi PCR MasterMix (Invitrogen) according to manufacturer's instructions. The following thermal profile was performed on a GeneAMP 2700 thermal cycler (ABI): 98°C for 1 min (initial denaturation), 98°C for 20 sec, 65/70°C for

30 sec, 72°C for 15 sec (x30 cycles), 72°C for 5 min (final extension). *Bam*HI restriction sites, as well as random nucleotide overhangs to assist with restriction enzyme digestion, flanked the mSA PCR fragment alongside a 48-nt linker at the 5' end (Table 9). Following gel purification using the QIAquick PCR Purification Kit (Qiagen), the mSA fragment was cloned into the intermediate lentiCRISPRv2 vector (Addgene) which was subject to restriction digest with Anza™ 5 *Bam*HI (Invitrogen). Double digest of lentiCRISPRv2-mSA and px459 was performed with *Eco*RV and *Stu*I (Invitrogen) followed by ligation of the mSA fragment into px459 using the Anza T4 DNA Ligase Master Mix. Recombinant px459 constructs were screened for the presence of mSA by *Bst*EII digestion (unique restriction site within the mSA insert). Presence of mSA resulted in a 473 bp long DNA fragment (Appendix B). All digestion reactions were performed according to manufacturer's instructions. Synthetic gRNA\_3 was synthesised by Sigma. A *Bbs*I cleavage site was introduced at the 5' end of the gRNA sequence which was also phosphorylated to facilitate cloning into the recombinant px459 vector following *Bbs*I digestion.

**Table 9.** Primer sequences used for mSA PCR amplification.

| Primer                      | Sequence (5'→3')  |
|-----------------------------|---|
| Overhang-BamHI-linker-mSA-F | TAAGCAGGATCCAGCGGTTTCAGAGACCCCAGGAACTAGC<br>GAGAGCGCTACACCGGAATCGGCGGAAGCGGGTATCACC |
| mSA-BamHI-overhang-R        | TGCTTAGGATCCAGACGCCGCAGACGGTTTAA  |

#### 4.2.3. Cell transfection/nucleofection

HEK293 cells were transfected with each CRISPR vector using the TransIT-LT1 transfection reagent (Mirus) whilst CWR22Rv1 were nucleofected with the same amount of each CRISPR vector using the Amaxa® Cell Line Nucleofector kit R (Lonza) using programme T-009 according to manufacturer's protocol. 6 µg of each CRISPR-Cas9 vector and 1 µg of ssODN were transfected individually or co-transfected into 2x10<sup>6</sup> cells seeded in 6-well plates (Corning). 48h post-transfection/nucleofection, puromycin selection (2µg/ml) was performed for 5 days and then CRISPR-modified cells were harvested and DNA extraction was performed

using the GenElute Mammalian Genomic DNA Miniprep Kit (Sigma) as described in Chapter 3.14.5. Selection and expansion of individual cell clones was performed in parallel. Puromycin-resistant clones were seeded into 96-well plates for continuous culture and downstream screening. CWR22Rv1 cells were treated with 30  $\mu$ M Scr7, where appropriate.

#### **4.2.4. Mutation detection assay (TIDE and SURVEYOR assay)**

AR exon 5 was amplified by PCR using the Platinum *Taq* DNA polymerase kit (Invitrogen). The primer sequences and PCR amplification protocol are provided in Tables 11 and 12 respectively. PCR products were cleaned-up using the QIAquick PCR Purification Kit (Qiagen). Purified PCR products were then outsourced and sequenced by Eurofins Genomics. Sequence chromatograms were analysed by TIDE (Tracking of Indels by DEcomposition) (Brinkman et al., 2014) to accurately determine the editing efficiency of each Cas9/gRNA complex. The SURVEYOR nuclease assay (IDT) was performed in parallel according to manufacturer's instructions. In brief, wild-type and CRISPR-edited PCR amplicons were mixed at 1:1 ratio, incubated at 95°C for 5 min and re-annealed by gradient cool down to RT. Resulting homo- or hetero- duplexes were then digested by the SURVEYOR nuclease as shown in Table 10.

**Table 10.** Reagents used for mutation detection analysis using the SURVEYOR assay.

| <b>Component</b>           | <b>Amount</b>             |
|----------------------------|---------------------------|
| DNA duplexes               | ~400 ng                   |
| MgCl <sub>2</sub> (150 mM) | 1/10 <sup>th</sup> volume |
| SURVEYOR Nuclease S        | 1 $\mu$ l                 |
| SURVEYOR Enhancer S        | 1 $\mu$ l                 |

The SURVEYOR digestion products were analysed on a 2% agarose gel. The Cas9 cleavage efficiency was estimated based on the relative intensity of the PCR fragments using ImageJ as described by Ran *et al.*.

$$f_{cut} = (b+c)/(a+b+c)$$

or

$$indel(\%) = 100 \times \left( 1 - \sqrt{\frac{b + c}{a + b + c}} \right)$$

where a = intensity of the undigested PCR amplicon/fragment

b and c = intensities of each digested fragment

**Table 11.** Primer sequences used for AR exon 5 amplification.

| Primer      | Sequence (5'→3')       |
|-------------|------------------------|
| AR exon 5-F | CCGAATACCAGAGCATCTCTG  |
| AR exon 5-R | TGAGCTAAGCTTCACTGTCACC |

**Table 12.** Reagents used per PCR reaction.

| Reagent                          | Volume (μl) |
|----------------------------------|-------------|
| Buffer (10x)                     | 5           |
| dNTPs (4 mM each)                | 1.25        |
| MgCl <sub>2</sub> (50 mM)        | 1.5         |
| Primer mix (10 μM each)          | 1.25        |
| <i>Taq</i> polymerase (100 U/μl) | 0.2         |
| Template DNA (50 ng/μl)          | 1           |
| dH <sub>2</sub> O                | Up to 50    |

#### 4.2.5. ssODN design

Three different W741L mutation-containing ssODNs (custom-designed and synthesised by Sigma) were used for the CRISPR knock-in optimisation experiments. All ssODNs were PAGE-purified to ensure length integrity. The desired W741L mutation was generated by the conversion of TGG to TTA (highlighted in red). The conversion created an *Mse*I restriction site in the sequence of all ssODNs (highlighted in red) for downstream restriction fragment length

polymorphism (RFLP) screening. The ssODN sequences contained CRISPR shield mutations/silent mutations (highlighted in magenta) to prevent binding of the gRNA upon ssODN incorporation and reversal of the knock-in. Homology arms ranging from 60-80 nucleotides upstream and downstream of the point mutation are highlighted in yellow and grey respectively. The ssODN sequences are given below:

- ssODN\_1: 180 bp, linear, symmetric, sense, no modification

GCTCAACCCGTCAGTACCCAGACTGACCACTGCCTCTGCCTCTTCTTCTCCAGG  
 CTTCCGCAACTTACACGTGGATGATCAATGGCTGTCATTCACTCCTTAAT  
 GGGGCTCATGGTGTGGCCATGGGCTGGCGATCCTTCACCAATGTCAACTCCA  
 GGATGCTCTACTTCGCCCC

- ssODN\_2: 151 bp, linear, asymmetric, sense, no modification

ACCCAGACTGACCACTGCCTCTGCCTCTTCTTCTCCAGGCTTCCGCAACTTACA  
 CGTGGATGATCAAATGGCTGTCATTCAATATTCGTTAATGGGGCTCATGGTGT  
 TGCCATGGGCTGGCGATCCTTCACCAATGTCAACTCCAGGATG

- ssODN\_3-F: 151 bp, linear, asymmetric, sense, 5'-biotinylated

ACCCAGACTGACCACTGCCTCTGCCTCTTCTTCTCCAGGCTTCCGCAACTTACA  
 CGTGGATGATCAAATGGCTGTCATTCAATATTCGTTAATGGGGCTCATGGTGT  
 TGCCATGGGCTGGCGATCCTTCACCAATGTCAACTCCAGGATG

- ssODN\_3-R: 151 bp, linear, asymmetric, antisense, no modification

CATCCTGGAGTTGACATTGGTGAAGGATCGCCAGCCCATGGCAAACACCATGA  
 GCCCATTAACGAATATTTGAATGACAGCCATTTGATCATCCACGTGTAAGTTGC  
 GGAAGCCTGGAGAAGAAGAGGCAGAGGCAGTGGTCAGTCTGGGT

#### **4.2.6. Flow cytometry and cell sorting**

Cells were harvested in 500 µl pre-warmed PBS, transferred into sterile FACS tubes (STEMCELL Technologies) and stained with 0.5 µg/ml sterile DAPI (BD Biosciences) for viability analysis. A sterile FACS Aria II sorter (BD Biosciences) was used to sort single, live (DAPI negative), GFP

positive cells which were ultimately ejected into 100  $\mu$ l of full medium per well of a 96-well plate. FlowJo was used for gating and data analysis.

#### **4.2.7. HDR detection**

Detection of potential CRISPR induced HDR events in the mixed cell population was carried out upon PCR amplification of *AR* exon 5 using a donor template specific primer (Table 13). The PCR reaction was performed under stringent annealing conditions ( $T_m=60^\circ\text{C}$ ) to guarantee specific binding of the primer exclusively to the template-containing sequence and not to the wild-type counterpart.

**Table 13.** Primer sequences used for PCR amplification of the knocked-in donor template in *AR* locus.

| <b>Primer</b> | <b>Sequence (5'→3')</b>  | <b>Detected ssODN</b> |
|---------------|--------------------------|-----------------------|
| Internal_1-F  | ACACGTGGATGATCAAATGG     | ssODN_1               |
| Internal_2-F  | ATGGCTGTCATTCAATATTCGTAA | ssODN_2 & ssODN3-F    |

#### **4.2.8. Colony screening – RFLP/Sanger sequencing**

Screening of potential CRISPR-edited clones was performed upon amplification of *AR* exon 5 as described in section X. PCR amplicons were purified using the QIAquick PCR purification kit (Qiagen) and were incubated with 1U *Mse*I (10 U/ $\mu$ l) (Invitrogen) at 37°C overnight in a GeneAMP PCR System 2700 (Applied Biosystems). Digested products were analysed on a 2% agarose gel. A batch of *AR* exon 5 PCR products was outsourced for Sanger sequencing (Eurofins Genomics) and sequence chromatograms were analysed using FinchTV.

#### **4.2.9. CRISPR off-target analysis**

Prediction of off-target effects was performed using the CRISPR design tool developed by MIT as well as the CCTop online tool developed by the University of Heidelberg.

#### **4.2.10. Subcellular Fractionation and WB**

HEK293 cells were transfected as in X. 48h post-transfection, cells were harvested and lysates were subject to subcellular fractionation using the Subcellular Protein Fractionation Kit

(Thermo Fischer Scientific). Briefly,  $2 \times 10^6$  cells were harvested, re-suspended in 200  $\mu$ l ice-cold CEB with protease inhibitors and placed on ice for 30 min with brief vortexing every 10 min. Lysates were spun at 500 x g for 5 min at 4°C and supernatants (cytoplasmic fraction) were transferred to new 1.5 ml Eppendorf tubes and stored at -20°C. Remaining cell pellets were re-suspended in 200  $\mu$ l ice-cold MEB with protease inhibitors, incubated on ice for 15 min and finally spun at 3000 x g for 5 min at 4°C. Following centrifugation, supernatants (membrane fraction) were removed and remaining cell pellets were re-suspended in 100  $\mu$ l ice-cold NEB with protease inhibitors and incubated on ice for 30 min with brief vortexing every 10 min. Following incubation, lysates were spun at 5000 x g for 5 min at 4°C and supernatants (soluble nuclear fraction) were transferred to new 1.5 ml Eppendorf tubes and stored at -20°C. Cell pellets were finally re-suspended in 100  $\mu$ l NEB supplemented with 5 mM CaCl<sub>2</sub> and 300 units of micrococcal nuclease and samples were incubated at 37°C for 10 min before being spun at 16,000 x g for 5 min at 4°C. Supernatants (chromatin fraction) were transferred to new 1.5 ml Eppendorf tubes and stored at -20°C. The different cell fractions were analysed on a 10% SDS-PAGE gel and immunoblotting was performed using an anti-Cas9 antibody (ab191468) (Abcam).

#### **4.2.11. siRNA knockdown**

CWR22Rv1 cells ( $2 \times 10^6$  cells per well of a 6-well plate) were reverse transfected with 25 nM of each siRNA (Table 14), either individually or as a pool for 24 hours prior to trypsinisation and subsequent nucleofection with the gRNA<sub>2</sub>/Cas9 pLV vector.

**Table 14.** siRNA sequences used to deplete components of the NHEJ pathway for precise genome editing.

| <b>Oligo</b> | <b>Sequence (5'→3')</b> |
|--------------|-------------------------|
| siScr        | UUCUCCGAACGUGUCACGU     |
| siKu70       | GAUGAGUCAUAAGAGGAUCAU   |
| siKu80       | CCUCAUAUCAAGCAUAACUAU   |
| siDNA-PKcs   | CCGGUAAAGAUCUAAUUCUA    |

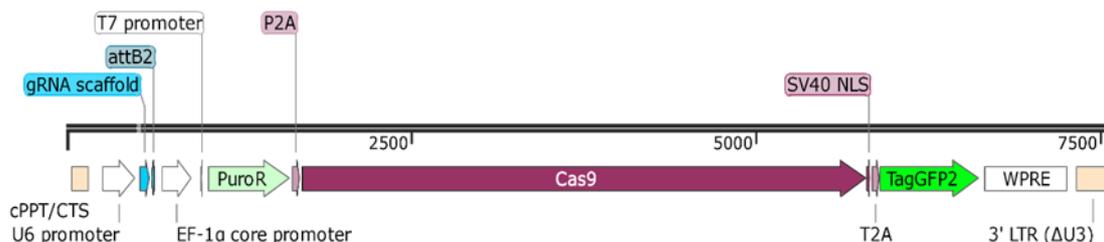
### 4.3. Results

The ultimate goal of this arm of the project was to generate, establish and validate an AR<sub>W741L</sub> mutant cell line, which would allow us to study the function of this particular mutant in a more physiological cellular background to ultimately identify and characterise its interacting partners and highlight potential therapeutic targets to benefit patient clinical outcome.

The recently developed CRISPR/Cas9 technology was exploited in an attempt to establish a PC cell line which would permanently harbour the desired mutation in its genome. Numerous optimisation steps were carried out and different approaches were applied (presented in this section) since the technology is fast-developing with novel findings emerging in quick succession to impact knock-in efficiencies. The main goal was to optimise the conditions which would increase the chances of getting the W741L mutation incorporated into the genome of the host PC cell line.

Firstly, two gRNAs were used to identify one which would demonstrate the highest editing efficiency. The all-in-one CRISPR/Cas9 vectors contained all the necessary CRISPR components: Cas9 cDNA under the control of the CMV promoter; a gRNA for RNA-guided recruitment of Cas9 to the target region for cleavage (under the control of a U6 promoter); and two selection markers, GFP and puromycin, to allow selection of the transfected and potentially edited cell population (Figure 17). The vectors were transfected and nucleofected in HEK293 and CWR22Rv1 cells, respectively, for 48h and upon selection with GFP and puromycin, the Cas9 editing efficiency was estimated by performing the mutation detection SURVEYOR assay and TIDE analysis.

pLV5-U6gRNA-EPCG (12,627 bp)



**Figure 17. Map of the all-in-one pLV-U6g-EPCG CRISPR vector used to ectopically express the gRNA and Cas9 nuclease in the host cell lines.** The gRNA (highlighted in blue) consists of a 19-nucleotide sequence and its expression is controlled by a U6 promoter (highlighted in white). The *S. pyogenes* Cas9 (highlighted in magenta) is fused to GFP for verification of Cas9-GFP expression in the host cell lines.

#### 4.3.1. The sgRNA\_2/Cas9 complex demonstrated high editing efficiency in HEK293 and CWR22Rv1 cells

Two different gRNAs were *in silico* designed and synthesized to target AR exon 5 in order to generate the clinically relevant AR<sup>W741L</sup> mutant in PC cells. They demonstrated similar quality score and their off-target sites across the human genome were also predicted (Figure 18).

Quality score: 95%

|             | Seed region             | PAM |        |
|-------------|-------------------------|-----|--------|
| On-target:  | AGGCTTCC [GCAACTTACACG] | TGG | gRNA_1 |
| Off-targets | AAGTTTA [GCAACTTACACG]  | AGG |        |
|             | AAGCCTTG [GCAACTTACACG] | TGG |        |
|             | AAGCTTTT [TCAACTTACACG] | CGG |        |
|             | ATGGAGCC [GCCACTTACACG] | AGG |        |
|             | AAGGGTCA [GTAAGTTACACG] | GGG |        |
|             | AAGCCTCG [GCAGCTTACACG] | TGG |        |
|             | AAGCCTCA [GCAGCTTACACG] | TGG |        |
|             | GTGCTGCT [GCCACTTACACG] | TGG |        |
|             | AAACATCT [GCCACTTACACG] | TGG |        |
|             | AAGCTTTG [GCAGCTTACACG] | TGG |        |
|             | ATGCTGAT [GCTACTTACACG] | TGG |        |
|             | TGGCAACG [GCAGCTTACACG] | TGG |        |
|             | AACTTTG [GCAGCTTACACG]  | TGG |        |
|             | AAGCCTTG [GCAGCTTACACG] | TGG |        |
|             | AAGCCTTG [GCAGCTTACACG] | AGG |        |
|             | AAGCCTTG [GCAGCTTACACG] | TGG |        |
|             | AAGCCTTG [GCAGCTTACACG] | TGG |        |
|             | AAGCCTTG [GCAGCTTACACG] | TGG |        |
|             | ATGCCTTG [GCAGCTTACACG] | TGG |        |

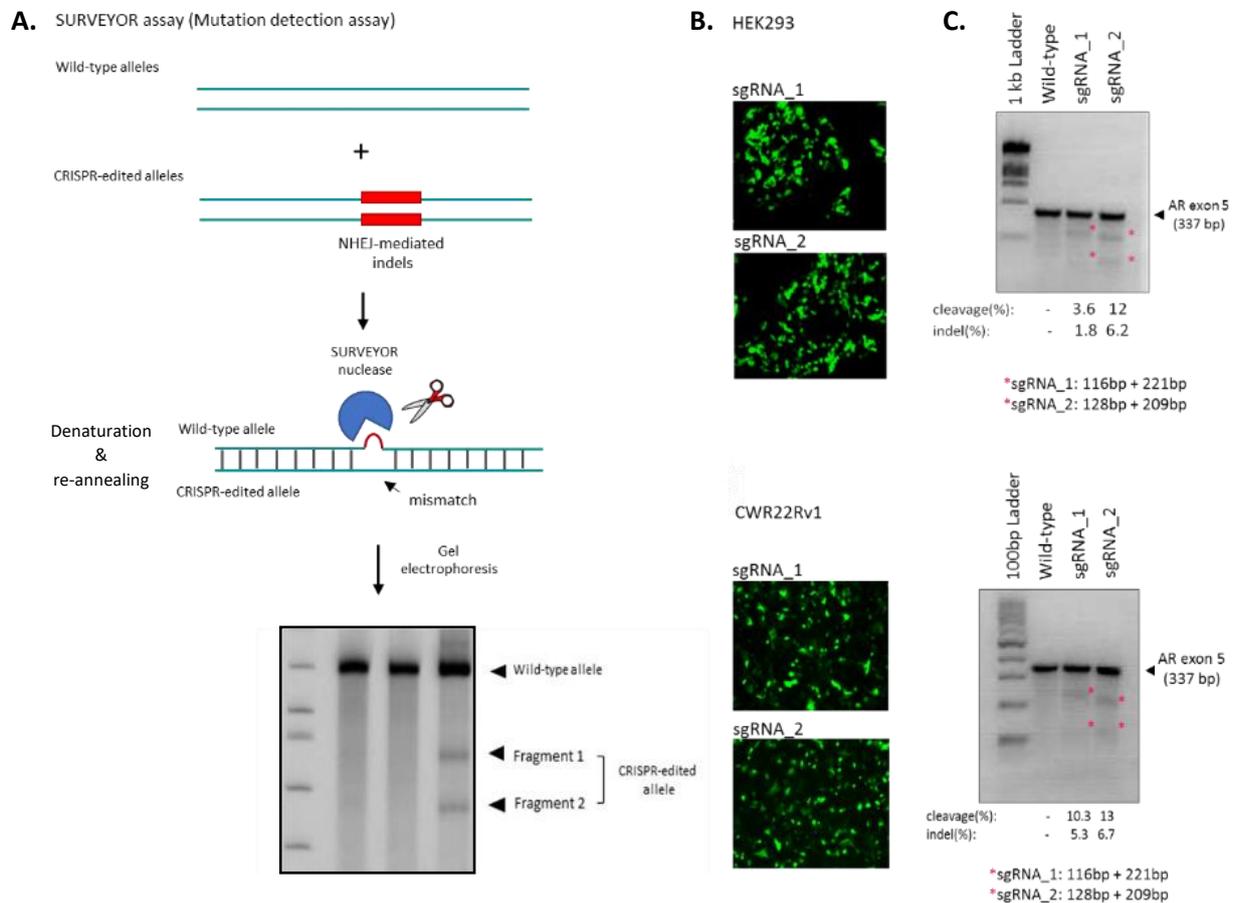
Quality score: 93%

|             | Seed region             | PAM |        |
|-------------|-------------------------|-----|--------|
| On-target:  | CTTACACG [TGGACGACCAGA] | TGG | gRNA_2 |
| Off-targets | TTGGCGCG [GGGACGACCAGA] | GGG |        |
|             | CTTGCACG [TATACGACCAGA] | TGG |        |
|             | CTTGCACG [TATACGACCAGA] | TGG |        |
|             | CTTGCACG [TATACGACCAGA] | TGG |        |
|             | TTCCACG [TGGAGACCAGA]   | AGG |        |
|             | TGTAATCG [TGGGCGACCAGA] | AGG |        |
|             | CTCCACA [TGGAGACCAGA]   | GGG |        |
|             | TTAACATG [GAGACGACCAGA] | GGG |        |
|             | GTGAAATG [TGGTCGACCAGA] | TGG |        |
|             | CACAGAAG [TGGTCGACCAGA] | AGG |        |
|             | TTTTCACA [TGGATGACCAGA] | TGG |        |
|             | ATAATTCG [TGGAGGACCAGA] | AGG |        |
|             | CTGAGACC [TGGAGGACCAGA] | AGG |        |
|             | CTTGCATG [TATACGACCAGA] | TGG |        |
|             | CTCCAGG [TATACGACCAGA]  | TGG |        |
|             | CTGCCACC [CAGACGACCAGA] | AGG |        |
|             | GAGATACG [TGGACAACCAGA] | GGG |        |
|             | GGTCACT [TGGAGGACCAGA]  | TGG |        |
|             | CTGACCC [TGGATGACCAGA]  | AGG |        |

**Figure 18. *In silico* analysis of AR exon 5 and surrounding sequences using the CCTop CRISPR/Cas9 target online predictor.** The two gRNAs with the highest quality score (%), as calculated by the CRISPR design tool (<http://crispr.mit.edu>), were synthesised and used to target AR exon 5. Scores higher than 75% indicated gRNAs with good on-target cleavage efficiency and low off-targeting based on DNA complementarity between the target sites and the gRNA sequence. The top 20 predicted off-target sites of each gRNA are listed and nucleotides highlighted in red depict mismatches between the gRNA sequences and their potential off-target sites.

To measure the editing efficiency of Cas9, HEK293 and CWR22Rv1 cells were transfected with 6 µg of each CRISPR vector. The editing efficiency of each gRNA/Cas9 complex was roughly estimated using the SURVEYOR assay, a mutation detection assay outlined in Figure 19A. Relative band intensity analysis indicated that gRNA\_2 outperformed gRNA\_1 in both cell

lines, despite the slightly lower quality score calculated by the CRISPR design tool (93 versus 95 respectively)(Figure 18). In particular, gRNA\_2 demonstrated cleavage efficiency of 12% versus 3.6% and 13% versus 10.3% in HEK293 and CWR22Rv1 cells, respectively (Figure 19C).

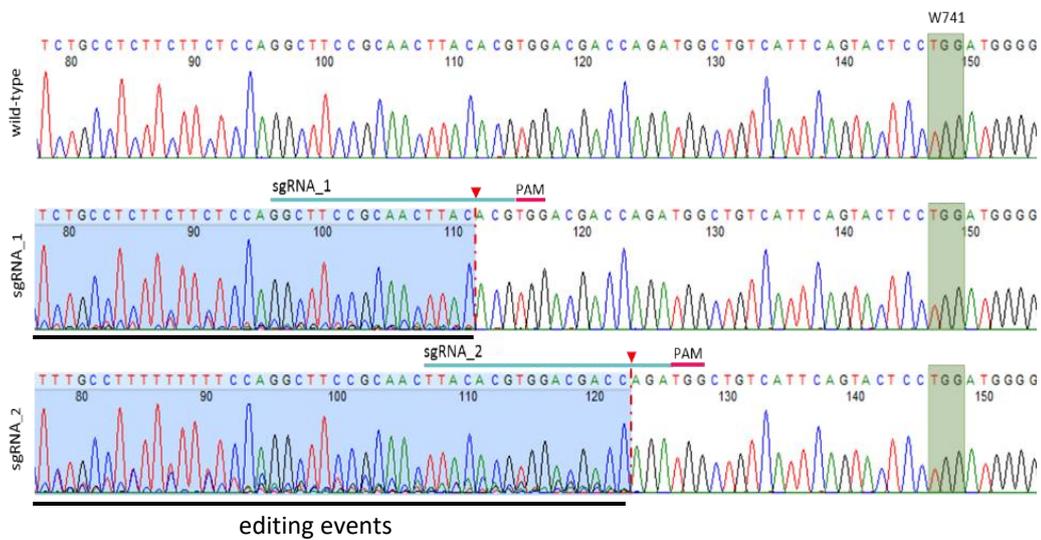


**Figure 19. Outline of the SURVEYOR assay used for the mutational analysis of the CRISPR engineered cell lines.** The wild-type and CRISPR-edited PCR amplicons were hybridised at 1:1 ratio to form heteroduplexes following denaturation and re-annealing. The SURVEYOR nuclease recognises and cleaves the NHEJ-mediated nucleotide mismatches (red loop) within the heteroduplexes, whilst homoduplexes remain intact. Successful cleavage generates two additional fragments which correspond to the CRISPR-edited allele. **B.** HEK293 (upper left) and CWR22Rv1 (bottom left) cells transfected and nucleofected, respectively, with the all-in-one pLV-U6g-EPCG CRISPR vectors. GFP fused to Cas9 was exploited for rapid verification of expression 8 h post-transfection/nucleofection. Representative 10x magnification images are shown. **C.** SURVEYOR nuclease assay for NHEJ-mediated indels within AR exon 5. Cleaved amplicons were analysed on a 2% agarose gel. Asterisks indicate the expected fragments upon cleavage by the SURVEYOR nuclease. NHEJ, Non-Homologous End Joining.

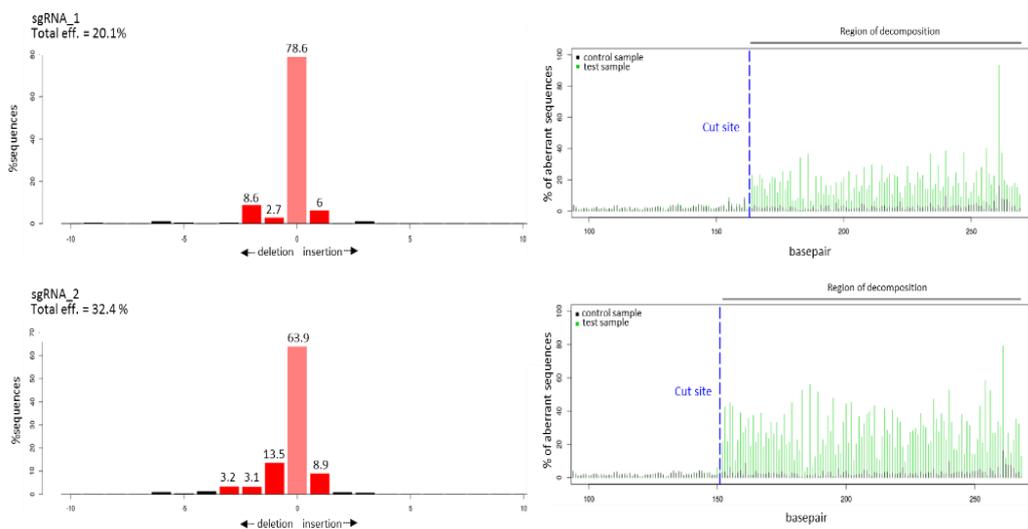
However, the SURVEYOR assay is not accurate enough as it only estimates the editing efficiency of each Cas9/gRNA complex. The latter was therefore precisely measured by performing Sanger sequencing of the AR exon 5 amplicons and downstream analysis of the

sequencing chromatograms using the TIDE online tool (Figure 20). Specific binding of the gRNA and cleavage 3bp upstream of the PAM site was confirmed. The gRNA\_2/Cas9 complex demonstrated significantly higher editing efficiency than the gRNA\_1 complex, ~30% over 20% respectively in the CWR22Rv1 cell line (Figure 20A-B). Lower editing efficiencies were observed in HEK293 cells despite the higher transfection efficiency (Figure 20C-D), potentially due to the tight conformation of chromatin around the AR locus and hence the limited access of Cas9 to exon 5 since the AR is not expressed in this cell line.

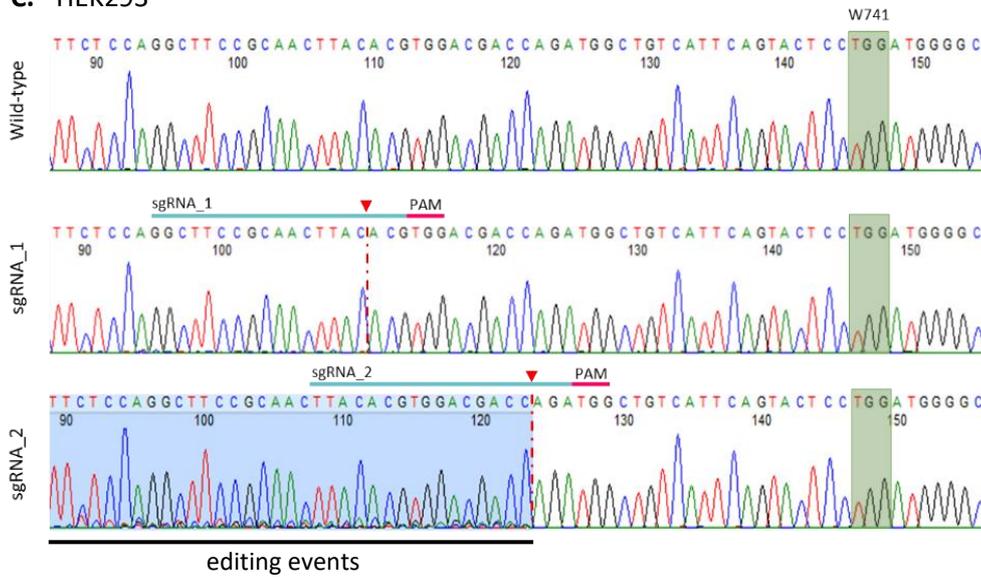
### A. CWR22Rv1



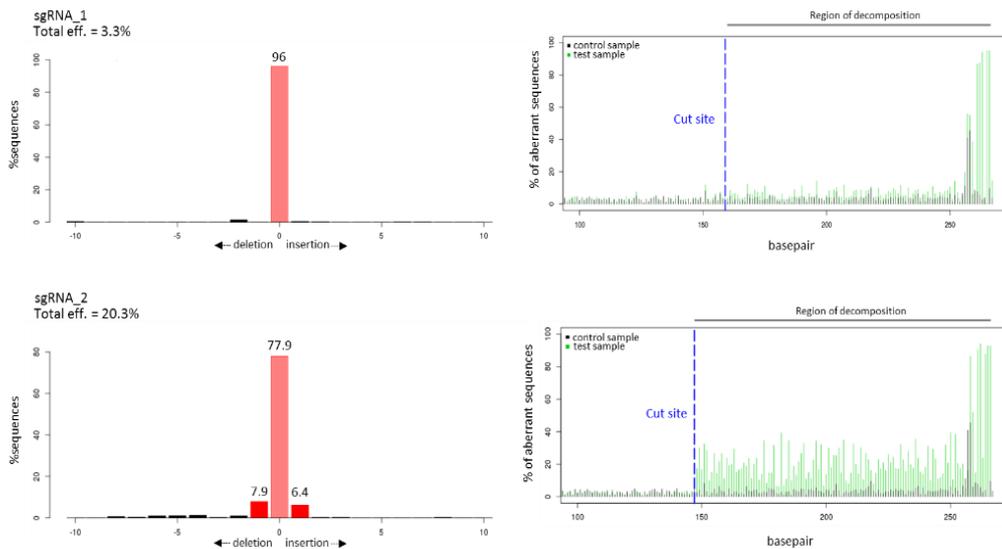
### B.



**C. HEK293**

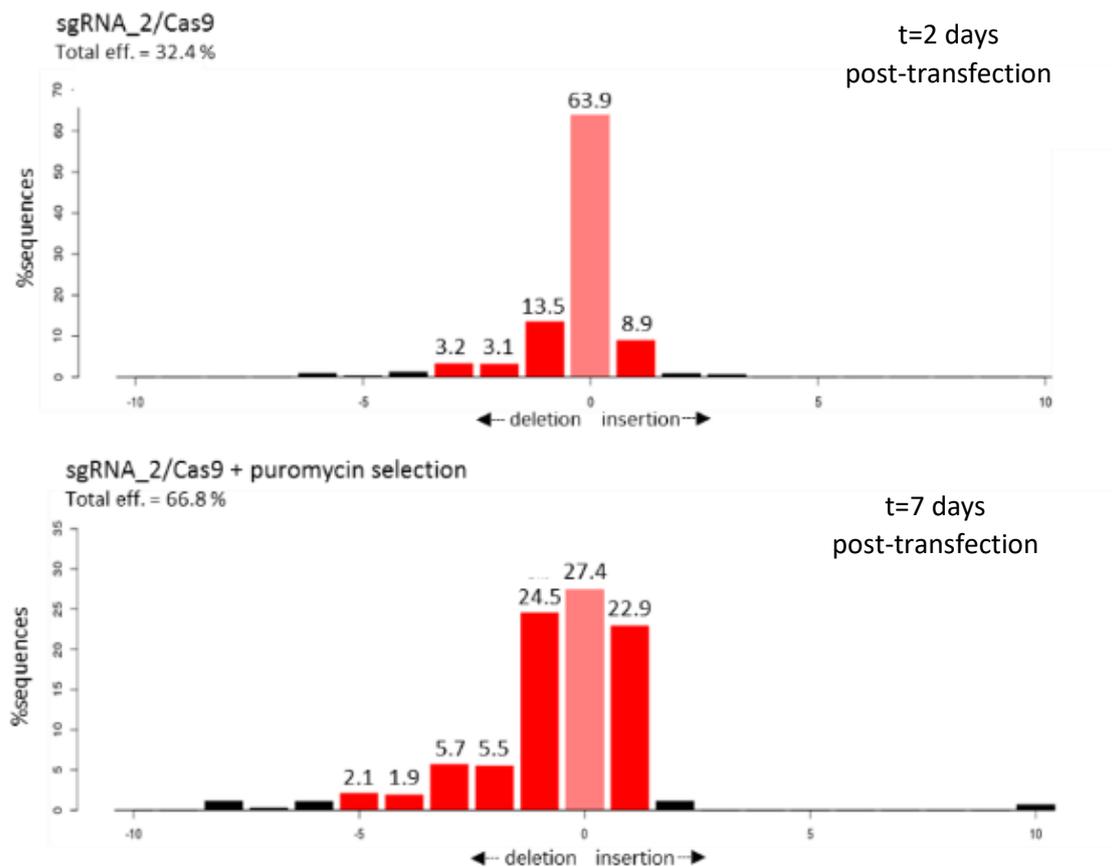


**D.**

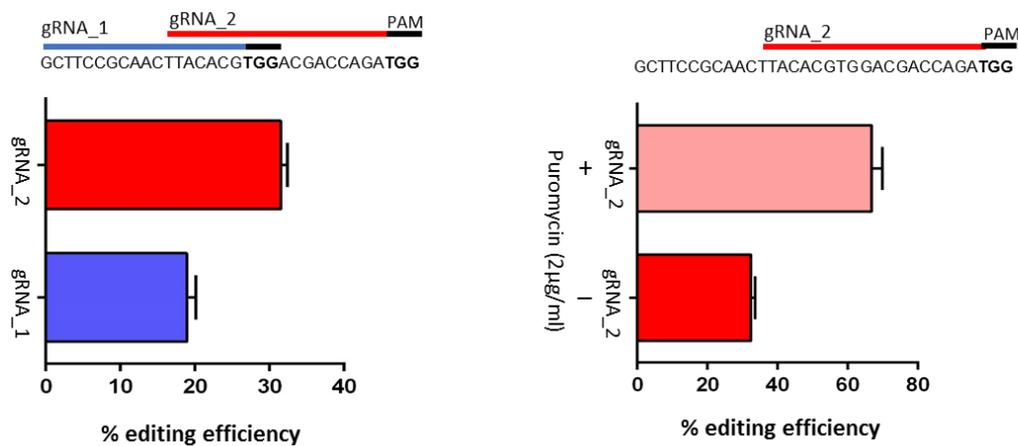


**Figure 20. Chromatograms of sequenced AR exon 5 upon editing by Cas9 in the CWR22Rv1 and HEK293 cell lines. A and C.** Noisy sequence regions with multiple underlying peaks indicate CRISPR-mediated indels (highlighted in blue). Cas9 binds to the expected PAM sites (magenta bars) and induces double strand breaks 3 bp upstream (indicated by red arrowheads and red dashed lines). **B and D.** The TIDE algorithm was used to calculate the percentage of NHEJ-mediated indels (total efficiency) (left panel) by aligning and decomposing the sequences of the edited versus the wild-type AR exon 5 (right panel). The intact sequence is represented by the middle coral bar whilst indels (insertions on the right, deletions on the left) are indicated by the red and black bars. The  $p$ -value associated with the estimated abundance of each indel is depicted by red or black for  $p < 0.05$  and  $p > 0.05$ , respectively.

Given that gRNA\_2 worked best in both cell lines, gRNA\_1 was dropped from downstream experiments. CWR22Rv1 cells were then subject to another round of nucleofection with the gRNA\_2/Cas9 vector, this time with puromycin selection for 5 days, to potentially enhance detection of genomic editing, prior to TIDE analysis. As expected, the editing efficiency of gRNA\_2/Cas9 markedly increased to 66.8% (2-fold) upon puromycin selection (Figure 21).



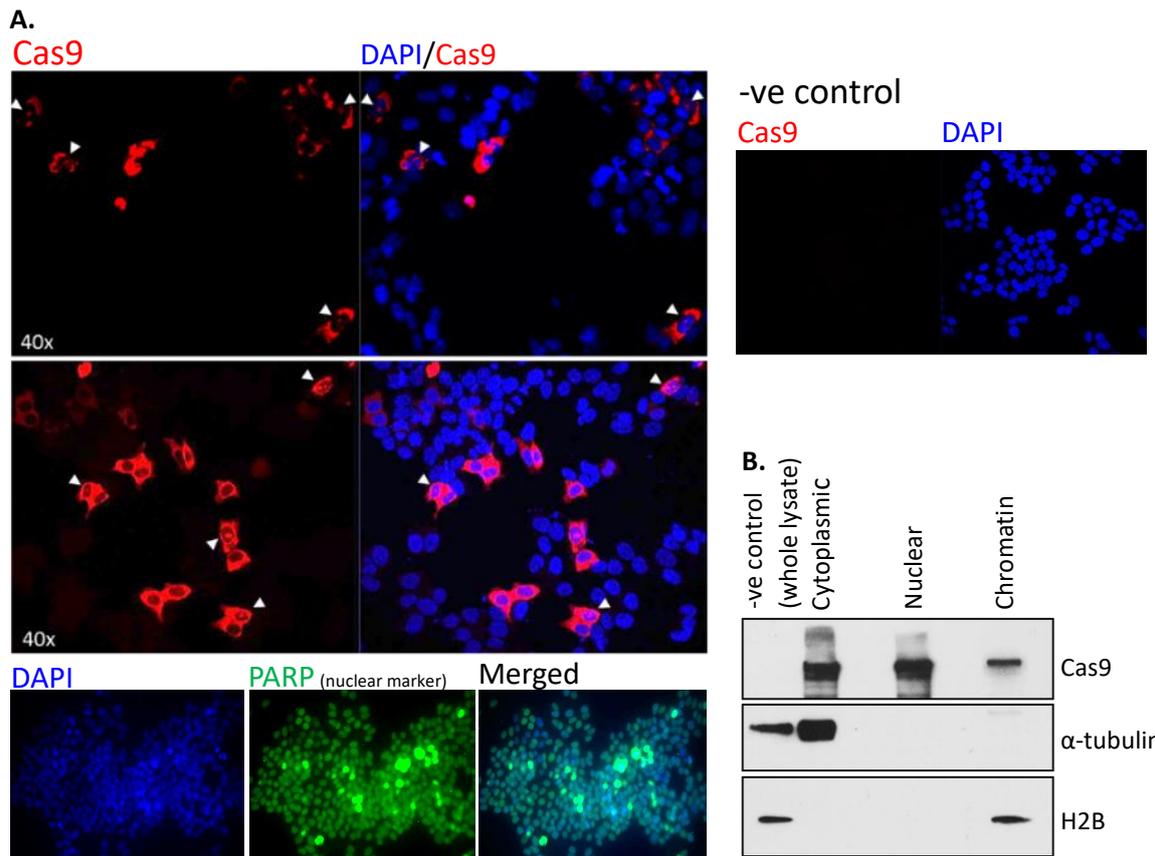
**Figure 21. CRISPR editing efficiency before and after puromycin selection in CWR22Rv1 cells.** Enrichment of the CRISPR engineered cell population via puromycin selection results in higher editing efficiency as assessed by TIDE.



**Figure 22. Bar charts of the editing efficiencies of gRNA\_1/Cas9 and gRNA\_2/Cas9 in CWR22Rv1 cells.** gRNA\_2 demonstrated significantly higher editing efficiency than sgRNA\_1 (left panel). Editing efficiency of gRNA\_2 is significantly elevated upon puromycin selection (right panel). 6 µg of each gRNA/Cas9 vector were transiently transfected in CWR22Rv1 cells by nucleofection. Cells were cultured for 48h in full media following nucleofection. DNA extraction, PCR amplification of AR exon 5, Sanger sequencing and finally TIDE analysis of the chromatograms were performed to assess the editing efficiency of each gRNA. Values represent means  $\pm$  SEM from two technical replicates.

#### **4.3.2. gRNA\_2 efficiently guides Cas9 to target chromatin sites**

Although high editing efficiency of gRNA\_2/Cas9 was observed at the DNA level, indicative of Cas9 presence in the nucleus, it was important to establish that Cas9 was present in the nucleus of target cells. To this end, gRNA\_2/Cas9-expressing CWR22Rv1 cells were immunostained with an anti-Cas9 antibody to examine the level of Cas9 translocation to the cell nucleus as well as binding events of the nuclease to DNA. Confocal fluorescence microscopy confirmed that Cas9 successfully enters the cell nucleus and binds to chromatin, where it is observed in foci (Figure 23A), to exert its editing function. Subcellular fractionation of HEK293 cells transfected with the same Cas9 vector was carried out, in parallel, to examine the cellular localisation of Cas9 by WB. Pancellular distribution of the nuclease was observed, but more importantly when one performs genome editing experiments, Cas9 was also present in the chromatin fraction indicating efficient RNA-guided recruitment of Cas9 to DNA (Figure 23B). A nuclear marker (PARP) was included in the immunofluorescence analysis to ensure that the observed nuclear speckles corresponded to genuine chromatin-bound Cas9 and the fixation/permeabilisation protocol that was used allowed access of the Cas9 antibody to the nucleus and hence the observed foci are not due to fixation artifacts.

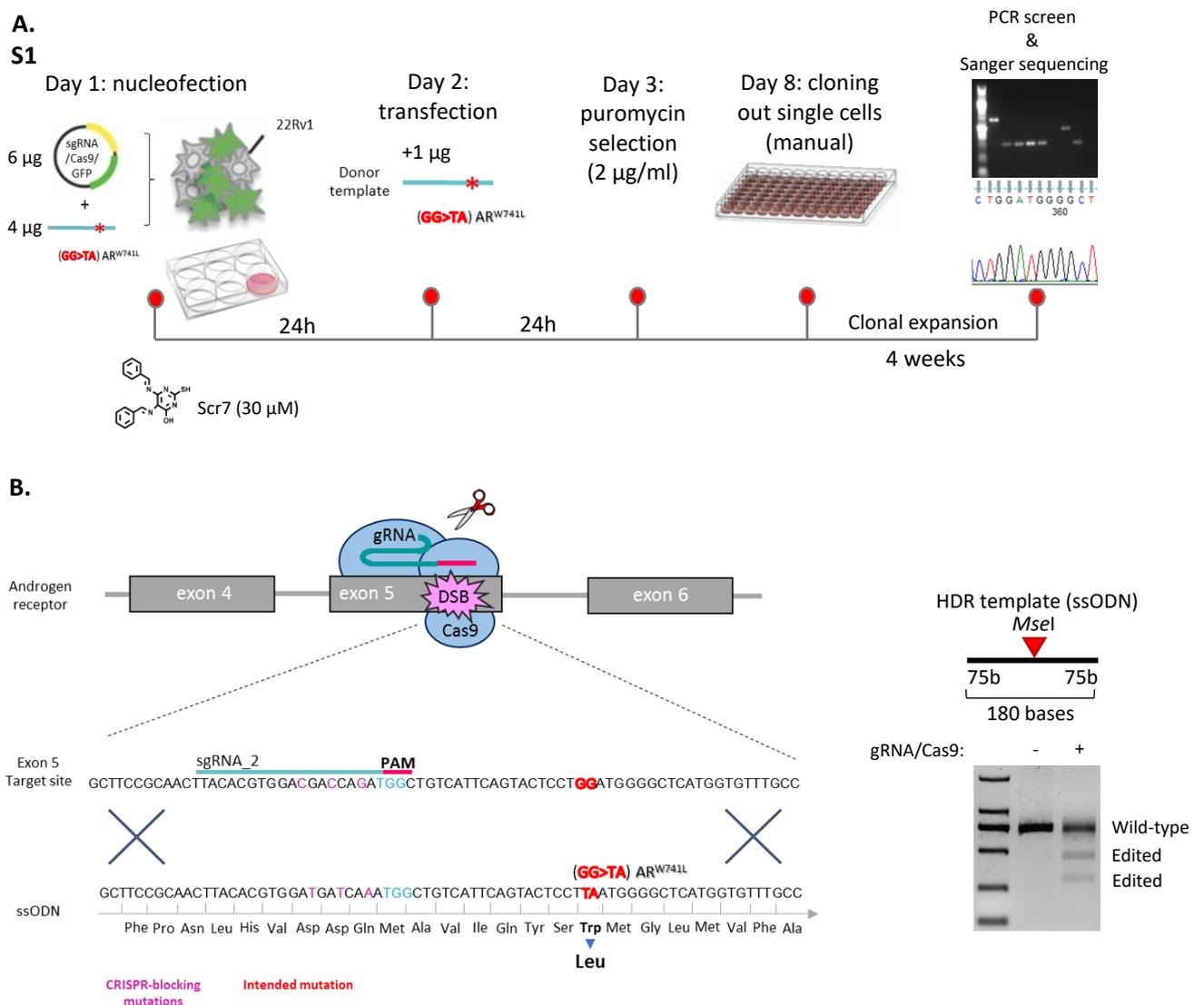


**Figure 23. Cas9-NLS successfully translocates to the cell nucleus, where it is observed in foci. A.** Representative confocal microscopy images of CWR22v1 cells expressing sgRNA\_2/Cas9. Arrows depict chromatin-bound Cas9. Whole lysate derived from non-transfected cells was used as a negative control (right panel) to ensure specific binding of the Cas9 antibody. Cells were immunostained with a PARP1 antibody, in parallel to assess antibody access to the nucleus (bottom panel). **B.** Subcellular fractionation of HEK293 cells transfected with the all-in-one gRNA\_2/Cas9 expressing vector and assessment of Cas9 cellular distribution by WB to confirm RNA-guided recruitment of Cas9 to target chromatin sites.

#### **4.3.3. sgRNA\_2/Cas9 and ssODN delivery in CWR22Rv1 cells to generate the clinically relevant $AR^{W741L}$ mutation via HDR**

In order to introduce the W741L mutation in the *AR* locus, a single-stranded donor template (ssODN) was introduced into CWR22Rv1 cells alongside the gRNA\_2/Cas9 encoding vector. Cells were treated with the Scr7 NHEJ inhibitor to enhance the frequency of HDR by transiently blocking the more prominent NHEJ DNA repair mechanism. A second round of transfection with extra ssODN was carried out to guarantee its presence and increase its availability while gene editing was still occurring (Figure 24A). The ssODN was synthesised to

be 180-bp long with flanking regions (75-bp long) of 100% homology to AR exon 5 on each side of the W741L mutation. The GG to TA substitution (W741L) creates a cut site for *MseI*, which was exploited for RFLP screening and identification of clones with the W741L knock-in mutation. Finally, single base substitutions were intentionally introduced upstream of the PAM site (positions: -2, -5 and -8) to generate silent mutations (CRISPR-blocking mutations) which would prevent binding of Cas9 and reversal of the W741L knock-in mutation back to wild-type upon editing (Figure 24B).



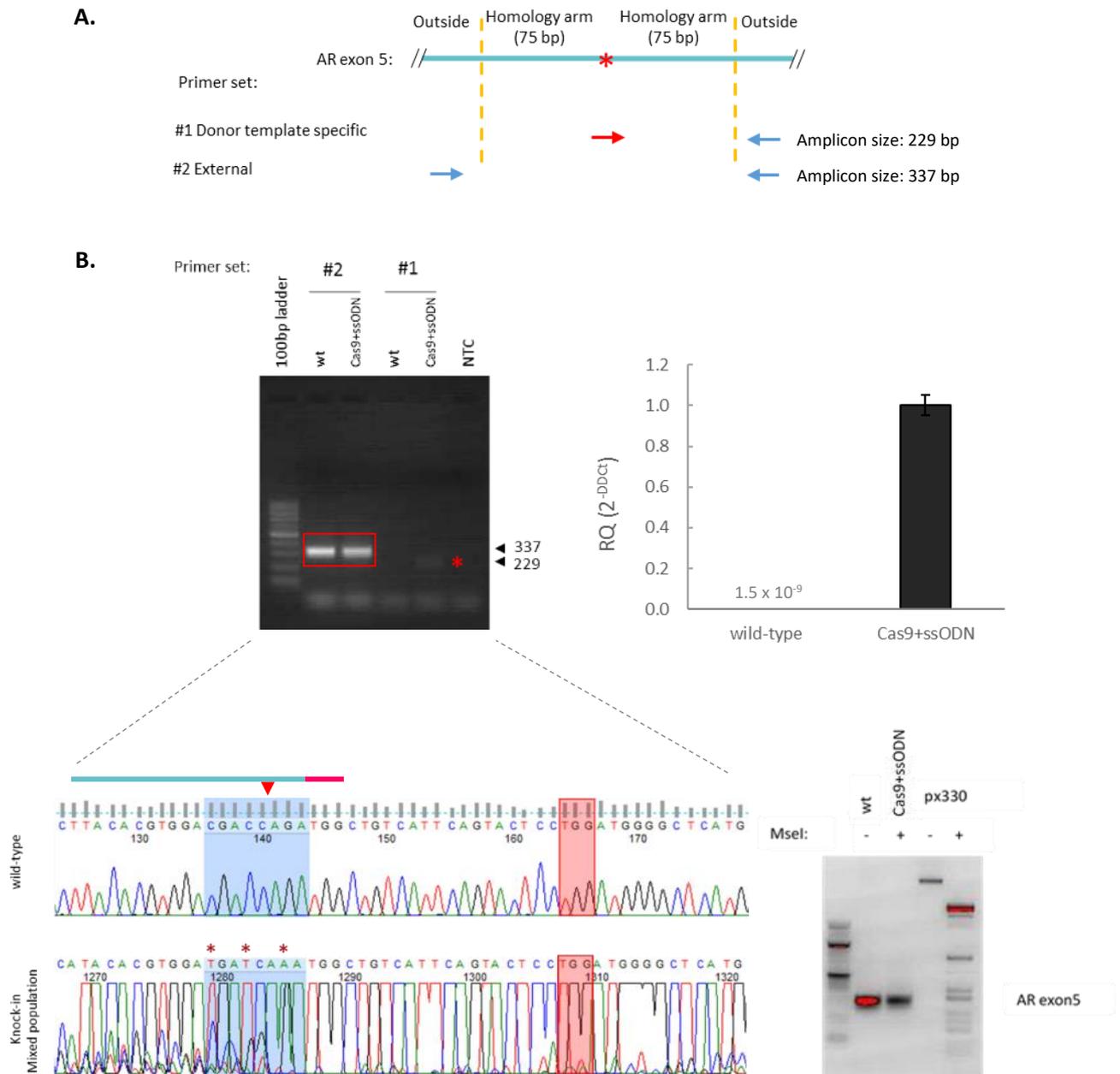
**Figure 24. Schematic outline of the strategy designed to CRISPR engineer and generate the CWR22Rv1 ARW741L cell line.** **A.** The indicated amounts of sequence-verified gRNA<sub>2</sub>/Cas9 vector and HDR donor template/ssODN were nucleofected into CWR22Rv1 cells, which were cultured in Scr7-containing media for 48h prior to puromycin selection. Cells were supplied with extra ssODN 24 hours post-nucleofection by conventional transfection. CRISPR edited cells were then clonally expanded to derive isogenic cell lines harbouring the AR<sub>W741L</sub> mutation. Finally, DNA extraction and PCR

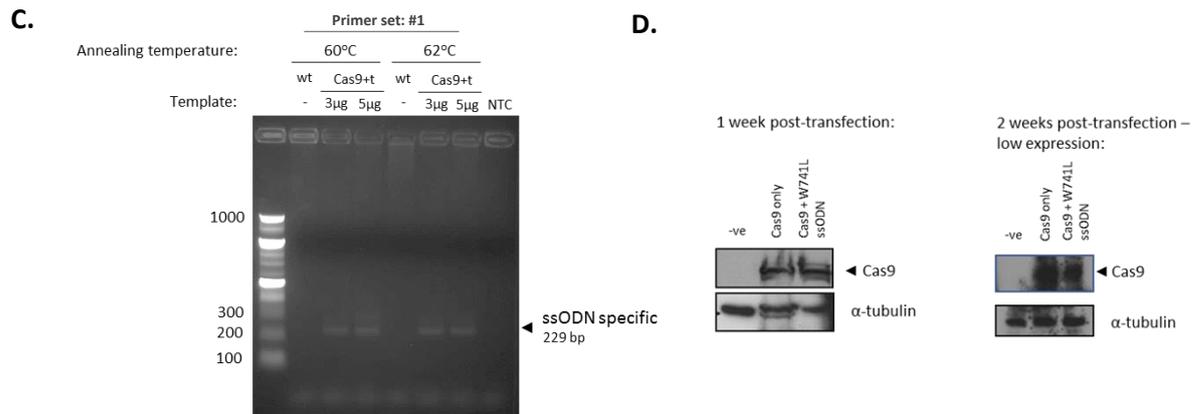
amplification of AR exon 5 were carried out and amplicons were sequenced to identify AR<sub>W741L</sub> positive clones. **B.** Representation of part of the donor template sequence used to introduce the AR<sub>W741L</sub> mutation via Homologous DNA Repair (HDR) in the CWR22Rv1 cell line. The template is 180 bp long and contains homologous arms of 75bp on each side of the cut site. Three silent mutations were introduced upstream of the PAM site (highlighted in magenta) to prevent reversal of the sequence to wild-type caused by the gRNA/Cas9 binding to the target site upon CRISPR knock-in. The GG to TA substitution (W741L), highlighted in red, creates an *Mse*I restriction cut site which was exploited for identifying AR<sup>W741L</sup> positive clones by RFLP screening. An example of the expected DNA fragments upon Cas9 cleavage is illustrated (right panel).

#### ***4.3.4. Successful HDR events following CRISPR induced DSBs were observed in CWR22Rv1 cells***

A PCR primer specific to the donor template was designed to examine the presence of the knock-in AR<sub>W741L</sub> allele in the puromycin-enriched CRISPR engineered cell population. The three silent mutations upstream of the PAM site were exploited so the primer was designed to bind to that mutation "hot spot" allowing amplification (Figure 25A). On the contrary, mismatches between the ssODN specific primer and the wild-type sequence would prevent binding under stringent annealing conditions and this would indicate the presence of unedited sequences and finally provide a measure for HDR efficiency. Therefore, genomic DNA from CRISPR engineered and wild-type mixed cell populations was extracted two weeks post-transfection and conventional as well as real-time PCR amplification of AR exon 5 was performed using the specific primer for the knock-in allele as well as a primer pair which binds to the genomic region which surrounds the ssODN (Figure 25A). Amplification was observed exclusively in the CRISPR edited sample using the ssODN specific primer, suggesting that efficient HDR and successful knock-in of the donor template had occurred in CWR22Rv1 cells (Figure 25B). Two different quantities of the donor template were introduced into the cells to test whether increased intracellular concentration of the ssODN correlates with higher knock-in rates. Of note, the supply of extra ssODN did not seem to have significant effect on the HDR events that occurred in CWR22Rv1 cells (Figure 25C). The PCR amplicon was also subject to Sanger sequencing and RFLP analysis to verify the incorporation of the AR<sub>W741L</sub> allele. Surprisingly, only the CRISPR-blocking mutations were efficiently incorporated as indicated by the strong corresponding peaks in the sequencing chromatogram. Low underlying peaks which correspond to the desired mutation may exist at the W741 site (Figure 25B), however they are unlikely to be detected by Sanger sequencing or RFLP due to the low abundance of the mutation in the mixed population in addition to the low sensitivity of the detection

methods. Finally, expression of Cas9 was assessed one and two weeks post-transfection. Cas9 was still detected in the mixed cell population two weeks post-transfection (Figure 25D), although at negligible levels indicating that (minimal) editing was still occurring. Whether and how this impacts HDR remains elusive.





**Figure 25. Detection of HDR events induced in CWR22Rv1 cells upon Cas9-mediated DSB. A.** Primer design to detect knock-in of the donor template (ssODN) by exploiting the mutation hot-spot (depicted by *\**) right upstream of the PAM. The presence of the AR<sub>W741L</sub> knock-in mutation was interrogated using an internal forward primer specific for the donor template and an external reverse primer specific to a sequence outside of the homology arm (primer set #1). The genomic region is targeted by external primers surrounding the ssODN as an internal control (primer set #2; reverse primer is the same as that in set #1). **B.** Genomic DNA was analysed by end-point and real-time PCR using primer sets #1 and #2 to assess the ssODN knock-in efficiency. A 229 bp band (depicted by an asterisk) is present in the Cas9+ssODN sample, indicating the successful knock-in of ssODN in the AR locus (left panel). Real-time PCR was used as a more sensitive detection method and the fold change in HDR efficiency between the wild-type and Cas9+ssODN samples was calculated using the  $\Delta\Delta C_t$  method (right panel). DNA extraction from the CRISPR engineered cell population was carried out and AR exon 5 was sequenced. Partial knock-in of the ssODN was detected in the mixed population with only the CRISPR-blocking mutations (indicated by asterisks) being incorporated. The W741 site (highlighted in red) remained intact as assessed by RFLP analysis of the AR exon 5 amplicon (obtained using primer set #2). No additional bands, indicative of ssODN knock-in and subsequent *MseI* cleavage, were observed in the Cas9+ssODN sample (bottom right panel). The px330 vector was used as a positive control for successful cleavage by *MseI*. **C.** CWR22Rv1 cells were transfected with 3  $\mu$ g and 5  $\mu$ g of ssODN and cell lysates derived from wild-type (wt) and CRISPR edited (Cas9+t) cell populations were analysed by end-point PCR using primer set #1 at two different annealing temperatures (60°C and 62°C). **D.** Expression of Cas9 in the mixed cell population was assessed one and two weeks post-transfection by Western blotting using the anti-Cas9 (7A9-3A3) antibody, 45 min exposure. DSB, Double Strand Break; t, template.

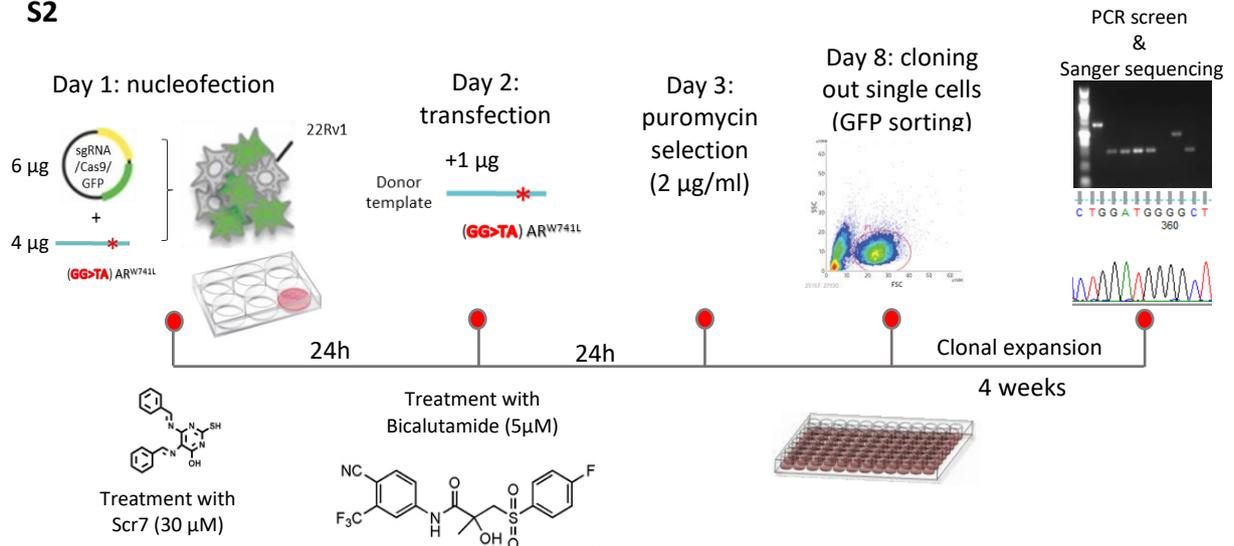
#### 4.3.5. Optimisation steps to enhance genome editing via HDR-mediated ssODN knock-in

ssODN knock-in is not an easy task to achieve hence three different strategies (hereafter referred as S1, S2 and S3) were followed to help generate the AR<sub>W741L</sub> mutation in the CWR22Rv1 cell line (Figures 24A and 26A). The major aim in all cases was to enhance HDR in

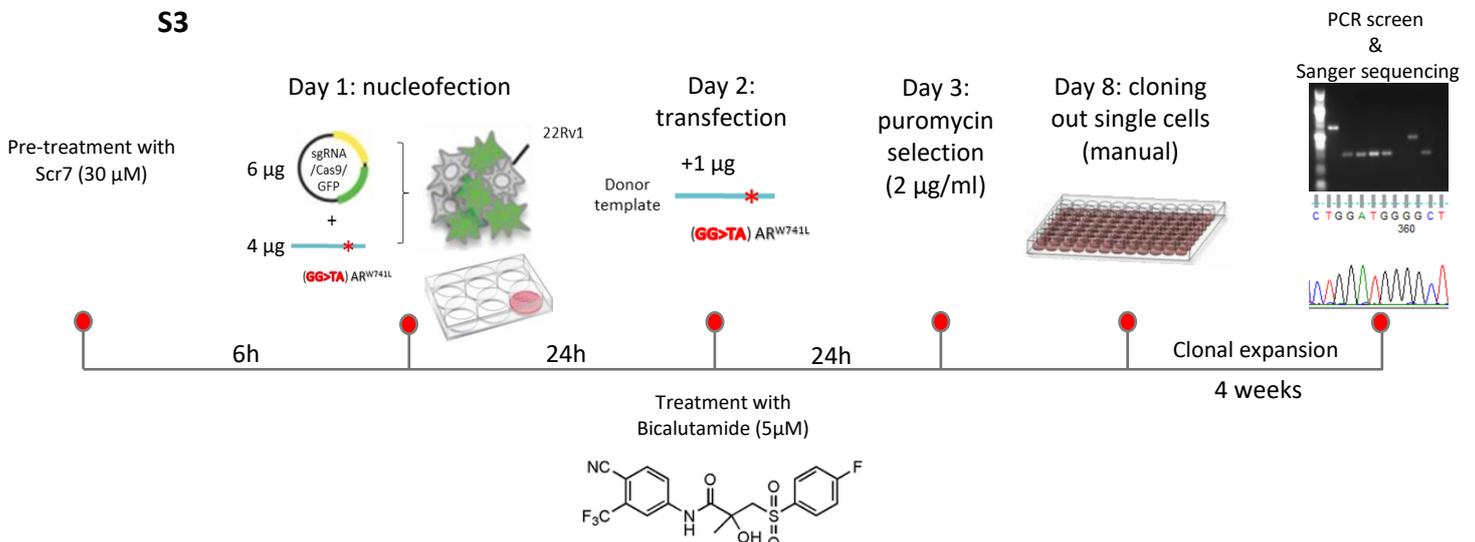
order to achieve precise knock-in of the AR<sub>W741L</sub> mutation in the host genome. The total amount of DNA introduced into the cells was kept constant. The mixed cell populations were treated with Scr7 either at transfection (S1 and S2) or 6h prior to transfection (S3) to potentially block NHEJ to a much higher extent. Additionally, cells were treated with 5 μM bicalutamide (S2 and S3), as AR<sub>W741L</sub> positive cells would in theory have a growth advantage in bicalutamide conditions and they would ideally out-grow non-edited cells potentiating identification of AR<sub>W741L</sub> positive clones in the downstream steps. Finally, single cell cloning was carried out upon puromycin selection (S1 and S3) or by FACS sorting of Cas9-GFP positive cells (S2) in order to rule out the clonal expansion of cells which did not express the CRISPR complex but acquired resistance and survived in puromycin conditions.

**A.**

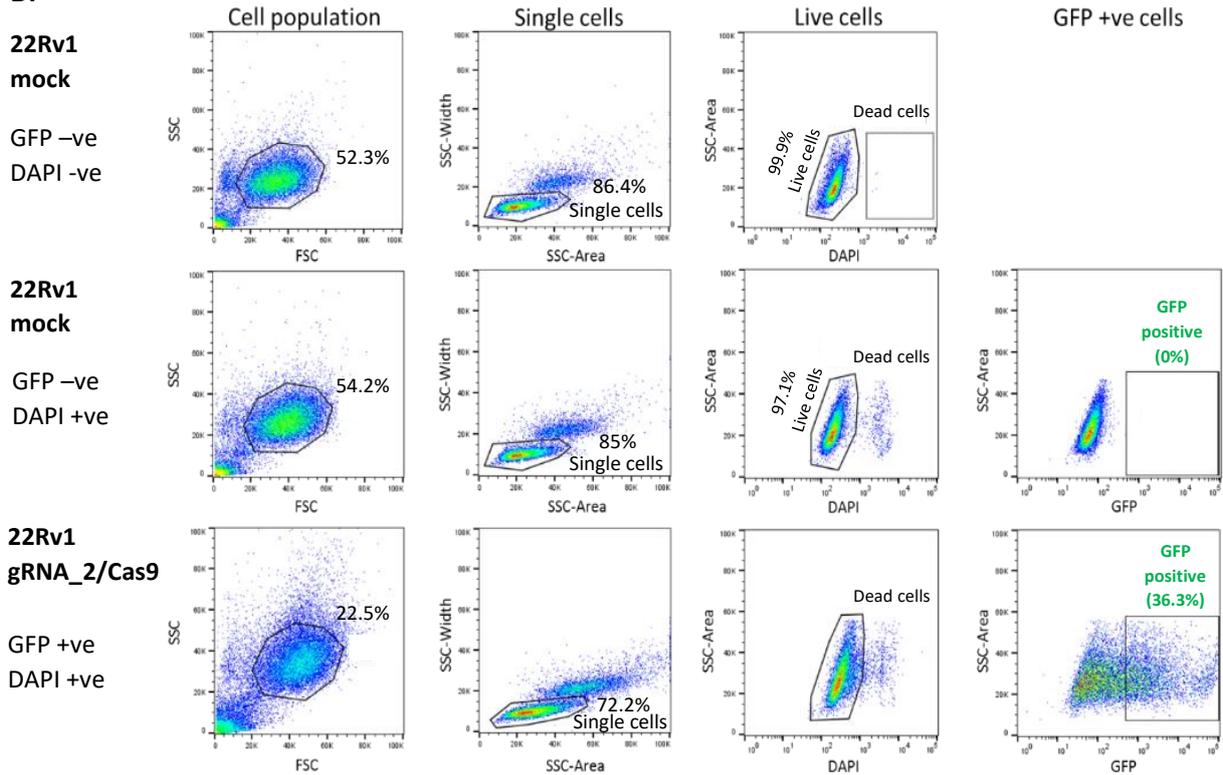
**S2**



**S3**

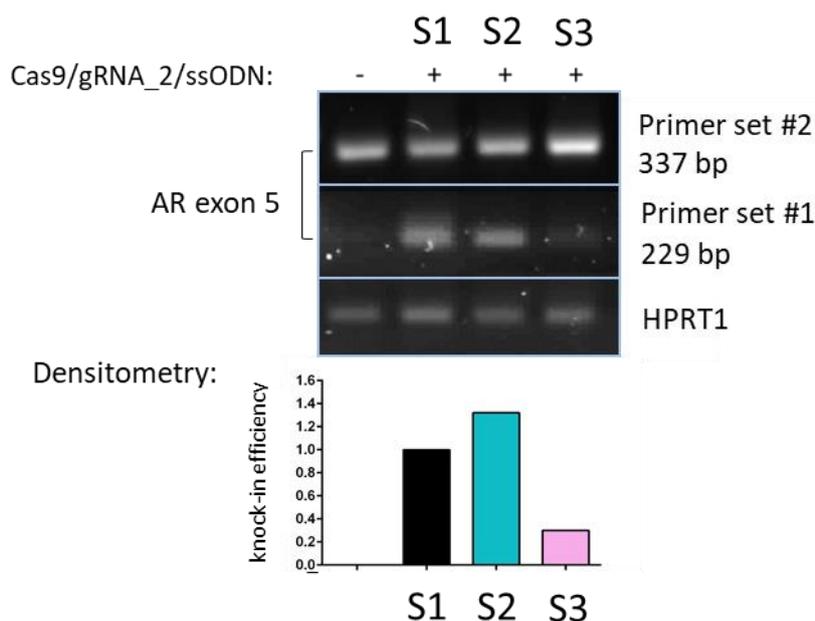


**B.**



**Figure 26. Schematic of the alternative strategies applied for the CRISPR-mediated generation of the CWR22Rv1 AR<sup>W741L</sup> cell line.** A. Cells were nucleofected with the indicated amounts of gRNA\_2/Cas9 vector and ssODN. The mixed cell population was treated with 5  $\mu$ M Bicalutamide 24 h post-nucleofection to favour growth of the AR<sub>W741L</sub> positive cells. Cas9-GFP positive cells were then FACS sorted and cloned out for subsequent screening by PCR and Sanger sequencing (S2). Cells were pre-treated with 30  $\mu$ M Scr7 NHEJ inhibitor for 6h prior to nucleofection as an attempt to enhance HDR and cloned out following conventional puromycin selection (S3). Cell transfections and treatments were performed as described in S1. **B.** FACS sorting of Cas9-GFP positive cells for single cell cloning and clonal expansion. CWR22Rv1 cells nucleofected with empty vector (mock) were used for setting the parameters for precise gating.

All mixed populations were comparatively interrogated for the presence of the AR<sub>W741L</sub> allele by conventional PCR using the ssODN specific primer. Successful HDR events with similar efficiencies (band intensities) occurred in the S1 and S2 cell populations as opposed to the S3 cell population which was treated with Scr7 prior to nucleofection with the CRISPR editing complex (Figure 27). This suggests that treatment with the inhibitor may have blocked ligase IV (member of the NHEJ complex), but cells may have upregulated other ligases to mediate NHEJ in response to longer exposure to Scr7.

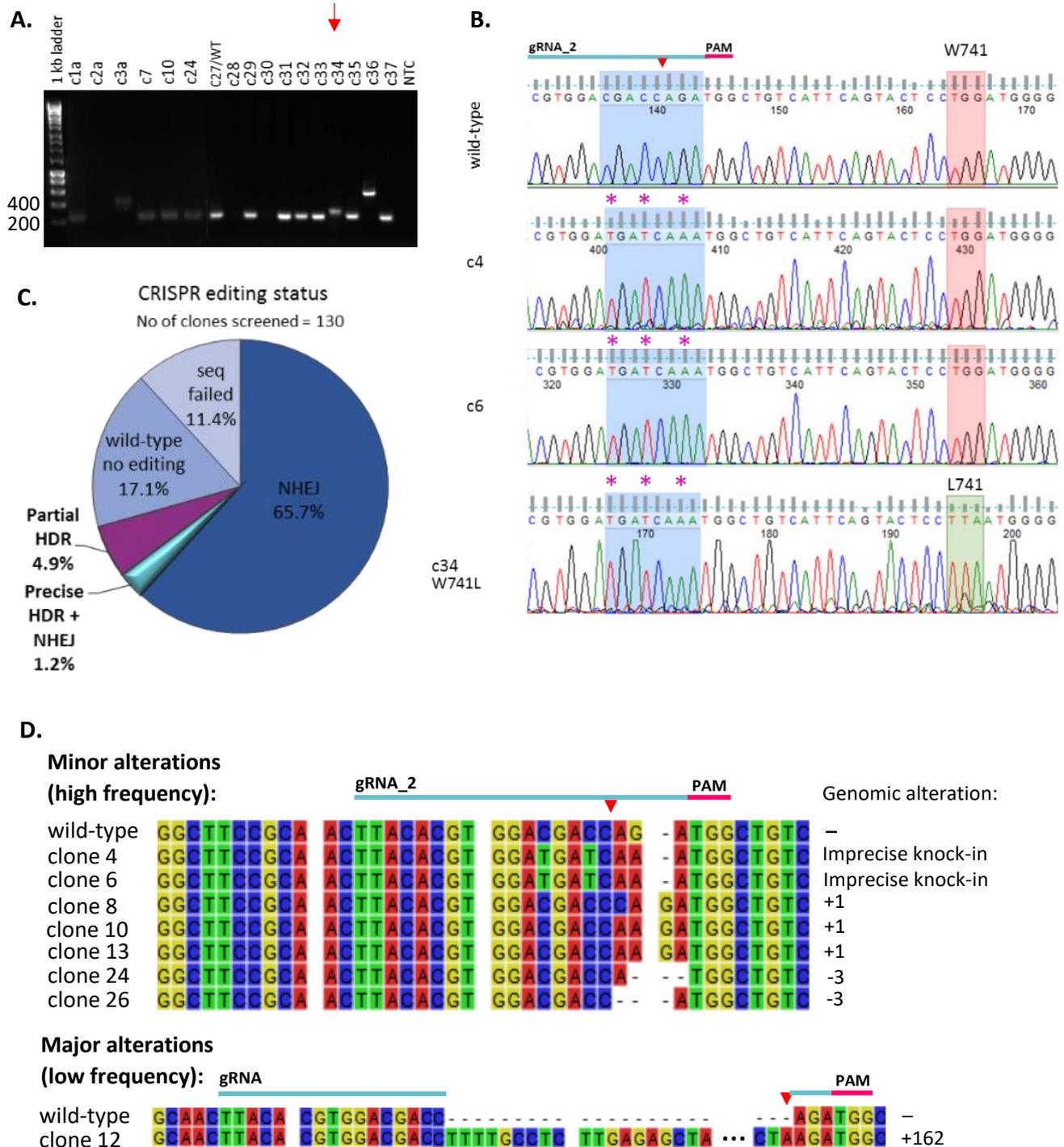


**Figure 27. Assessment of the HDR events and ssODN knock-in efficiencies in the mixed cell populations obtained from all three experimental strategies outlined in Figures 24A and 26A.** The AR exon 5 genomic region was amplified using an ssODN specific primer (primer set #1) as well as an external primer set (primer set #2). Densitometry analysis of the gel bands was performed to estimate the knock-in efficiency in each cell population (S1-S3). AR exon 5 band intensities were normalised to HPRT1.

#### **4.3.6. Mutation-to-cut-site distance dictates ssODN knock-in fate**

Single cell cloning was performed to isolate and screen clonal cell populations which encompass the AR<sup>W741L</sup> mutation. A large number of clones was expanded and screened using PCR coupled with RFLP or Sanger sequencing (Figure 28A). CWR22Rv1 clones harbouring only the CRISPR-blocking mutations (c4, c6 and c34) with one of them also harbouring the W741L mutation (c34) were obtained. Partial incorporation of the ssODN evident of the presence of the CRISPR-blocking mutations and absence of the W741L mutation was observed (Figure 28B), indicating that incomplete HDR may have occurred. In particular, the CRISPR-blocking mutations which reside upstream of the cut site were predominantly observed in the mixed cell population as evident by the high chromatogram peaks (Figure 28B) as well as in 4.9% of the screened clones whereas the W741L mutation was observed in 1.2% of the screened clones (Figure 28C), suggesting that mutations closer to the cut site are highly likely to be incorporated into the genome as opposed to distal mutations which lie further upstream or downstream of the cut site. Notably, NHEJ was predominantly observed in most cell clones,

accounting for 65.7% of total editing events (Figure 28C). Screening indicated that most of the NHEJ mediated edits were short nucleotide deletions or insertions whilst very few clones contained larger DNA fragment alterations (Figure 28D).

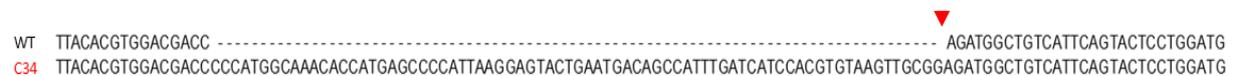


**Figure 28. Spectrum of CRISPR editing events in CWR22Rv1 clonal cell populations.** **A.** Genomic DNA was extracted from the CWR22Rv1 clonal populations (n=130) obtained from all three experimental strategies outlined in Figures 24A and 26A and AR exon 5 was amplified by PCR. **B.** PCR amplicons were subject to Sanger sequencing for detection of AR<sub>W741L</sub> positive cell clones. Sequencing chromatograms

of representative HDR clones are shown and the AR exon 5 wild-type sequence at the top is used as reference. Clones 4 and 6 contained the CRISPR-blocking mutations (depicted by asterisks) but not the W741L mutation (red box) (incomplete HDR). Clone 34 contained the desired W741L mutation (green box) in addition to the CRISPR-blocking mutations (depicted by asterisks). **C.** The pie chart summarises the mutation patterns (NHEJ, HDR, HDR+NHEJ or no editing) of 130 single-cell clones as determined by Sanger sequencing. **D.** Spectrum of genomic alterations caused by Cas9 in edited clonal cell populations. Alignment to the wild-type sequence of AR exon 5 was carried out using CLC sequence viewer (Qiagen).

The AR<sub>W741L</sub> positive clone (c34) was further examined for its mutation status. The presence of NHEJ-mediated indels upstream and/or downstream of the knock-in mutation was suspected by an apparent upshift of the c34 AR exon 5 PCR band when the latter was compared to the wild-type counterpart (unedited clone c27) (Figure 28A). However, further investigation was required to identify the precise genomic alterations caused in clone c34. Essentially, amplification of AR exon 5 with a high-fidelity polymerase, Sanger sequencing and subsequent alignment to the wild-type sequence were performed and revealed the extent of the scar left in the genome as a result of Cas9 editing. Surprisingly, co-existing NHEJ-mediated insertions upstream of the cut site were observed (Figure 29) suggesting that repair mechanisms can be coupled or act simultaneously to fix the Cas9-induced break.

WT TTACACGTGGACGACC ----- AGATGGCTGTCATTAGTACTCCTGGATG  
 C34 TTACACGTGGACGACCCCATGGCAAACACCATGAGCCCATTAAGGAGTACTGAATGACAGCCATTTGATCATCCACGTGTAAGTTGCGGAGATGGCTGTCAATCAGTACTCCTGGATG

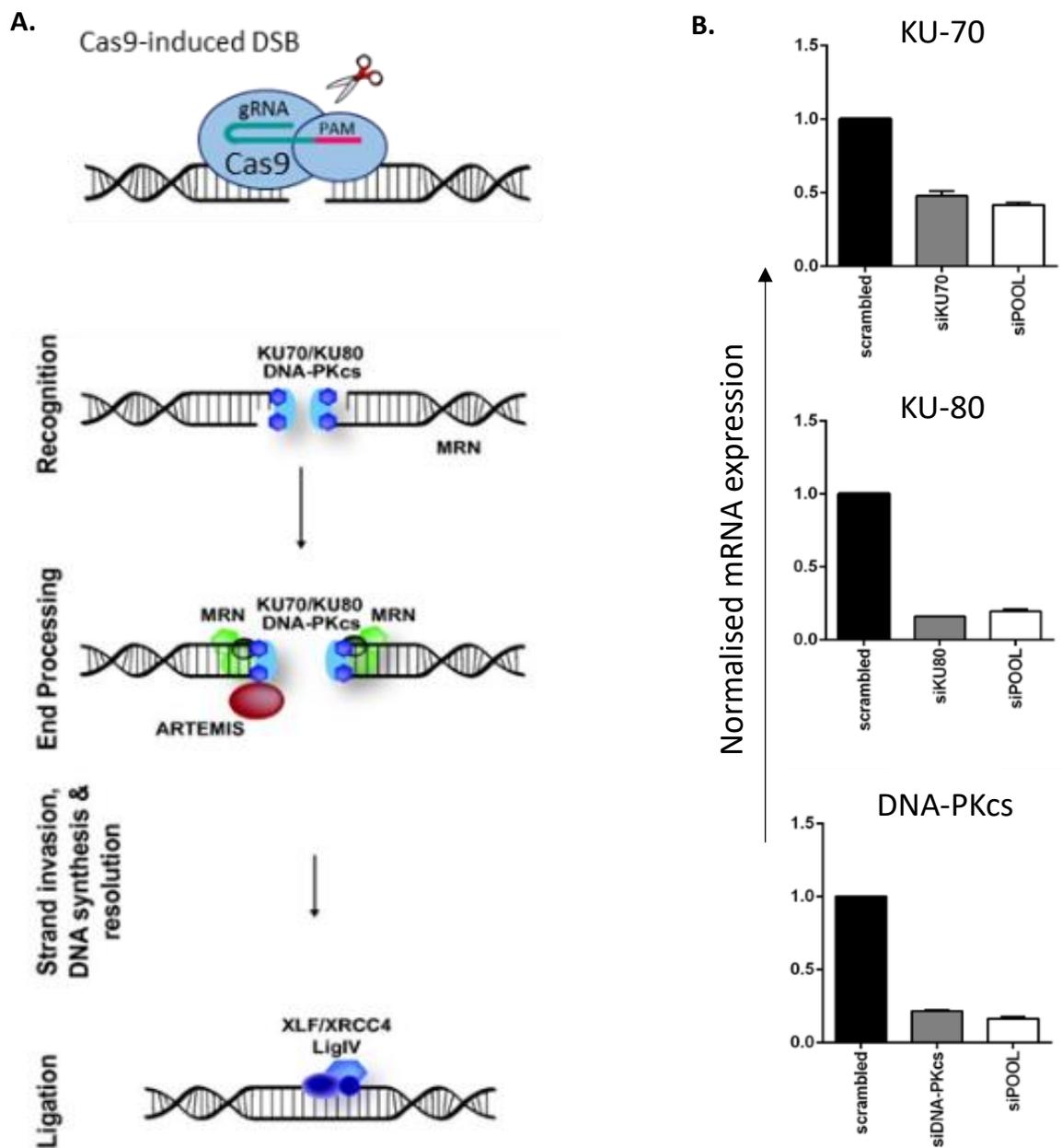


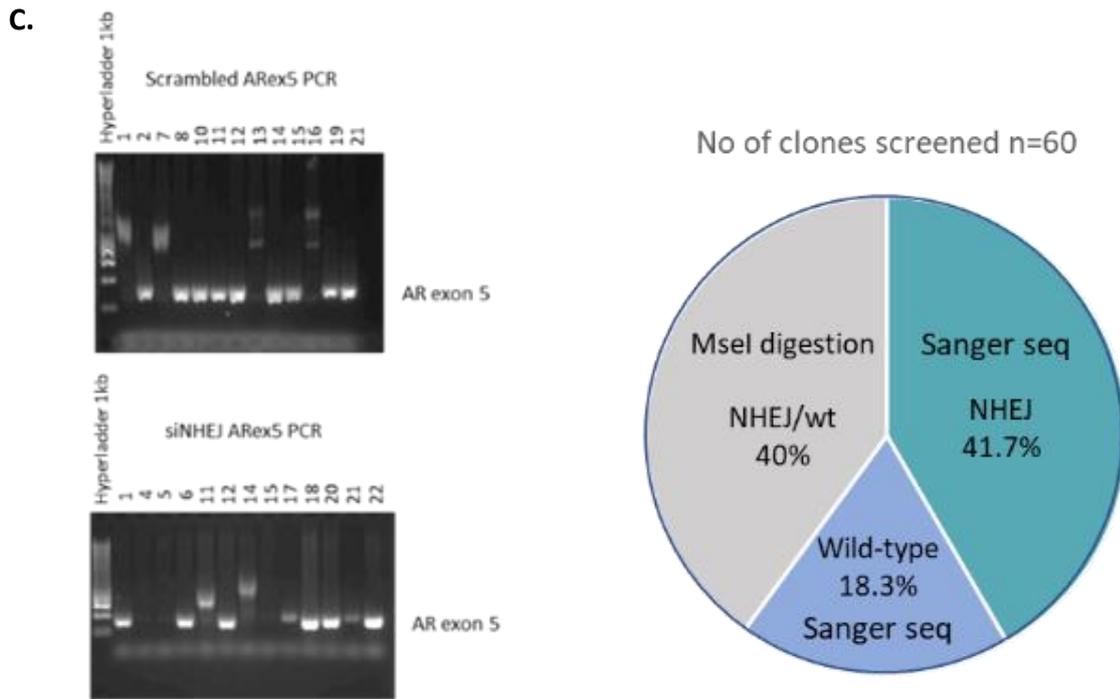
**Figure 29.** Sequence alignment of the wild-type and C34 AR exon 5 using ClustalW shows a NHEJ-mediated insertion of 75 bases upstream of the cut site (depicted by red arrowhead).

#### ***4.3.7. Depletion of key components of the NHEJ pathway does not enhance HDR***

To tackle the apparent prevalence of NHEJ, cells were treated with the ligase IV inhibitor Scr7 in an attempt to limit the activity of the pathway and allow HDR to occur more prevalently. However, inhibition of the NHEJ pathway by Scr7 has been controversial. In fact, it has been reported that ligase IV is not a genuine target of Scr7 (Greco et al., 2016). Therefore, other key members of the pathway which are rapidly triggered upon DSB and function upstream in the pathway, such as Ku70, Ku80 and DNA-PKcs (Figure 30A), were depleted by siRNA-mediated knockdown. Each NHEJ component was knocked-down 24h prior to nucleofection, either individually or as a pool to prevent functional compensation events. Knockdown of the NHEJ components was confirmed by qPCR (Figure 30B). Following puromycin selection, individual CWR22Rv1 cell clones were screened for the presence of the W741L mutation by conventional PCR coupled with Sanger sequencing or RFLP (Figure 30 C). More precisely, 36

out 60 clones were screened by Sanger sequencing whereas the remaining 24 clones were RFLP analysed by *MseI* digestion. Depletion of endogenous Ku70, Ku80 and DNA-PKcs did not seem to improve HDR rates over NHEJ rates. In fact, no knock-in positive clones were observed in this round whilst NHEJ events were still quite prevalent, despite the apparent compromise in pathway activity. Clones with wild-type AR sequence were finally observed indicating that a number of cells in the mixed population managed to bypass puromycin treatment and survive in such conditions without being edited (Figure 30C).





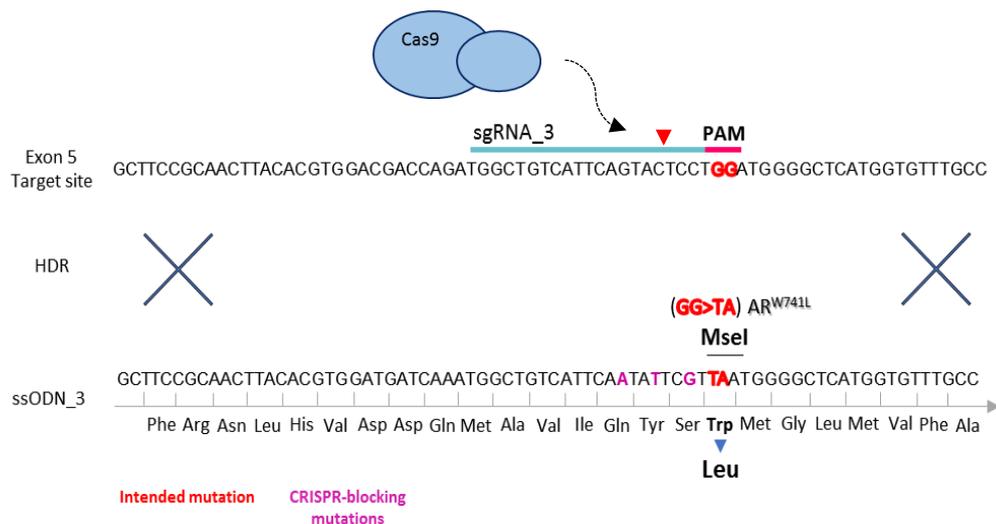
**Figure 30. Inhibition of NHEJ does not enhance HDR.** **A.** Schematic of the NHEJ signalling cascade with its different components acting at different levels across the cascade to repair the DSB upon Cas9-induced cleavage. **B.** CWR22Rv1 cells were reverse transfected with KU-70, KU-80 and DNA-PKcs-targeting siRNAs as indicated 24h prior to nucleofection with the CRISPR editing complex. The level of knockdown was evaluated by qPCR 72h post-transfection. **C.** Representative agarose gel images of AR exon 5 PCR-amplified clones analysed by Sanger sequencing and *MseI* digestion (left panel). The pie chart summarises the mutation patterns (NHEJ or no editing/wild-type) observed in 60 single-cell clones as determined by Sanger sequencing and *MseI* digestion (right panel).

#### ***4.3.8. Reducing the mutation-to-cut-site distance in concert with biotinylation of the donor template and covalent tagging of Cas9 with mSA facilitate HDR***

The importance of the distance between the cut site and the desired mutation has been highlighted as a major contributing factor in HDR frequency and proficiency, with distal edits being less frequently knocked-in to the desired loci due to various reasons discussed in the next section of this chapter. Equally important, the presence of ssODN in close proximity to the cut site seems to be a crucial determinant of the repair process to be chosen by the insulted cells upon DSB. More precisely, a novel approach which exploits the link between streptavidin-Cas9 and biotin-ssODN conjugates has been successfully applied in mice and markedly improved HDR rates. In line with these recent findings, a new gRNA and ssODN were designed as shown in Figure 31A. The new gRNA, hereafter named gRNA\_3, binds adjacent to the PAM site (TGG) which comprised the target codon to be mutated to TAA to ultimately generate the W741L mutation within *AR* exon 5. Hence, the distance between the desired

mutation and the cut site was restricted to 4 nucleotides. The gRNA oligo was cloned into the px459 vector following the strategy shown in Appendix A. The performance of gRNA\_3 was assessed using TIDE and the editing efficiency of the gRNA\_3/Cas9 complex was 22.4% in CWR22Rv1 cells prior to puromycin selection (Figure 31C). Off-target sites were predicted using CCTop and mismatches between the off-target sites and the gRNA sequence were permissive to continue (Figure 31B). Moreover, the new ssODN was 151 bp long, asymmetric and harboured three CRISPR shield mutations right upstream of the W741L mutation (Figure 31A) to prevent reversal of knock-in to wild-type. In addition, the ssODN was biotinylated at 5' end to form a complex with streptavidin-tagged Cas9 at the cut site. Finally, two alternative strep-Cas9/biotin-ssODN complexes were formed whereby a sense 5'-biotinylated ssODN was either utilised alone or as a double stranded template with its complementary (antisense) unmodified strand following recent trends in CRISPR knock-in experiments (Renaud et al., 2016).

**A.**

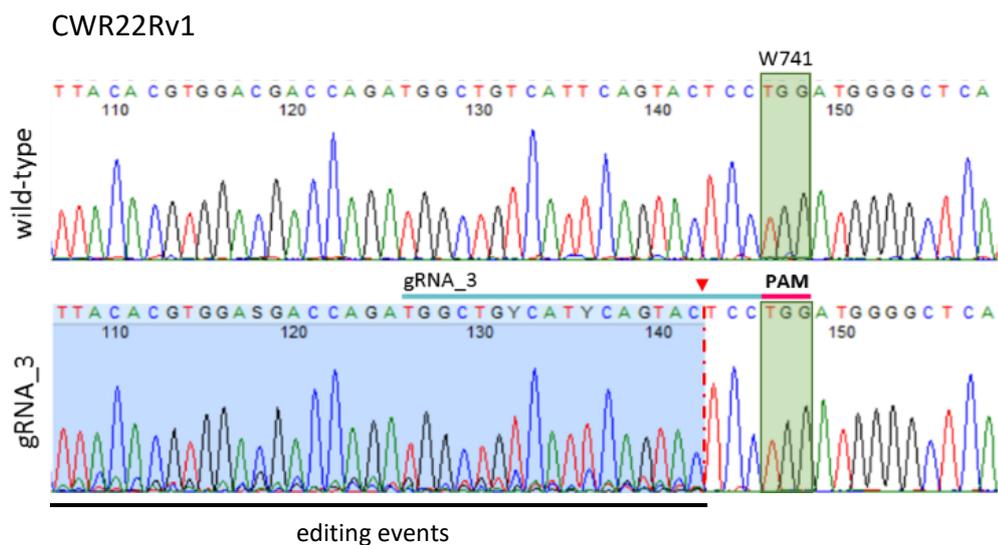


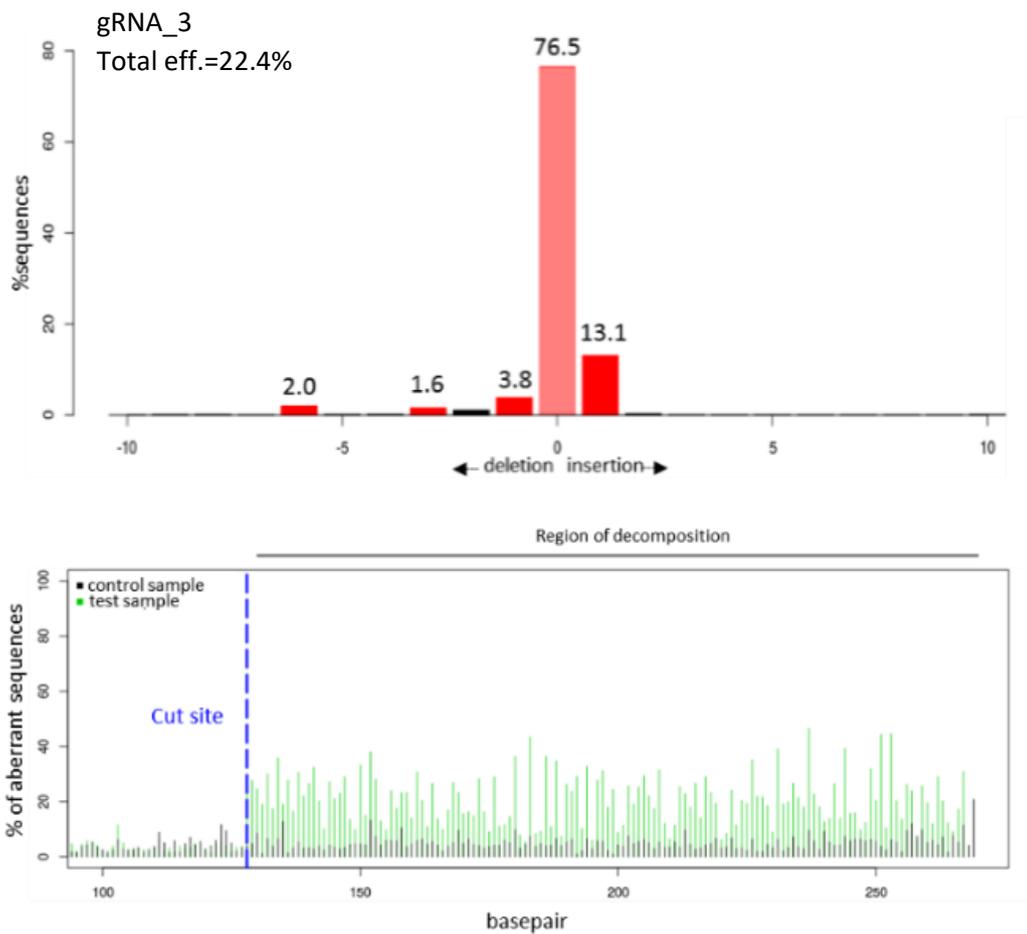
B.

Quality score: 66%

|             | Mismatches | Seed region    | PAM |        |
|-------------|------------|----------------|-----|--------|
| On-target   | TGGCTGTC   | [ATTCAGTACTCC] | TGG | gRNA_3 |
|             | TGGCAGAC   | [ACTCAGTACTCC] | GGG |        |
|             | AGTTTCTG   | [ATTCAGTACTCC] | TGG |        |
|             | TGCTTTC    | [ACTCAGTACTCC] | AGG |        |
|             | TTGCTTT    | [ATTCAGTACTCC] | TGG |        |
|             | TGAGGGTC   | [ATACAGTACTCC] | TGG |        |
|             | AGTTTGCC   | [TTCAGTACTCC]  | AGG |        |
|             | TAGCTTT    | [CTTCAGTACTCC] | TGG |        |
|             | CCGGTGCC   | [AGTCAGTACTCC] | AGG |        |
|             | TATCTGC    | [CTTCAGTACTCC] | AGG |        |
| Off-targets | AGGGAATC   | [AATCAGTACTCC] | AGG |        |
|             | TCTCTTA    | [TTCAGTACTCC]  | TGG |        |
|             | GGCTTG     | [ATGCAGTACTCC] | AGG |        |
|             | TAACTGAG   | [GTCAGTACTCC]  | AGG |        |
|             | AGGTC      | [ATTCTGTACTCC] | AGG |        |
|             | AGGTC      | [ATTCGTACTCC]  | AGG |        |
|             | CCGTTGTC   | [TTCAGTACTCC]  | AGG |        |
|             | TGGCTAGA   | [ATCCAGTACTCC] | AGG |        |
|             | CTGCGGTC   | [ACCCAGTACTCC] | GGG |        |
|             | TGACCATG   | [ATACAGTACTCC] | AGG |        |

C.



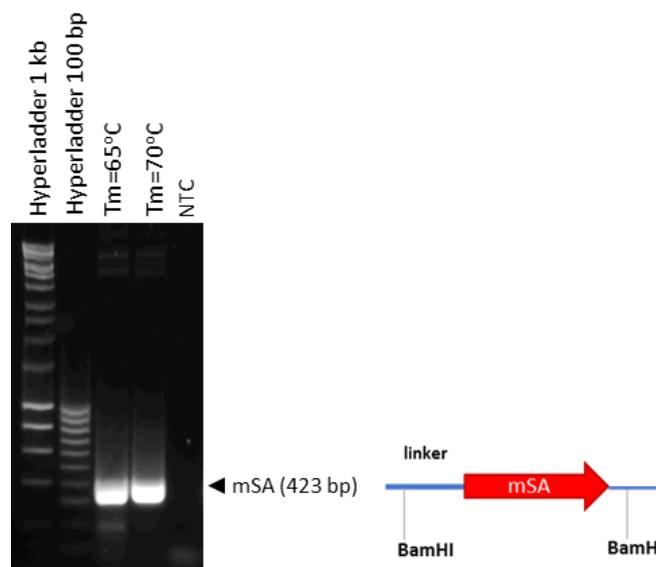


**Figure 31. Partial sequence of the ssODN\_3 designed to improve ARW741L knock-in rates. A.** The CRISPR-blocking mutations are highlighted in magenta whilst the desired GG>TA (W741L) mutation is highlighted in red. gRNA\_3 binds to the target DNA sequence (indicated by the turquoise bar) and recruits Cas9 to cleave DNA 3 bp upstream of the PAM site (red arrow head), which comprises the CRISPR target codon. The GG>TA substitutions create a unique *MseI* restriction site which is exploited for RFLP screening of clonal cell populations for detection of the W741L mutation. **B.** The top 20 off-target sites of gRNA\_3 were predicted using CCTop. Nucleotides highlighted in red indicate mismatches between the gRNA\_3 sequence and the potential off-target genomic sites. **C.** Sequencing chromatograms of AR exon 5 in CWR22Rv1 cells nucleofected with the gRNA\_3/Cas9 (px459) vector. CRISPR-induced indels (highlighted in blue) are shown as background peaks upstream of the cut site which is depicted by the red arrowhead and the dashed line (top panel). The chromatograms were analysed using TIDE to calculate the editing efficiency of the CRISPR complex by aligning and decomposing the sequences of the edited (green) versus the wild-type (black) AR exon 5 (middle panel). The intact (unedited) sequence is represented by the middle coral bar whilst  $\pm 10$  indels are indicated by the red and black bars (bottom panel).

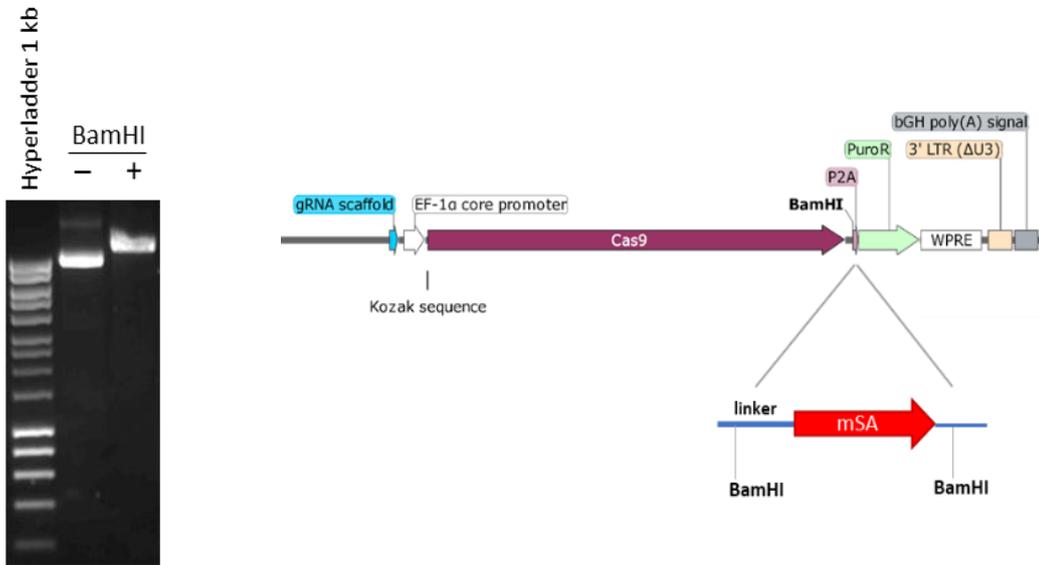
#### 4.3.9. Generation of an in-house mSA-tagged Cas9 expressing vector

A streptavidin-tagged Cas9-encoding vector was not available at the time and so was constructed in-house (as described in Figure 32) using the commercially available px459 as the recipient vector. Previous work in our lab highlighted a number of pitfalls to be avoided when generating Cas9 fusions. More specifically, the construction of an RFP-fused Cas9 expressing plasmid was attempted with the RFP insert cloned either upstream or downstream of the Cas9 ORF. Despite its in-frame insertion and detection of Cas9 by WB, no RFP was observed by fluorescence microscopy, suggesting that the presence of such a bulky moiety may hinder appropriate folding of the fused proteins. To tackle this, a linker sequence between Cas9 and RFP was added and only in the case of C terminal RFP insertion, red fluorescence was detected, suggesting blocking of the NLS signal of Cas9 (located at the N terminus) by RFP when the latter was inserted at the N terminus. These observations were taken into account to design the Cas9-mSA fusion. Hence, the mSA coding sequence was cloned at the C-terminus of the Cas9 ORF and a 48-nt linker was included between the two proteins to provide enough 3D space for them to fold appropriately preventing potential allosteric hindrance and ultimately preserving their function. Finally, gRNA\_3 was cloned into px459 (pSpCas9-mSA-2A-Puro) following *BbsI* digestion.

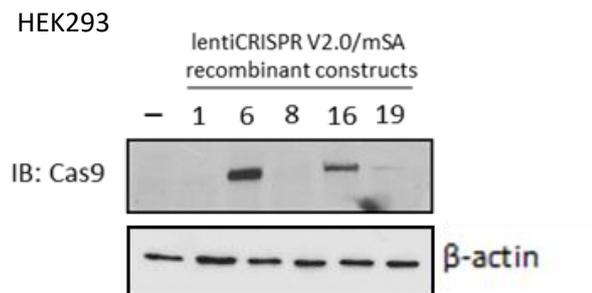
##### 1. mSA amplification



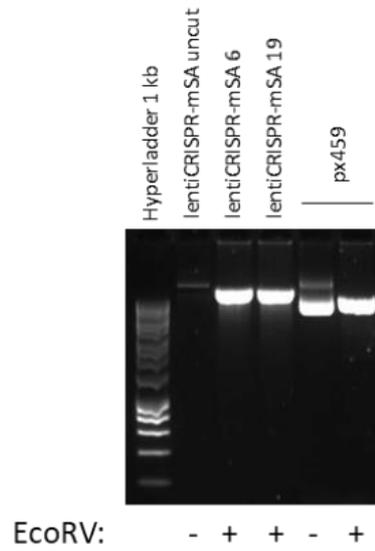
2. BamHI digestion & cloning of mSA in lentiCRSIPR V2.0 (intermediate vector)



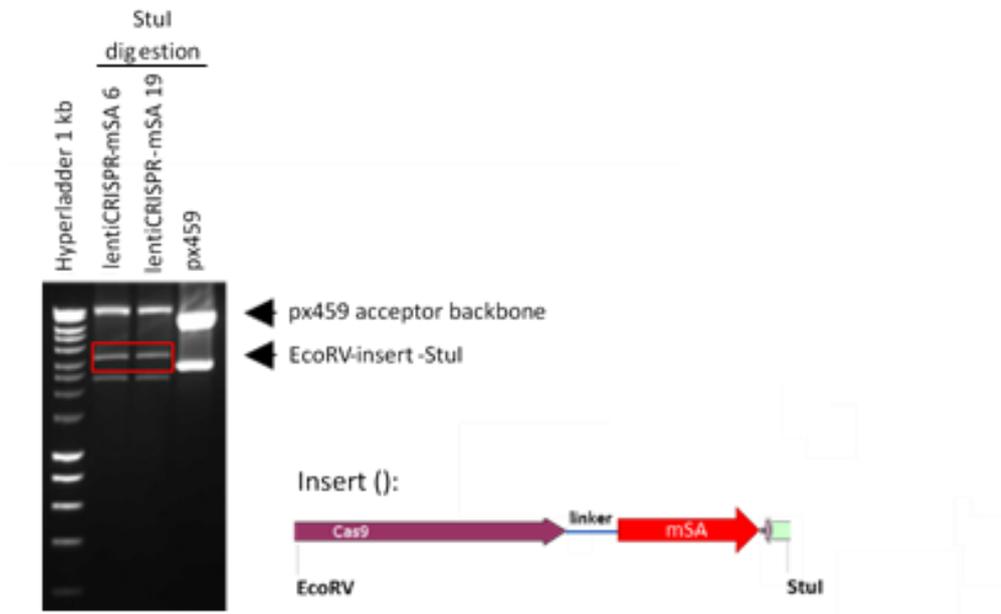
3. Validation of intact Cas9-mSA



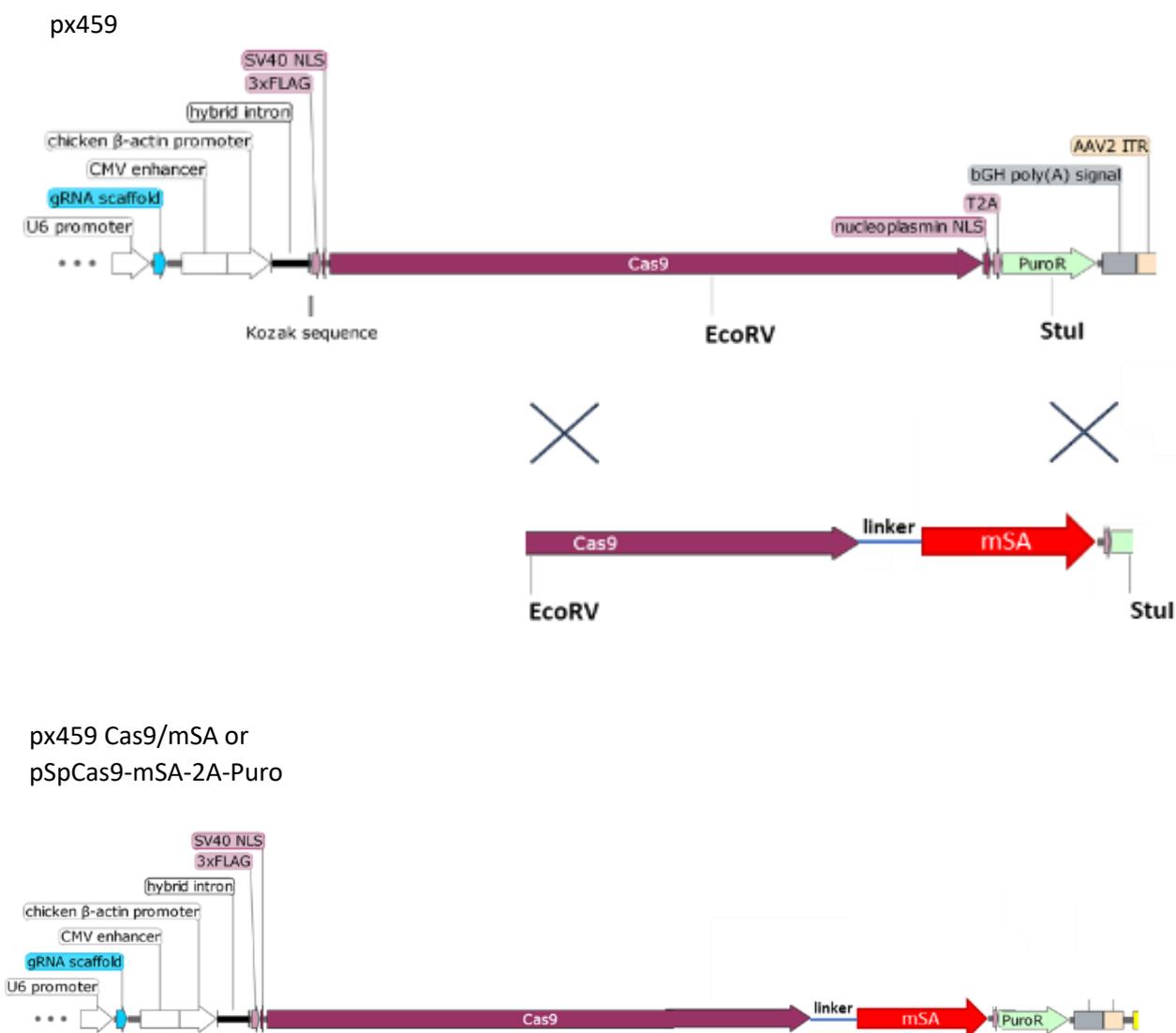
4. EcoRV digestion of recombinant lentiCRISPR V2.0/mSA



5. Concurrent Stul digestion of recombinant lentiCRISPR V2.0/mSA



6. Generation of recombinant px459  
Cas9/mSA (pSpCas9-mSA-2A-Puro) &

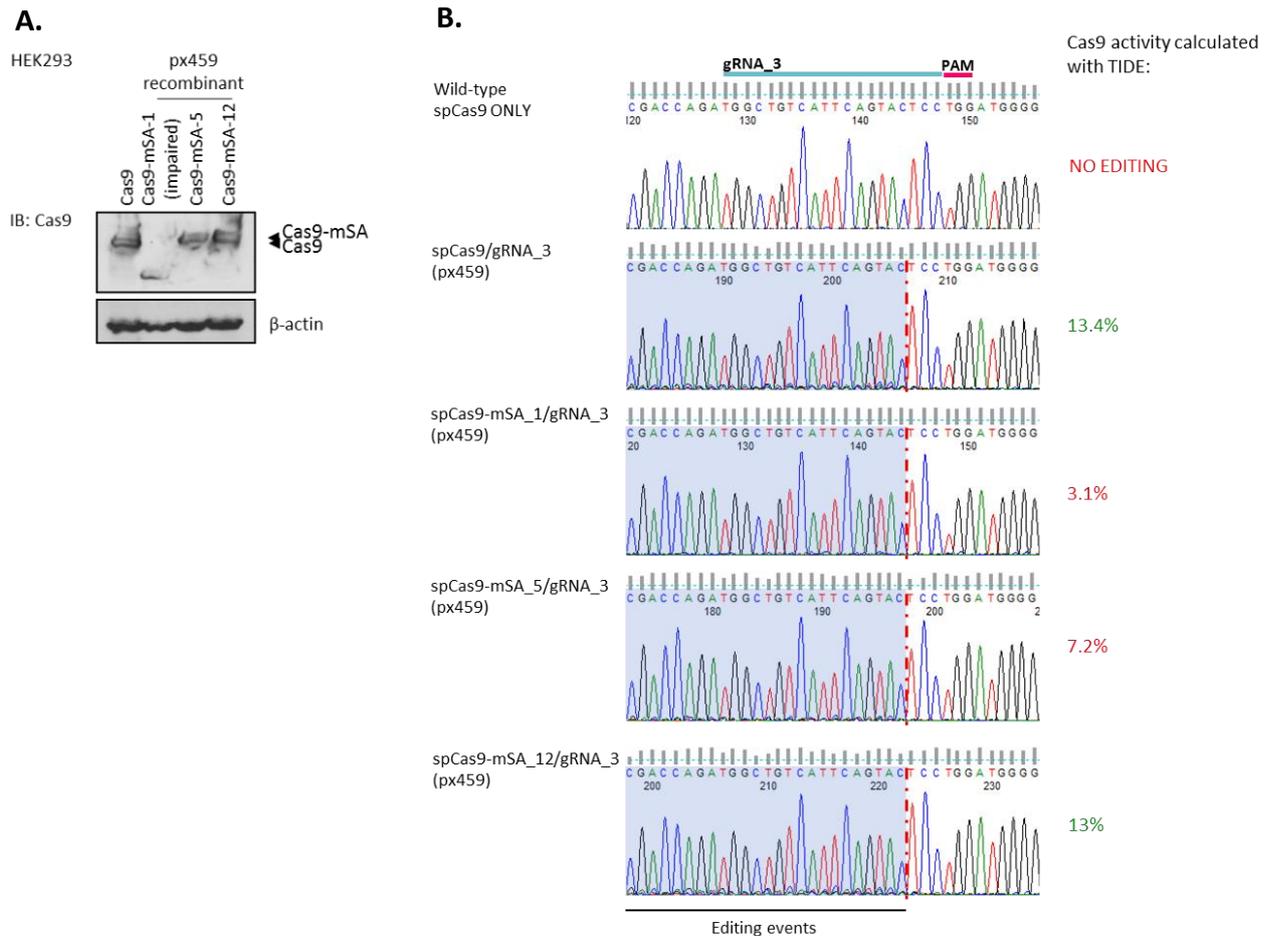


**Figure 32. Summary of the cloning strategy applied to generate the in-house px459 Cas9-mSA expressing construct.** The mSA fragment was amplified by end-point PCR using the pCS2+Cas9-mSA vector as template. A 48-nt linker was added to the 5' end of the forward PCR primer while *Bam*HI restriction sites were flanking the linker-mSA amplicon. Two different melting temperatures ( $T_m$ ) were applied to achieve high target specificity (step 1). The amplified linker-mSA insert as well as the lentiCRISPR V2.0 vector were subject to *Bam*HI digestion and were subsequently ligated to generate the recombinant lentiCRISPR V2.0/mSA construct (step 2). Intact expression of Cas9 from the new recombinant vectors was examined by WB upon overexpression in HEK293 cells (step 3). The intermediate lentiCRISPR V2.0/mSA and px459 vectors were incubated with *Eco*RV and *Stu*I (steps 4 and 5) and the Cas9-linker-mSA fragment derived from the digestion reactions of lentiCRISPR V2.0/mSA was ligated into the px459 backbone to generate the final px459 Cas9-mSA construct (step 6).

#### **4.3.10. Validation of successful Cas9-mSA expression**

After all cloning steps were complete, expression of Cas9-mSA was assessed by transfecting the resultant gRNA\_3/Cas9-mSA (px459) constructs in HEK293 cells. Three recombinant constructs were tested (Cas9-mS-1, Cas9-mSA-5 and Cas9-mSA-12). Sequencing of the constructs was performed to verify whether the linker-mSA fusion was in-frame with Cas9 (Appendix B). The coding frame of Cas9-mSA-1 was severely impacted as evidenced by the short Cas9-like isoform detected by WB analysis (depicted by \*) using a 15% acrylamide gel. On the contrary, Cas9-mSA-5 and Cas9-mSA-12 remained intact and a slight band upshift, which corresponded to the additional coding sequence of the linker-mSA (~15-20 kDa) was observed (Figure 33A).

The editing efficiency of each gRNA\_3/Cas9-mSA complex was compared to that of the original gRNA\_3 /Cas9 complex to examine whether the mSA fusion impaired the ability of Cas9 to cleave target DNA. The constructs were therefore transfected into HEK293 cells and upon DNA extraction, *AR* exon 5 was amplified and sequenced and the editing efficiency of each gRNA\_3/Cas9-mSA complex was assessed using TIDE (Figure 33B). As expected, Cas9-mSA-1 was incompetent and no editing events were observed. Surprisingly, Cas9-mSA-5 demonstrated reduced editing efficiency of about 50% despite its intact coding sequence and finally Cas9-mSA-12 retained full activity (13%) which was comparable to that of the original, unmodified Cas9 (13.4%) and therefore was picked for the downstream knock-in experiments.

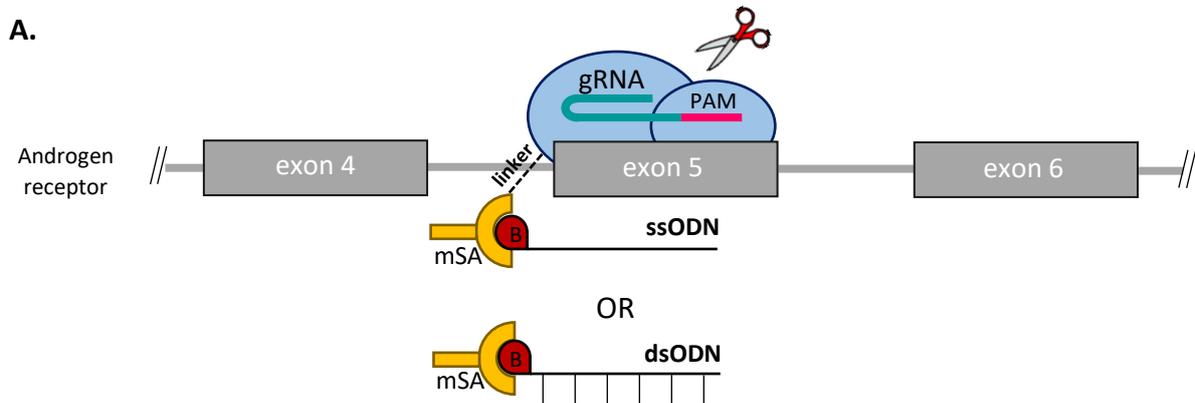


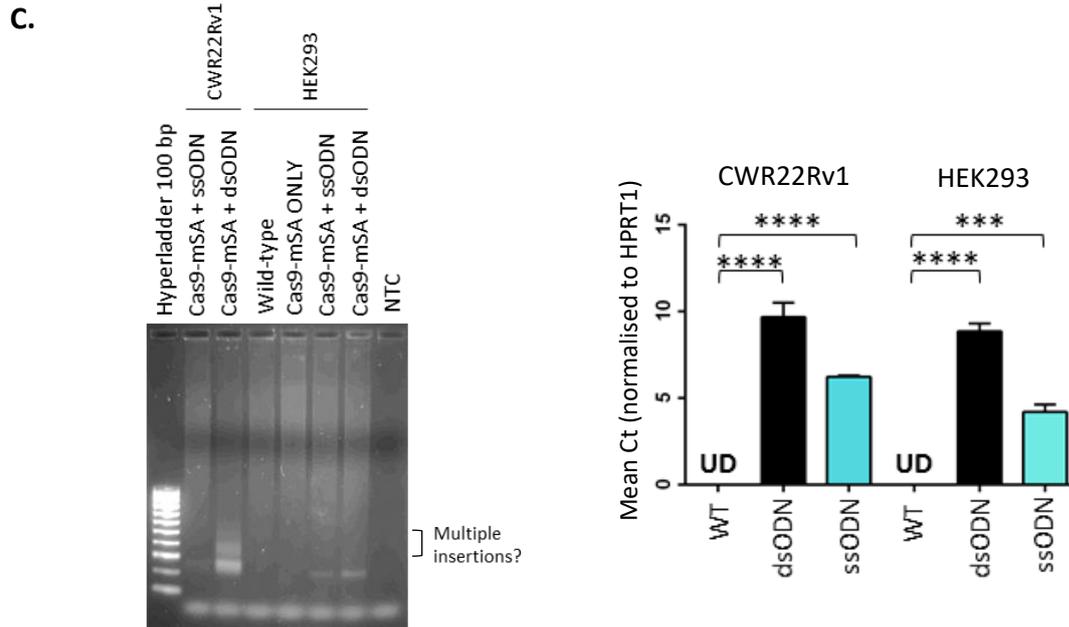
**Figure 33. Impact of the mSA fusion on Cas9 expression and function in HEK293 cells.** **A.** Expression of recombinant Cas9-mSA derived from the px459 vector was assessed by WB using an anti-Cas9 antibody. HEK293 cells were transfected with three different recombinant px459 Cas9-mSA (1, 5 and 12) and the original px459 Cas9 expressing constructs. Cells were harvested 48h post-transfection and resulting lysates were analysed on a 15% SDS-PAGE gel to observe mSA-fused Cas9 species. **B.** Retention of activity of the mSA-fused Cas9 was assessed by comparing the editing efficiency of the recombinant nuclease to the activity of the non-recombinant Cas9 expressed by the original px459 vector. HEK293 cells were transfected as in A. DNA extracted from CRISPR-edited cells was used as template to amplify and then sequence AR exon 5. Sequence chromatograms were analysed using TIDE. Underlying peaks indicate indels.

#### **4.3.11. Conjugation of mSA-tagged Cas9 with a biotinylated donor template yielded elevated template knock-in rates**

HEK293 and CWR22Rv1 cells were transfected and nucleofected respectively with the gRNA\_3/Cas9-mSA-12 vector alongside the two different versions of the donor template, single stranded (ssODN) or double stranded (dsODN) (Figure 34A). Upon puromycin selection, the mixed populations of CRISPR-edited cells were subject to DNA extraction followed by AR exon 5 amplification by PCR. A donor specific PCR primer was designed as shown in Figure

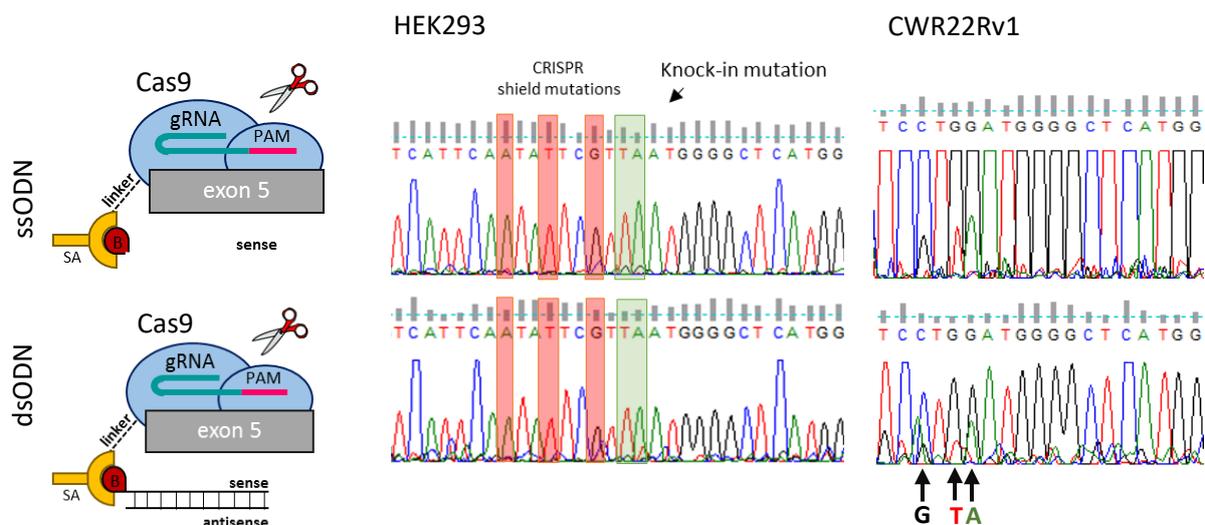
34B. Amplification using the donor specific primer under stringent PCR conditions (i.e. high primer annealing temperature) would indicate successful HDR events in the interrogated mixed cell population. Satisfying levels and more importantly, specific amplification, was observed in both cell lines. Extremely encouraging was the fact that the observed amplification was restricted to those DNA samples which were derived from the CRISPR edited cell populations and not the wild-type or the Cas9 only transfected cell populations ( ) indicating that genuine knock-in events occurred in those cell populations. Remarkably, more efficient HDR was mediated by the dsODN rather than the ssODN as evidenced by the increased intensity of the gel band as well as the increased fluorescent signal in qPCR (Figure 34C). Interestingly, a quite intense smear was observed in the case of dsODN which could be explained by potential multiple insertions of the template in the target locus.





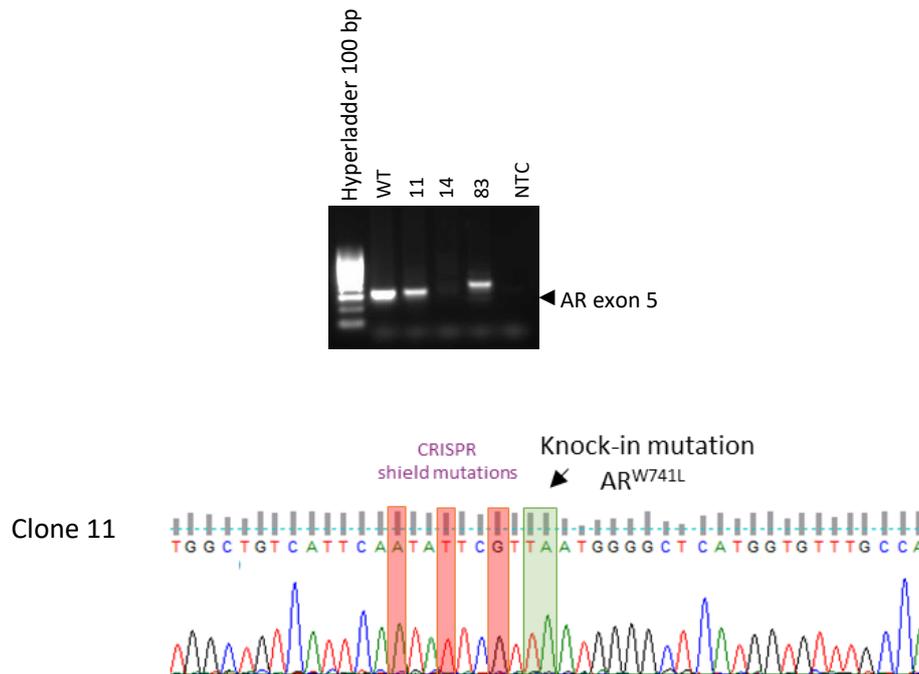
**Figure 34. Detection of HDR events in CWR22Rv1 cells using a biotinylated donor template and mSA-tagged Cas9.** **A.** Graphical representation of the streptavidin (mSA)-tagged Cas9 in conjunction with the single and double stranded donor templates. **B.** A donor specific primer was designed for screening and detection of knock-in positive CWR22Rv1 cell clones. Mismatches (depicted by loops/hairpins) between the primer and the wild type AR exon 5 sequence would prevent amplification from unedited cell clones. In contrast, perfect annealing of the primer and amplification would only occur in the presence of the knock-in mutations in the PCR template sequence. **C.** HEK293 and CWR22Rv1 cells were transfected with Cas9-mSA + SSODN\_3, Cas9-mSA + dsODN\_3 or Cas9-mSA ONLY for 48h prior to DNA analysis by end-point (left panel) and quantitative real-time PCR (right panel) using the donor template specific primer. Resultant PCR products from wild-type and CRISPR edited cells were analysed on a 2% agarose gel alongside a 100 bp DNA marker for size reference.

The precise edits introduced by Cas9 were assessed by Sanger sequencing, whereby AR exon 5 amplicons derived from the HDR positive cell populations demonstrated a spectrum of mutations. However, it was the first time that apparent knock-in events were observed in the mixed CRISPR-edited cell populations. Despite the background noise, distinct, sharp peaks were present in the sequence chromatograms and corresponded precisely to the desired W741L base substitutions (TGG>TTA) as well as the CRISPR blocking mutations in both cell lines (Figure 35). HDR efficiency was ~100% in HEK293 cells as opposed to CWR22Rv1 cells which demonstrated less pronounced knock-in mutations as evidenced by weaker W741L and CRISPR shield mutation peaks, but still sufficient to carry out downstream single cell cloning experiments and screening.



**Figure 35. HEK293 and CWR22Rv1 mixed cell populations were screened to identify HDR mediated donor template knock-in events.** Cells transfected or nucleofected with gRNA\_3/Cas9-mSA and ssODN or dsODN were subject to DNA extraction, AR exon 5 PCR amplification and Sanger sequencing. The CRISPR shield mutations are highlighted in red whilst the W741L mutation is highlighted in green in the HEK293 derived sequence (middle panel). Similarly, the GG to TA tandem substitution is highlighted in green, the more downstream CRISPR shield mutation in the donor template sequence is highlighted in black and all point mutations are depicted by arrows in the CWR22Rv1 derived sequence (right panel).

This observation significantly enhanced our confidence in finding a high number of AR<sup>W741L</sup> positive cells in the mixed cell population. Hence, CWR22Rv1 cells nucleofected with gRNA\_3/Cas9-mSA-12 and dsODN were cloned-out and expanded for 4 weeks in culture prior to screening by PCR amplification of AR exon 5 and subsequent Sanger sequencing. 54 clones were screened in total. CRISPR editing events and more precisely large insertions and single nucleotide indels were observed in the screened clones. Importantly, three W741L positive clones (clones 11, 14 and 83) were identified raising the incidence of knock-in from 1.2% in S1 to 5.5% using this latter approach. Two of the AR<sup>W741L</sup> positive clones contained additional indels whereas clone 11 harboured the desired W741L mutation in AR exon 5 alongside the CRISPR shield mutations without any additional CRISPR-induced mutations upstream or downstream of the target codons (Figure 36).



**Figure 36. Identification of AR<sup>W741L</sup> positive CWR22Rv1 single cell clones.** DNA was extracted from CWR22Rv1 single cell clones and PCR amplified AR exon 5 sequences were analysed on a 2% agarose gel (representative clones are shown). Amplicons were gel-purified and screened for detection of the W741L mutation. Sequence chromatogram of the AR<sup>W741L</sup> positive clone 11. Knocked-in CRISPR shield mutations are highlighted in red. The intended tandem W741L mutation is highlighted in green.

After multiple rounds of optimisation experiments an AR<sub>W741L</sub> knock-in CWR22Rv1 derivative cell line was finally generated (clone 11). However, further validation experiments are required to establish whether it will comprise a reliable cell line model for studying the role of the AR<sub>W741L</sub> mutation in CRPC. CWR22Rv1 cells express multiple truncated isoforms of AR alongside AR-FL. Hence, in order to attribute the observed phenotypes specifically to mutant activity, AR variants will need to be transiently depleted by siRNA knockdown. This approach may not provide the best strategy to study AR-FL mutations, however it is a proof-of-principle study and highlights the potential and background knowledge to more efficiently generate knock-in mutations in genes of interest, i.e. regulators of AR which play a key role in CRPC.

#### 4.4. Discussion

Incurable PC as well as poor understanding of the molecular mechanisms that drive PC progression highlight the yet unmet need for improved and physiologically relevant models to study AR signalling in PC. So far, research around AR mutations which confer resistance to pharmacological agents has been limited to overexpression of the mutant of interest upon lentiviral transduction of LNCaP cells (O'Neill et al., 2015; Sun et al., 2006). This approach is very often associated with high levels of cytotoxicity, mislocalisation of the protein of interest due to overexpression and biased read-outs due to random genomic integration of the transgene (Thomas et al., 2003). CRISPR has overcome these challenges and has offered scientists the opportunity to generate theoretically more efficiently the desired preclinical cell line and mouse models in order to better understand the molecular background of the studied disease.

We aimed to use the technology to generate a CRISPR engineered PC cell line which would permanently express the AR<sub>W741L</sub> mutant following knock-in of a donor template, which encompassed the point mutation, in the genome of CWR22Rv1 cells via homologous recombination. This way endogenous levels of the mutant would be produced avoiding all the risks associated with protein overexpression. Ideally an LNCaP AR<sub>W741L</sub> derivative cell line would have been generated alongside the CWR22Rv1 counterpart. LNCaP cells rely exclusively on AR-FL for growth and they would comprise a more suitable model to study AR-FL mutations. However, various limitations such as poor transfection efficiency and delivery of Cas9 in combination with limited clonogenicity of LNCaP cells did not allow the generation of an AR<sub>W741L</sub> mutant LNCaP cell line and hence our study was limited to CWR22Rv1 engineered cells.

The technology was relatively novel at the start of this project and is still fast-developing. A plethora of strategies can be attempted in order to achieve the ultimate goal. Each strategy, however, requires often long and multiple optimisation steps and each of them may comprise a significant challenge.

Major determinant of success in carrying out CRISPR knock-in projects is the careful design of the chosen strategy starting with gRNA and donor template design, the method of Cas9 delivery as well as the stability of the different CRISPR machinery components.

In our experiments, plasmid-derived Cas9 and transcribed gRNA from an all-in-one plasmid vector were used to edit the target *AR* exon 5 sequence. This kind of approach is linked with an increased risk of off-target mutations introduced by promiscuous activity of Cas9 due to high expression levels and prolonged presence of Cas9 (Liang et al., 2015). The nuclease was indeed present in our mixed cell population even 2 weeks post-transfection, albeit at trace levels. Although CRISPR/Cas orthologs, such as Cpf1 nickases are proposed to reduce off-target effects due to their compromised nuclease activity, they are associated with poorer editing outcomes and so were not preferred in our experimental set-up whereby high editing efficiency was desired (Bin Moon et al., 2018).

It is of note that the majority of the predicted off-target sites were located within intronic regions of the genome and mutations in these genomic sequences would not, at least in theory, cause any deleterious implications in protein function. In addition, intronic regions and subsequently off-target sites are associated with heterochromatin and it has been suggested that Cas9 has limited, if not eliminated accessibility to these tightly structured chromatin regions due to high abundance of nucleosomes (Horlbeck et al., 2016). For those off-target sites identified within gene bodies, caution is required. Several studies have highlighted the importance of specific targeting by Cas9 and have examined its promiscuity by interrogating whole genome sequences performing unbiased deep sequencing. It has been consistently demonstrated that a single mismatch between the guide RNA and the off-target sequence can be tolerated by Cas9. However, Cas9 activity significantly drops when two mismatches are present and finally is completely eliminated when more than two mismatches occur, irrespective of their position within the otherwise complementary sequences (Anderson et al., 2015; Hsu et al., 2013). In keeping with the latter finding, structural studies of the entire CRISPR complex in mammalian cells have indicated that perfect matching between the qRNA and the target sequence is required in order for Cas9 to transiently switch from its loose inactive conformation to its DNA-docked active one, which is compatible with its editing inducing conformation (Sternberg et al., 2015). Given these criteria, the gRNAs used in this study should not raise any concerns regarding Cas9 specificity as the number of mismatches between the gRNAs and the potential off-target sites was either three or more than three, which is not permissive for CRISPR editing.

There is always a balance between low off-target risk and high editing efficiency. Since gRNA off-target effects were not a major concern given the number of mismatches within their sequence, editing efficiency was the most relevant variable to look at. According to a recent study gRNA activity is dictated by TT- and GCC-rich motifs located at certain positions within its sequence. More specifically, presence of such motifs at positions 17-20 of the gRNA sequence blocks editing (Graf et al., 2019). The gRNAs used in this study were analysed for their sequence content to confirm that they did not harbour any of these two inhibitory motifs and hence were sufficiently potent to induce Cas9 recruitment and subsequent gene alterations.

There were definitely alternative, more promising approaches which could have been used and could have potentially resulted in higher editing and HDR efficiencies. More specifically, a direct comparison between plasmid derived Cas9 and purified Cas9 protein in complex with a chemically modified and hence more stable gRNA (ribonucleoprotein complex, RNP) was carried out. In keeping with a previous report (Okamoto et al., 2019), the RNP complex demonstrated remarkable editing efficiency over the plasmid derived Cas9 and the HDR rates were also significantly elevated when RNP was used (Appendix I). However the main limitation in this kind of approach is the lack of a fluorescent or antibiotic marker to allow selection of the transfected and potentially CRISPR edited cells, which is critical when the transfection efficiency of the host cell line is low and only a limited number of cells in the mixed population will eventually contain the desired edit.

It is now well established that the cut-to-mutation distance must be minimised to achieve high HDR rates. Proximal to the cut site edits are more likely to be incorporated compared to distal edits as shown previously (Paquet et al., 2016; Wang et al., 2016). That was also evident in our initial set-up (gRNA\_2 and ssODN\_1), whereby the CRISPR-blocking mutations located right upstream of the cut site (proximal edits) were incorporated in 3/35 single cell clones as opposed to the desired mutation which was located 24 bp downstream of the cut site (distal edit) and was only found in 1/35 cell clones. The observed partial incorporation of the ssODN could be explained by multiple rounds of annealing between the homologous regions of the ssODN and the target sequence as suggested by other groups and highlights the complexity of the repair mechanisms that can take place upon DSB (Boel et al., 2018; Wang et al., 2016). To avoid incomplete HDR (knock-in of exclusively proximal edits) and enhance insertion

frequencies of both proximal and distal edits, silent mutations between the distal edits and the cut site should be introduced to potentially reduce homology with the target locus and prevent re-annealing and exclusion of the distal edit (Paix et al., 2017). Despite the apparent benefits of this approach the incorporation efficiency of distal edits remained higher than that of proximal edits. Therefore design of a new gRNA and ssODN was preferred (gRNA\_3 and ssODN\_3), whereby the gRNA directed Cas9 to cut 3 bp upstream of the TGG target codon and hence the cut-to-mutation distance was then limited to 4 bases, making the desired mutation a proximal edit. In this case, the target codon was also the PAM site which would be mutated to TTA upon knock-in of the ssODN and would prevent alongside the other proximal CRISPR- blocking mutations re-binding of Cas9 and reversal of knock-in.

This approach was combined with the realisation that high concentrations of the ssODN are desired around the cut site to favour knock-in to the target locus (Carlson-Stevermer et al., 2017) and therefore a Cas9-mSA expressing plasmid was constructed in the lab.

Monomeric streptavidin (mSA) was fused through a linker sequence of 48 bases to the C terminal end of Cas9. The C terminus was preferred over the N terminus since the latter encompasses the NLS sequence required for Cas9 import to the nucleus and hence potential risk of NLS masking by the inserted mSA peptide was prevented. A combination of short cut-to-mutation distance and increasing the availability of the ssODN around the DSB by exploiting the high affinity between biotin-streptavidin finally led us to the desired outcome and a CWR22Rv1 AR<sub>W741L</sub> mutant cell line derivative was generated. This approach is gaining much more popularity now with different combinations of covalent modifications of Cas9, gRNA and donor template (Aird et al., 2018; Gutierrez-Triana et al., 2018; Liu et al., 2017; Roche et al., 2018)

The most common limiting factor when one performs CRISPR knock-in experiments is the occurrence of error-prone NHEJ mediated repair. NHEJ is the most prominent DNA repair mechanism and the most rapid cellular response triggered upon DSB. It occurs in a frequency of almost 99% and therefore outnumbers precise HDR events which are limited to <1%. HDR only occurs during a short period of the cell cycle, in the G2 and S phases of the cell cycle and that further limits its prevalence. HDR efficiencies have been improved but are still far from optimal to enable easy knock-ins.

Even with improved HDR rates though, NHEJ events were still persistent, occurring in more than 50% of interrogated cell clones. Most clones appeared to have short insertions or deletions of 1-5 nucleotides around the cut site which is consistent with published data (Allen et al., 2018). Surprisingly, NHEJ events were observed even upon depletion of major components of the pathway highlighting the functional redundancy of NHEJ mediators in the insulted cells, an observation which contradicts previous reports which claim remarkable improvement of precision editing by blocking the NHEJ pathway (Li et al., 2018).

Finally, even when HDR occurs and the precise knock-in template gets successfully incorporated to the host genome, it is accompanied by imprecise monoallelic and/or biallelic NHEJ induced edits either upstream or downstream of the precise mutation. In our experimental strategy, it was quite apparent that clones positive for the AR<sup>W741L</sup> mutation also harboured indels, suggesting coupling of repair mechanisms, a quite common observation from different groups (Paix et al., 2017; Paquet et al., 2016).

It is apparent that multiple adjustments were required to determine the optimal conditions to favour HDR. Time-consuming and labour-intensive experiments were performed as they required manual handling of hundreds of clones and the identification of a knock-in positive clone was very much dependent on chance. Despite its challenges the project was accomplished with success and paved the way for more efficient CRISPR editing work in the lab.

## Chapter 5. Generation of an AR-FL knock out CWR22Rv1 cell line derivative using CRISPR

### 5.1. Introduction

The emergence of truncated AR proteins called AR splice variants (AR-Vs) is a common mechanism of acquired resistance in patients treated with second line antiandrogens, such as enzalutamide and abiraterone (Antonarakis et al., 2014). The most commonly identified AR-V with clinical relevance is AR-V7 which remains the best characterised due to frequent detection in clinical specimens and availability of antibodies for laboratory research. However, a plethora of different AR-Vs have been identified in cell lines, xenograft models of CRPC and PC biopsies (Cao et al., 2016). They drive PC growth in the presence of antiandrogens and their expression is associated with shorter patient survival. There are two proposed mechanisms which result in the generation of these shorter versions of the full-length receptor: i) intragenic rearrangements, such as deletions and tandem duplications of exonic regions in the AR locus (Li et al., 2011) and ii) alternative splicing of cryptic exons which was shown to be induced by ADT (Dehm et al., 2008).

Co-expression of AR-Vs with AR-FL is a common phenomenon observed in CRPC cell lines and clinical specimens. The major question which inevitably arises by the presence of both truncated and full-length forms is whether AR-Vs and AR-FL engage with one another to exert their function or whether they act independently to promote androgenic signals and ultimately disease progression in the absence of androgens. There is currently an ongoing scientific debate on whether AR-FL and AR-Vs associate and co-operate; with controversial findings from *in vitro* studies across different laboratories (Luo et al., 2018).

It has been shown that AR-Vs have diverged from AR-FL and are capable to regulate their own distinct transcriptome to drive CRPC. More specifically, AR-V7 drives a cell-cycle associated gene signature without requiring AR-FL activity (Hu et al., 2012). Additionally, the CRPC relevant LuCaP 86.2 bladder metastasis derived xenograft, which lacks AR exons 5-7 due to an intragenic deletion which in turn abrogates AR-FL synthesis but favours expression of a truncated AR-V isoform, highlights that AR-FL activity may not be required in CRPC and AR-V activity is sufficient to sustain and drive tumour growth (Li et al., 2012). This is in contrast with

an interdependent relationship between AR-Vs and AR-FL proposed by Cato *et al.* and Watson *et al.* (Cato *et al.*, 2019; Watson *et al.*, 2010).

Physical association between the AR-FL and different variants such as the AR-V7 and ARv567es has been reported previously on and off chromatin, suggesting an interdependent relationship (Cao *et al.*, 2014; Cato *et al.*, 2019; Sun *et al.*, 2010). However, there is no clear picture that interaction between the truncated and full-length receptors occurs as there is much contradictory data to suggest no heterodimerisation events. More precisely, AR-V7 and AR-FL co-occupancy of the PSA promoter has been shown in ChIP experiments in CWR22Rv1 cells grown in castrate conditions (Cao *et al.*, 2014). However, the Dehm group, as well as our in-house findings, suggest that such an interaction is not occurring as no heterodimers could be detected by immunoprecipitation experiments in the same cell line. In addition, the identification of unique AR-V7 binding sites and the distinct transcriptomes support the notion that AR-FL is not essential for AR-V function. On the contrary, prostate cancer cells can survive and proliferate even when a single AR-V is expressed (Nyquist *et al.*, 2013).

The majority of groups are focusing on AR-V7 simply because it is frequently detected in the clinic and specific antibodies have been developed against it. Despite our limited knowledge about its precise role in CRPC, AR-V7 is now a potential clinical biomarker exploited for guiding treatment options. The Dehm group have recently developed a CWR-AD1 derivative cell line, named R1-D567, which only expresses the clinically-relevant AR-v567es receptor variant and have provided significant insight to our understanding of AR-V-driven transcriptomics and drug sensitivities in an *in vitro* setting (Nyquist *et al.*, 2013).

However, the repertoire of AR-Vs is extremely broad. The plethora of AR-Vs may imply different functions and hence contribution of each variant to disease progression. Different variants demonstrate different cellular localisation patterns depending on the presence or absence of a canonical NLS. This diversity is also associated with different responses to pharmacological agents such as taxanes. Taxanes inhibit cell mitosis by altering the dynamics of microtubules that make up the mitotic spindle. In PC, taxanes such as docetaxel and cabazitaxel have been used to block AR-FL translocation to the nucleus by impairing trafficking of the receptor on the microtubules. AR-FL binds to microtubules via its hinge region and LBD. It was therefore expected that AR isoforms that lack these domains will not be impacted by taxane treatment. In reality, treatment of LuCaP xenografts with docetaxel caused

sequestration of inactive AR-FL monomers in the cytoplasm, compromised ARv567es nuclear translocation as this splice variant retains the hinge region which is the minimum feature for binding to microtubules and finally left AR-V7 nuclear localisation unaffected due to the absence of both the hinge region and LBD (Bai et al., n.d.; Tagawa et al., 2019)

Despite the apparent diversity, a recent study has demonstrated that the AR-V7 and AR-V9 splice variants are co-expressed at comparable levels in patient-derived xenograft (PDX) tissues, clinical CRPC biopsies and circulating tumour cells (CTCs) (Kallio et al., 2018; Kohli et al., 2017). The same study highlighted that AR-V9 could sustain tumour growth and conferred resistance to abiraterone treatment, indicating the importance of this other AR-V variant in disease progression which remains understudied compared to AR-V7 (Kohli et al., 2017). In addition, analysis of CTCs from mCRPC patients has shown that different combinations of AR splice variants are expressed in CTCs (De Laere et al., 2017), further supporting the notion that AR-Vs may need to work cooperatively in order to function and orchestrate AR signalling. Hence, studying each variant individually might not be the most appropriate approach to understand their biology.

This heterogeneity and plethora of different receptor species present in pre-clinical models and clinical specimens indicates that there might be a more complex network that necessitates research into understanding AR-V combined activity, rather than focusing on a single variant at a time.

A major limitation in the study of AR-V biology, however, is the paucity of models that allow discriminate AR-V-specific functional and phenotypic read-outs that are not influenced by AR-FL. Performing either AR-FL siRNA-mediated depletion or enzalutamide treatment in AR-FL- and AR-V- expressing CWR22Rv1 and VCaP cell lines, several groups including ours have attempted to establish models for interrogating AR-V transcriptomics and co-regulator requirements (Chaytor et al., 2019; Dehm et al., 2008; He et al., 2018). Although useful, incomplete AR-FL depletion or anti-androgen-mediated inactivation in these systems is likely to compromise read-outs believed to be AR-V specific and may be a contributing factor to the controversy regarding whether AR-FL and AR-Vs have distinct transcriptional programmes. However, given that multiple AR-Vs have been detected in individual circulating CRPC tumour cells consistent with the CWR22Rv1 and VCaP cell lines, there remains a requirement to

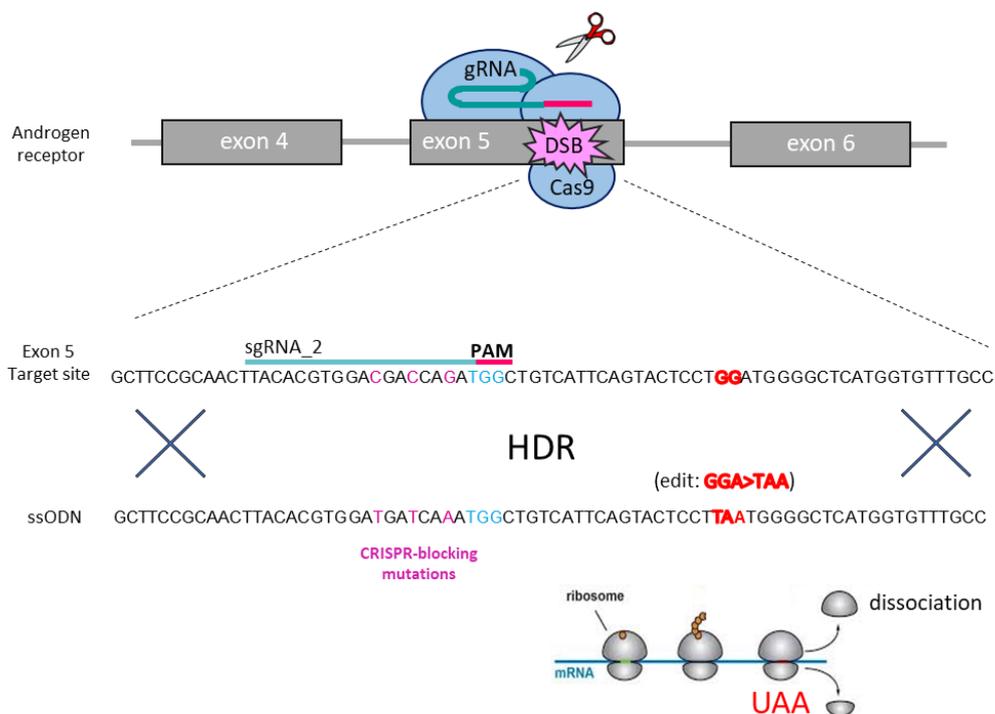
develop additional clinically-relevant models that express multiple AR-Vs in the absence of AR-FL to enable more robust studies of AR-V biology in advanced disease.

To this end, we have developed the first of its kind CRISPR-derived AR-FL knockout CWR22Rv1 cell line that retains expression of all endogenous AR-Vs making it a valuable model for the study of receptor splice variants. This new derivative called CWR22Rv1-AR-EK (Exon Knockout) is dependent upon AR-Vs for growth, is refractory to all AR-FL-targeting agents and displays a gene expression programme similar to parental CWR22Rv1 cells consistent with AR-FL and AR-V transcriptional mimicry.

## 5.2. Specific Materials & Methods

### 5.2.1. Generation of the CWR22Rv1-AR-EK cell line using CRISPR

The Cas9/gRNA\_2 pLV-U6g-EPCG vector alongside a 180 bp long single stranded donor template (ssODN\_1) (Sigma) were utilised to engineer the 5' end of AR exon 5 in CWR22Rv1 cells (Figure 37). Nucleofection, single cell cloning and downstream colony screening were carried out as described in 4.2.3. PCR amplification of AR exon 5 was performed using 50 ng of gDNA template and the high-fidelity Platinum SuperFi DNA polymerase (Invitrogen) according to manufacturer's instructions. CRISPR editing efficiency was assessed by TIDE analysis.



**Figure 37. CRISPR knock-in strategy to introduce a translational stop codon in AR exon 5 to abrogate AR-FL expression.**

### 5.2.2. CRISPR off-target analysis

Prediction of off-target effects was performed using the CRISPR design tool developed by the MIT as well as the CCTop online tool developed by the University of Heidelberg. The top 20 off-target sites were predicted based on nucleotide complementarity. The DNA sequences of potentially Cas9-insulted genes were downloaded from NCBI. The binding sites of gRNA\_2 were identified within the DNA sequences of these genes. PCR primers spanning those sites

were designed (Table 15) and the off-target sites were then amplified using the high-fidelity Platinum SuperFi DNA polymerase (Thermo Fischer Scientific) according to manufacturer's instructions. 50 ng of gDNA were used as template input and resultant amplicons were sequenced to rule out any unintended CRISPR induced mutations.

**Table 15.** Primers used in end-point PCR for CRISPR off-target analysis.

| Oligo Name    | Forward sequence (5'→3') | Reverse sequence (5'→3') |
|---------------|--------------------------|--------------------------|
| OFF1 (COL1A2) | TGTCTATATGCAATGGGCTTG    | CGAGCTGGGTTCTTTCTAGAG    |
| OFF2 (KMT2B)  | GAAGTTCTACGATGGGAAGGTG   | GCAGCTTGTTGCTGGCAT       |
| OFF3 (SLC7A8) | TTGAGGTTGCAATCCTACTGAG   | AGCTCAGGTGACTGACAATCC    |

The genomic regions surrounding AR exon 5 were also PCR amplified (see primer sequences in Table 16) and interrogated by Sanger sequencing.

**Table 16.** Primer sequences used to amplify (AR\_1) and sequence (AR\_2) the genomic region of AR intron 4 to intron 5.

| Oligo Name | Forward sequence (5'→3') | Reverse sequence (5'→3') |
|------------|--------------------------|--------------------------|
| AR_1       | TCAAACCAGATCGTAGCTGGAA   | AGAGAGCGGTAGAGGTACTCC    |
| AR_2       | GTGAAAAGCTGCTGTTTATGTGG  | TGACGCAGAAGAGCTGAGA      |

### **1.5. RNA sequencing and transcriptomic analysis**

2x10<sup>5</sup> CWR22Rv1 C34 cells were transiently transfected in triplicate for 48 hours with control or AR-targeting siRNAs in steroid depleted RPMI media (Gibco) prior to RNA extraction using the RNeasy Mini Kit (Qiagen) according to manufacturer's instructions. 500 ng RNA of each triplicate sample was subject to QC (Bioanalyzer, Agilent) and subsequent library preparation using the TruSeq Standed mRNA library prep kit (Illumina) (performed by Otogenetics, Atlanta, USA). Resultant libraries were subject to paired-end sequencing (100-125) on an

Illumina HiSeq 2500 sequencer, generating an average of 30 million reads per sample. Each data-set was mapped against the reference human genome (GRCh37/Hg19 genome assembly) utilising Star and then analysed with HTSeq to extract counts and DESeq2 (R package) to perform differential gene expression analysis between pooled control and AR-depleted samples. Differentially-expressed genes were annotated using a fold change threshold of 1.5 between control and AR-V knockdown arms. False discovery rate (FDR) threshold was set at 0.01; hence up- and down-regulated genes were identified as those with  $FDR \leq 0.01$  and respective FC of  $\geq 1.5$  and  $\leq -1.5$ .

### **5.2.3. Data Availability**

RNA-seq data has been deposited into NCBI Gene Expression Omnibus (GEO) (accession number: GSE126306).

### **5.2.4. Cell proliferation and Colony Formation assays**

$4 \times 10^6$  CWR22Rv1 and CWR22Rv1 C34 cells were seeded in 96-well plates on day 1 and were transfected with 25 nM of control or AR-V targeting siRNAs for 72h. A second round of siRNA transfection was performed at  $t=72h$  and cells were allowed to grow for an additional 48h. At end-point, proliferation was assessed using Sulforhodamine B (SRB) assays (as described in Chapter 3.15.1). For colony formation assays, CWR22Rv1 C34 cells were transiently transfected with either control or AR-targeting siRNAs for 48 hour prior to being subject to 2 Gy of ionizing radiation and re-seeded at densities of 500 and 1000 cells/well in 6-well plates (Corning) for two weeks. Colonies were fixed with 10% neutral buffered formalin solution (Sigma) for 30 min and were subsequently stained with 0.01% (w/v) crystal violet before counting using an automated colony counter.

### **5.2.5. Statistics**

Unless stated otherwise, graphical data shown in each figure represents the mean of three independent experiments and error bars indicate +/- standard error of the mean (SEM). For analysing the effect of siRNA-mediated knockdown or PARP inhibitor treatment on AR-mediated gene expression, chromatin enrichment and cell viability by qRT-PCR, ChIP and clonogenics experiments, respectively, one-way ANOVA and two-tailed student T-tests were conducted depending on the number of variables and \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$  and \*\*\*\*  $p < 0.0001$  were classified as statistically significant.]

### 5.3. Results

#### 5.3.1. CWR22Rv1 cells were genetically modified to knock-out AR-FL

A CRISPR/Cas9 knock-in strategy was developed in order to introduce a translational stop codon in exon 5 of the *AR* gene in CWR22Rv1 cells to ablate native AR-FL levels while maintaining expression of all AR-Vs endogenous to the parental cell line. The all-in-one plasmid vector expressing the Cas9/gRNA\_2 CRISPR complex as well as the ssODN\_1 donor template were used to transiently nucleofect parental CWR22Rv1 cells. Individual clones were screened and *AR* exon 5 was sequenced to identify the genomic alterations introduced by Cas9. As expected, there was a large number of clones bearing indels which could potentially result in open reading frame shifts and AR-FL gene knock-out. Two cell clones, C34 and C36 demonstrated an interesting sequence pattern with a premature TAA STOP codon resting in the 5' end of *AR* exon 5 (Figure 38A). The observed sequence variation to the wild-type sequence was followed-up by WB analysis to confirm absence of AR-FL at protein level and ruled out the generation of spurious truncated AR species. Given that the stop codon was introduced in exon 5, there was potential to generate an additional AR isoform consisting of exons 1-4. To test whether that was the case, cell lysates derived from CWR22Rv1 parental cells as well as the C34 and C36 clonal populations were analysed on a 15% SDS-PAGE which would allow separation and detection of close in size protein species. An N-terminal AR antibody was used to detect all AR species. As expected, no AR-FL was detected in C34 and C36, however a band of a suspected AR species consisting of exons 1-4 appeared between the AR-Vs and AR-FL corresponding bands in C36 (Figure 38B). The band was diminished by siARex1 and siARex4 knockdown confirming that it genuinely comprised a *de novo* synthesised AR species. The size of this non-canonical AR isoform correlated with the size of the aberrant sequence (32 DNA bases or 11 aa up to the stop codon) introduced by NHEJ upon Cas9-induced DNA DSB (Figure 38A). Importantly, no non-canonical AR species were detected in C34 despite the fact that there was an additional sequence of 26 DNA bases (or 8 amino acids) upstream of the STOP codon which could have given rise to a truncated AR species similar to the one detected in C36. QPCR analysis using an *AR* exon 3 – exon 4 amplifying primer set matched the WB findings by showing severely compromised AR-FL mRNA levels in C34 cells, while elevated AR mRNA expression derived from the *de novo* AR exon1 – exon 4 species was observed in C36 (Figure 38B). Hence, the C34 AR-FL knock-out cell clone was

expanded and utilised in further validation experiments as opposed to clone C36 which was dropped from subsequent studies.

**A.**

```

WT TTACACGTGGACGACC-----AGATGGCTGTCATTAGTACTCCTGGATG
C34 TTACACGTGGACGACCCCATGGCAAACACCATGAGCCCAATAAAGGAGTACTGAATGACAGCCATTTGATCATCCACGTGTAAGTTGCGGAGATGCTGTCATTAGTACTCCTGGATG
                26 bases

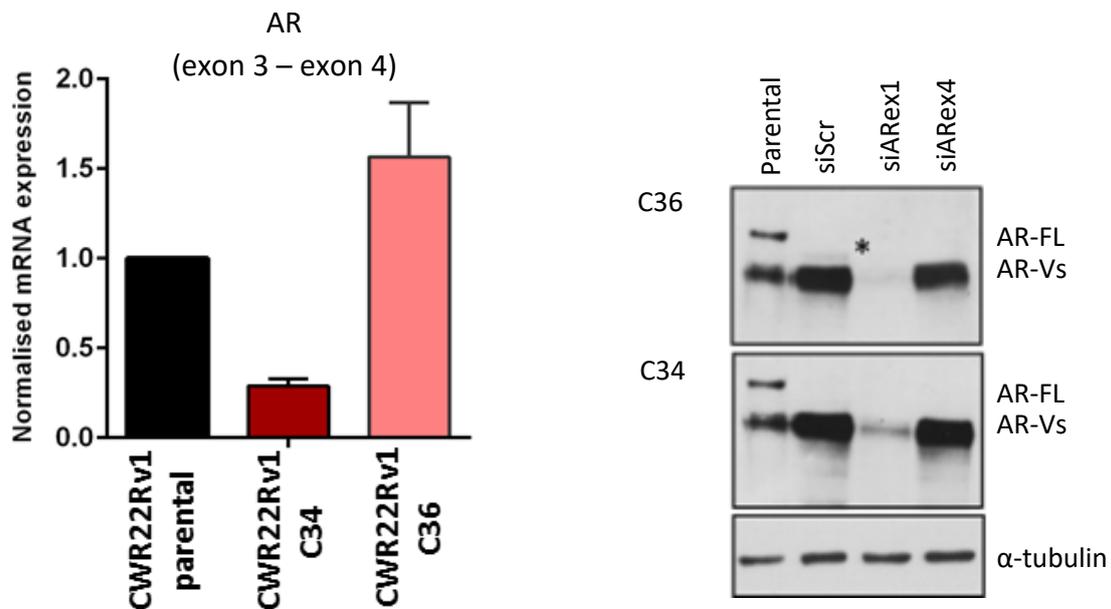
WT TTACACGTGGACGACC-----
C36 TTACACGTGGACGACCCAAAAGACCGGAAATACCCCAAGCCTTATTAACTAACCAATCAGCTCGCTTCTGCTTCTGTACCCGCGCTTTTGTCCCCAGTCTAGCCCTATAAAAAAGG
                32 bases

WT -----
C36 GGTAAGAACTCCACTGGGCGGCCAGTCATCCGATAGACTGAGTCGCCCGGGTACCCGTGTCCCAATAAAGCCTTTGTGTTTGCATCCGAAGCGTGGCCTCGTGTCTTGGGAGG

WT -----
C36 GTCTCTCATGAAAAAGTACCAGAGCTGAGTTCTAAAAGTTACAAGGAAGTTAATTAAGAATAAGGCTGAATAACACTGGGACAGGGGCCAAACAGGATATCTGTAGTCAGGCACCTG

WT -----AGATGGCTGTCATTAGTACTCCTGG
C36 GGCCCCGGCTCAGGGCCAAGAACAGATGGCTGTCATTAGTACTCCTGG
  
```

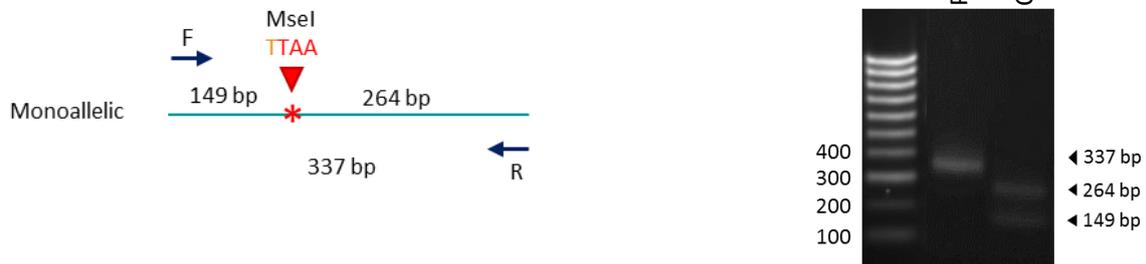
**B.**



**Figure 38. Loss of full-length AR expression in C34 and C36 clonal populations.** **A.** Sequence alignment of AR exon 5 from wild-type (WT) CWR22Rv1 cells and CRISPR engineered C34 and C36 CWR22Rv1 clones. A premature translational stop codon (TAA, highlighted in red) was introduced in AR exon 5 resulting in AR-FL gene knock-out. **B.** Loss of AR-FL in C34 and C36 was validated by qPCR using an AR exon 3 – exon 4 amplifying primer set (left panel). Parental CWR22Rv1, C34 and C36 cells were subject to control (siScr), AR exon 1 (siARex1) and AR exon 4 (siARex4) siRNA knockdown for 48 hours. Resulting whole cell lysates were analysed by WB using an N-terminal AR antibody to verify the presence of de novo AR species (depicted by asterisk) emerging from the incorporation of the TAA stop codon in AR exon 5.  $\alpha$ -tubulin was used as a loading control.



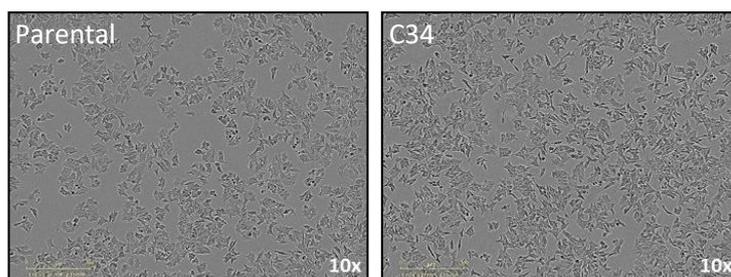
**B.**



**Figure 39. Diagrammatic representation of the DNA repair events in clone C34 upon Cas9-induced DNA DSB. A.** In theory (left panel), the entire ssODN (red bar) was expected to be integrated into the host locus via HDR. In reality (right panel), the ssODN was exploited as a template for DNA replication and the newly synthesised DNA fragment (yellow bar) was partially flipped/inserted into the AR sequence via incomplete HDR. **B.** The incorporation of the stop codon (highlighted in red) created an *MseI* cut site (depicted by the red arrowhead) in AR exon 5, which was exploited for RFLP analysis following PCR amplification of AR exon 5. Digestion products derived from parental and AR-EK CWR22Rv1 DNA were analysed on a 2% agarose gel. A 100 bp DNA marker was used for size reference.

The next step was to determine the size of the Cas9 footprint left in the *AR* locus upon editing. PCR primers were designed to amplify the flanking regions of up to 750 bp upstream and downstream of AR exon 5 (Table 16). Subsequent Sanger sequencing analysis using AR\_2Forward and AR\_2Reverse primers (Table 16) indicated no stochastic sequence insults (CRISPR-induced indels) on either side of *AR* exon 5 (example shown in Appendix C). Hence, intact and wild-type sequences around *AR* exon 5 were confirmed. The identity of the AR-FL knock-out clone was finally assessed by STR profiling. Ten core STR loci (Figure 40) were analysed (analysis was performed by Dr Liloglou's lab, University of Liverpool, UK) and showed 100% parity to the parental CWR22Rv1 cells, confirming that this new cell line derivative demonstrates a genetic background equivalent to parental cells. Finally, the morphology of clone 34 cells was not impacted by CRISPR editing. In fact, they looked similar to parental CWR22Rv1 cells when observed under the microscope (Figure 40), suggesting that no adhesion associated proteins were off-targeted by the CRISPR complex.

| Markers    | CWR22Rv1 parental | CWR22Rv1 C34 |
|------------|-------------------|--------------|
| Amelogenin | X,Y               | X,Y          |
| vWA        | 15,21             | 15,21        |
| TPOX       | 8                 | 8            |
| THO1       | 6,9.3             | 6,9.3        |
| D21S11     | 30                | 30           |
| D5S818     | 11,12             | 11,12        |
| D13S317    | 9,12              | 9,12         |
| D7S820     | 10,11             | 10,11        |
| D16S539    | 12                | 12           |
| CSF1PO     | 10,11             | 10,11        |



**Figure 40. C34 demonstrates similar phenotype to parental CWR22Rv1 cells.** Cell line authentication by STR loci profiling of CWR22Rv1 parental and clone 34 cells (left panel). Bright field images of parental CWR22Rv1 and clone 34 cells at 10x magnification taken using Incucyte live cell analysis system (right panel).

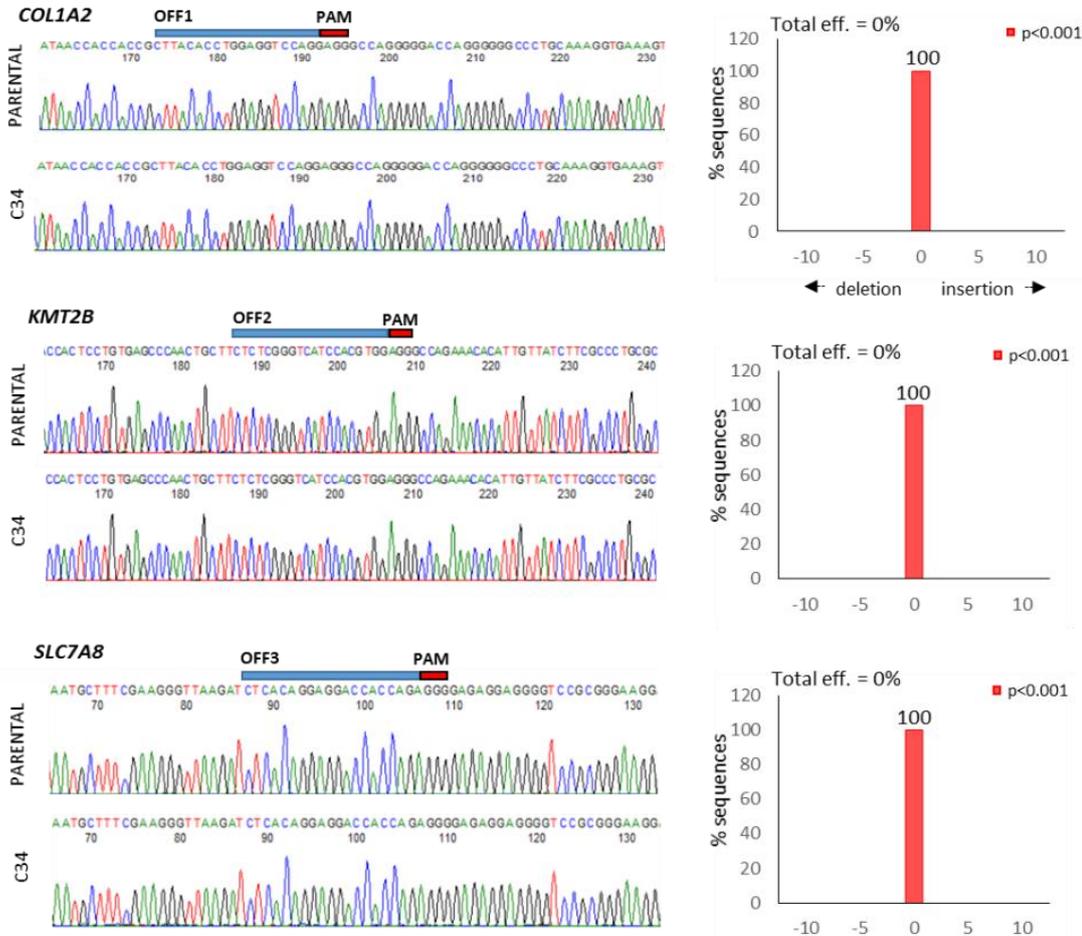
### 5.3.3. CRISPR off-target analysis shows no promiscuous activity of Cas9

Despite the high number of nucleotide mismatches (beyond 3) between the gRNA sequence and the target genomic sequence (Figure 41A), the predicted promiscuous activity of Cas9 was assessed by experimentally interrogating the top exonic off-target sites individually. More specifically, the *COL1A2*, *KMT2B* and *SLC7A8* loci were PCR amplified and the resulting amplicons were subject to conventional sequencing. Importantly, Sanger sequencing and subsequent TIDE analysis of the sequence chromatograms showed no alterations in the tested sequences when the latter were directly compared and aligned to the counterpart wild-type sequences (Figure 41B), highlighting the on-target activity of Cas9. Potential off-target genes expressed in tissues other than the prostate, as indicated by expression analysis using the Human Protein Atlas database were excluded from the analysis. Finally, intronic off-target sites were not pursued as they are predicted to have minimal if no impact on gene expression. The list of intronic Cas9 off-target regions, as predicted using the CRISPR design tool is provided in Appendix D.

#### A.

|       | MM | Mismatches<br>[Seed region] | PAM | Gene           | Gene_ID      |                                       |
|-------|----|-----------------------------|-----|----------------|--------------|---------------------------------------|
| chrX  | 0  | CTTACACG[TGGACGACCAGA]      | TGG | E AR           | NM_001011645 |                                       |
| chr14 | 4  | CCTACATG[TGGGAGACCAGA]      | GGG | E GZMB         | NM_004131    | ★ Exclusively expressed in WBC        |
| chr19 | 4  | CCTCCACG[TGGATGACCCGA]      | GAG | E KMT2B        | NM_014727    |                                       |
| chr14 | 4  | CTCACAGG[AGGACCACCAGA]      | GGG | E SLC7A8       | NM_012244    |                                       |
| chr7  | 4  | CTTACACC[TGGAGGTCAGG]       | AGG | E COL1A2       | NM_000089    |                                       |
| chr5  | 4  | CTTACAGG[TGGAATACCTGA]      | GAG | E ARHGAP26-AS1 | NR_046680    | ★ Exclusively expressed in the testis |

**B.**



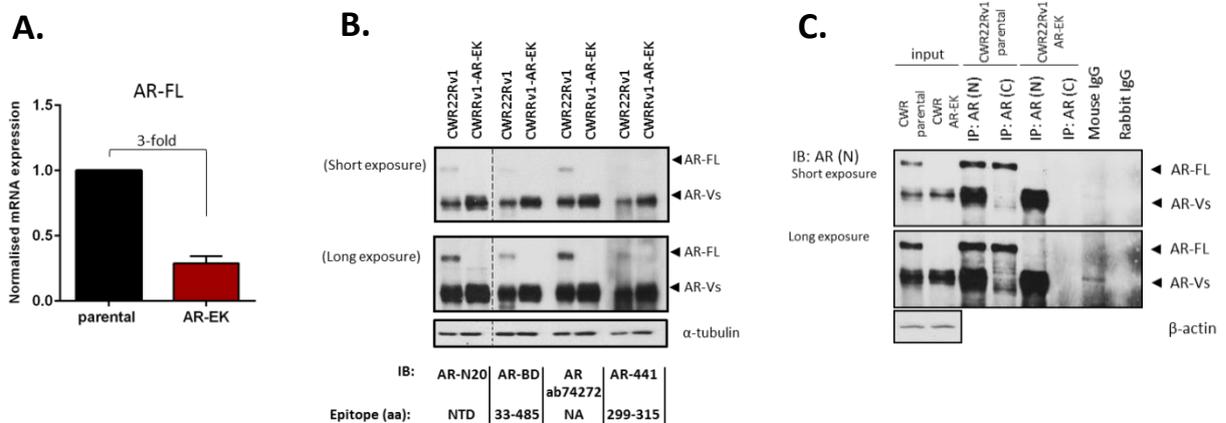
**Figure 41. Sequence analysis of the top 5 predicted CRISPR off-target genes demonstrate no off-target CRISPR activity. A.** The nucleotide mismatches between the gRNA sequence and each off-target site sequence are highlighted in red. Red asterisks depict genes expressed in tissues other than the prostate. **B.** PCR amplicons of the predicted exonic off-target sites (OFF1, OFF2 and OFF3) were analysed on a 2% agarose gel. Gel bands were excised, purified and sequenced. TIDE analysis was performed to assess the potential genomic alterations introduced upon genome editing. Blue and red bars depict the gRNA<sub>2</sub> binding sites. WBC, White Blood Cells; E, exonic.

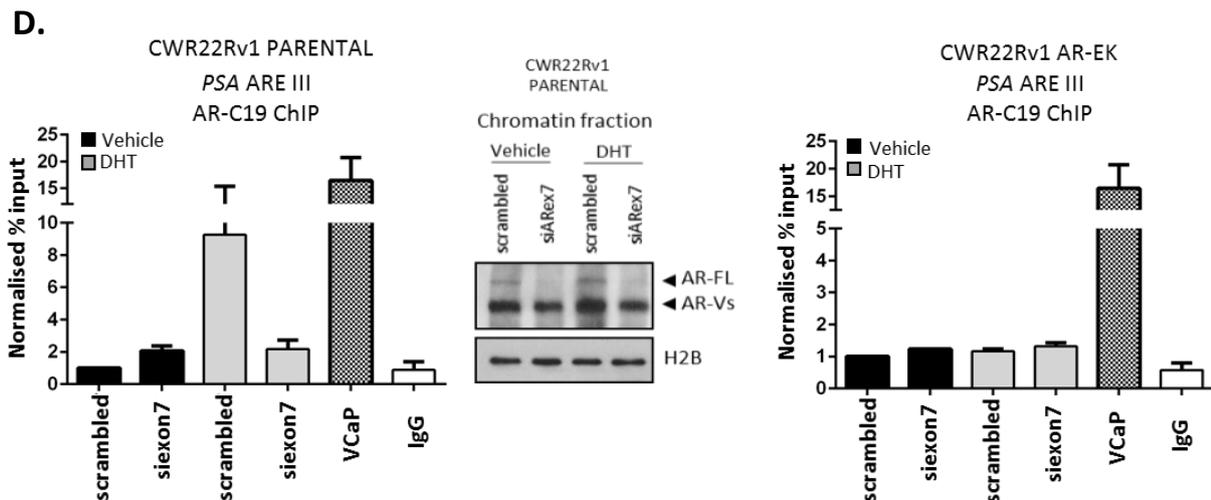
#### **5.3.4. AR-FL is no longer expressed in CWR22Rv1-AR-EK cells upon CRISPR editing**

The CWR22Rv1 derivative C34 was subsequently named CWR22Rv1-AR-EK (AR Exon Knock-out) as a result of our initial studies showing loss of AR-FL at mRNA and protein level. AR-FL mRNA levels were assessed in CWR22Rv1-AR-EK cells and CWR22Rv1 parental cells by qPCR using primers to exons 7 (forward primer) and 8 (reverse primer). AR-FL mRNA levels dropped significantly in AR-EK cells (Figure 42A) suggesting activation of a surveillance mechanism, such as the non-sense mediated mRNA decay, which drives elimination of spurious, non-physiological mRNA transcripts (Perrin-Vidoz, 2002) and hence may be responsible for diminished AR-FL mRNA as a consequence of incorporation of a premature stop codon within

AR exon 5. Interestingly, other truncated AR mRNA transcripts such as ARv5es with a premature stop codon in AR exon 6 which resulted upon exon 5 skipping have been shown to be subject to nonsense-mediated decay (Uo et al., 2017).

In addition, multiple AR antibodies recognising different epitopes within the N-terminal domain of the AR protein were utilised to detect the AR species expressed in CWR22Rv1-AR-EK cells by western blotting (Figure 42B). As predicted, no AR-FL was detected, but AR-V levels were maintained. To rule out the existence of trace levels of AR-FL protein potentially missed by WB (even after 15 min of exposure), we demonstrated that in contrast to CWR22Rv1 parental cells, no AR-FL was immunoprecipitated using a C-terminal epitope-targeting AR antibody in the CWR22Rv1-AR-EK cell line; and only AR-Vs, but not AR-FL, were immunoprecipitated using an anti-N-terminal AR antibody (Figure 42C). Importantly, ChIP experiments demonstrated that successful dihydrotestosterone (DHT)-induced enrichment of AR-FL to the *PSA* enhancer in CWR22Rv1 cells, using a C-terminal AR-binding antibody, was attenuated by an AR exon 7-targeting siRNA, but could not be replicated in the CWR22Rv1-AR-EK derivative (Figure 42D) confirming loss of AR-FL in this cell line.



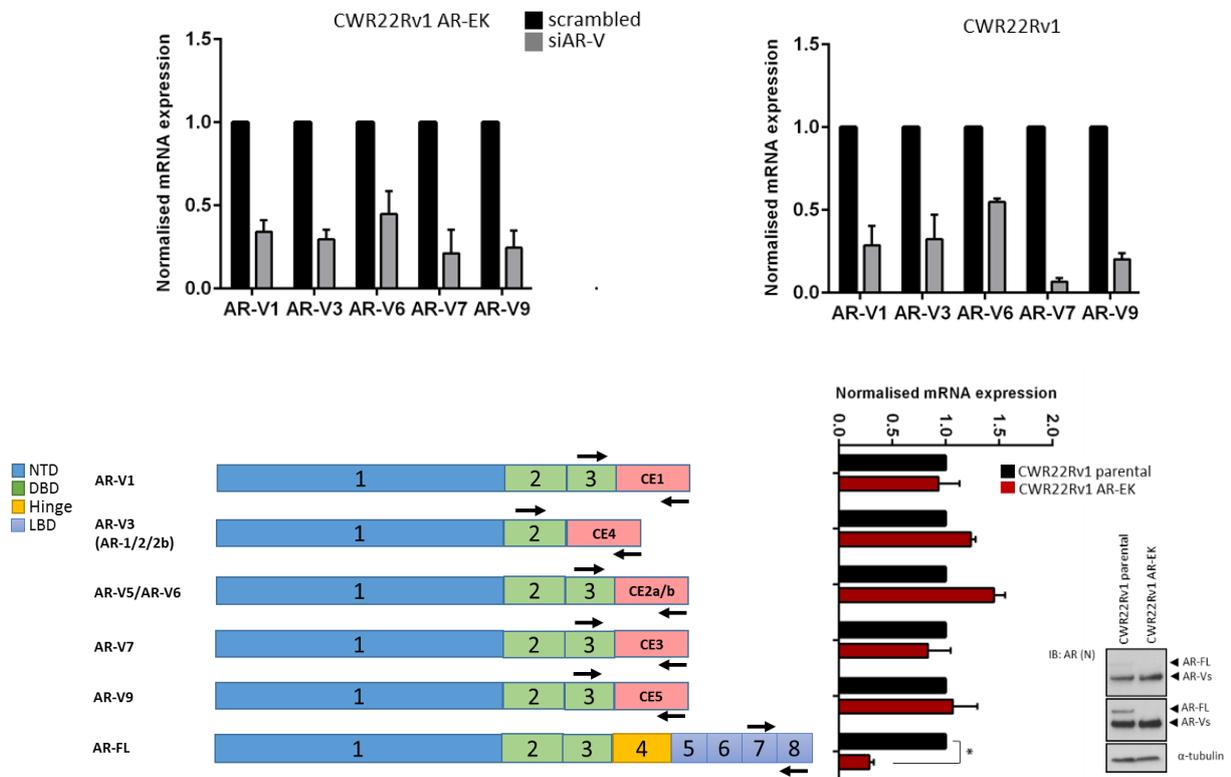


**Figure 42. Validation of AR-FL gene knock-out in CWR22Rv1-AR-EK cells. A.** Assessment of AR-FL mRNA levels in parental CWR22Rv1 and CWR22Rv1-AR-EK cells by qPCR using primers against exons 7 and 8. **B.** CWR22Rv1- and CWR22Rv1-AR-EK-derived whole lysates were subject to immunoblotting using four different AR N-terminal-targeting antibodies: AR N20 (Santa Cruz Biotechnology - discontinued), AR-BD (BD Pharmingen), AR ab74272 (Abcam) and AR-441 (Santa Cruz Biotechnology). Short and long exposures were performed to enable detection of AR-FL in CWR22Rv1 parental cells. Dashed line indicates cropped immunoblot. **C.** CWR22Rv1 and CWR22Rv1-AR-EK cells were subject to immunoprecipitation (IP) using either N- or C-terminal-binding AR antibodies and resultant immunoprecipitates were probed with an N-terminal-binding AR antibody. Input samples (10%) were ran alongside IP samples and were probed with  $\alpha$ -tubulin to demonstrate parity in protein quantities between the IP experimental arms. **D.** CWR22Rv1 (left panel) and CWR22Rv1-AR-EK (right panel) cells grown in steroid-depleted media supplemented with and without 10 nM dihydrotestosterone (DHT) were subject to either siScr or siAReX7 knockdown for 48 hours prior to chromatin immunoprecipitation (ChIP) using C-terminal-binding AR or control (IgG) antibodies. Quantitative PCR was then performed to assess enrichment of AR-FL to the PSA enhancer chromatin region. VCaP cells treated with 10 nM DHT for 4 hours were used as a positive control for enrichment of AR-FL. Data represents the mean of three independent experiments  $\pm$  SEM. Validation of siRNA-mediated AR-FL knockdown was demonstrated by immunoblotting of CWR22Rv1 chromatin fractions using anti-AR and -histone H2B antibodies.

### 5.3.5. AR-V expression remains intact in AR-EK cells

Given the objective of this work was to generate a new model to enable study of AR-Vs without AR-FL, it was extremely important to assess if we had maintained robust expression of all native AR-Vs in the CRISPR edited CWR22Rv1-AR-EK cell line. Therefore, expression of each AR splice variant was assessed individually by designing reverse qPCR primers which discriminately bind to the unique C-terminal cryptic exon of each AR-V. Importantly, the clinically relevant AR splice variants AR-V1, AR-V3, AR-V5/AR-V6, AR-V7 and AR-V9 were all expressed in the CWR22Rv1-AR-EK derivative and could be depleted by siRNA mediated

knockdown targeting a 3' UTR polyadenylation signal sequence found in all AR-Vs (Figure 43). Moreover, the mRNA levels of each AR-V was compared to that of the parental cell line and, importantly they remained unchanged upon CRISPR editing, showing comparable levels of expression to parental cells (Figure 43).

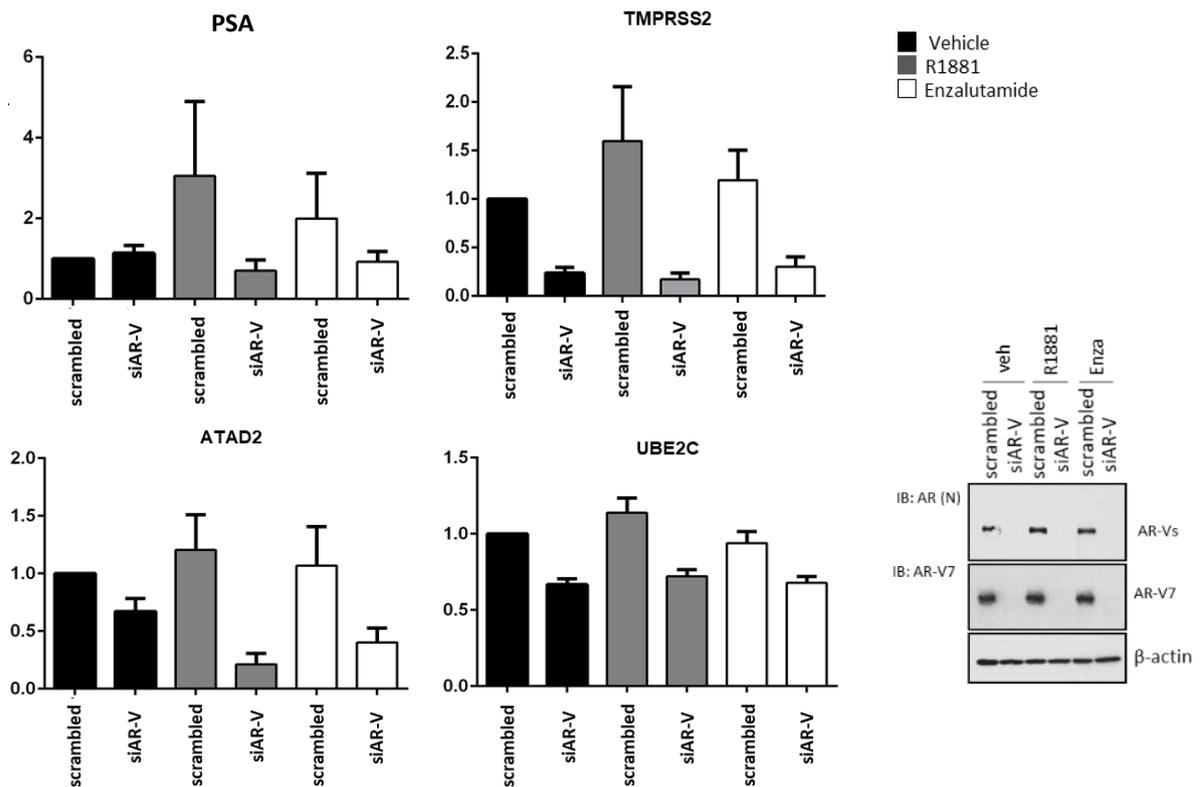


**Figure 43. AR-V levels were maintained in CRISPR edited AR-EK cells.** CWR22Rv1 parental and CWR22Rv1-AR-EK cells were reverse transfected with non-silencing (scrambled) and AR-V-targeting siRNAs for 48h prior to cell harvest and expression analysis of AR-Vs by qPCR (top panel) using specific reverse primers (depicted by arrows) to the unique cryptic exons present in the 3'-end of each AR-V (bottom left panel). CWR22Rv1 parental and CWR22Rv1-AR-EK cells were cultured in full media (containing 10% FBS)/serum-containing media for 72h prior to cell harvest, RNA isolation, reverse transcription and mRNA expression analysis of AR-Vs. Data represent the mean of three independent experiments  $\pm$  SEM. Expression levels of AR-Vs at protein level were compared between parental and AR-EK cells by WB using an N-terminal-binding AR antibody and  $\alpha$ -tubulin as loading control.

### 5.3.6. AR-Vs maintain expression of AR target genes in the absence of AR-FL

The capacity of AR-Vs to function as transcriptional regulators independently of AR-FL is currently debated with evidence suggesting that AR-Vs remain sensitive to next-generation anti-androgens while other studies, particularly that of the R1-D567 cell line, indicates that AR-Vs autonomously mediate androgenic signalling without the need of the full-length receptor. To investigate this further, CWR22Rv1-AR-EK cells grown in the presence and

absence of the synthetic androgen R1881, and enzalutamide were transiently transfected with either scrambled or AR-V-targeting (siAR-V) siRNAs and AR-target gene expression was assessed. Quantitative PCR analysis demonstrated that *PSA*, *TMPRSS2*, *UBE2C* and *ATAD2* remained unaltered in the presence of R1881 and enzalutamide, which is consistent with loss of AR-FL and activity of AR-Vs, but were all diminished upon depletion of AR-Vs, indicating that AR-Vs maintain transactivation of canonical AR target genes in this cell line ().

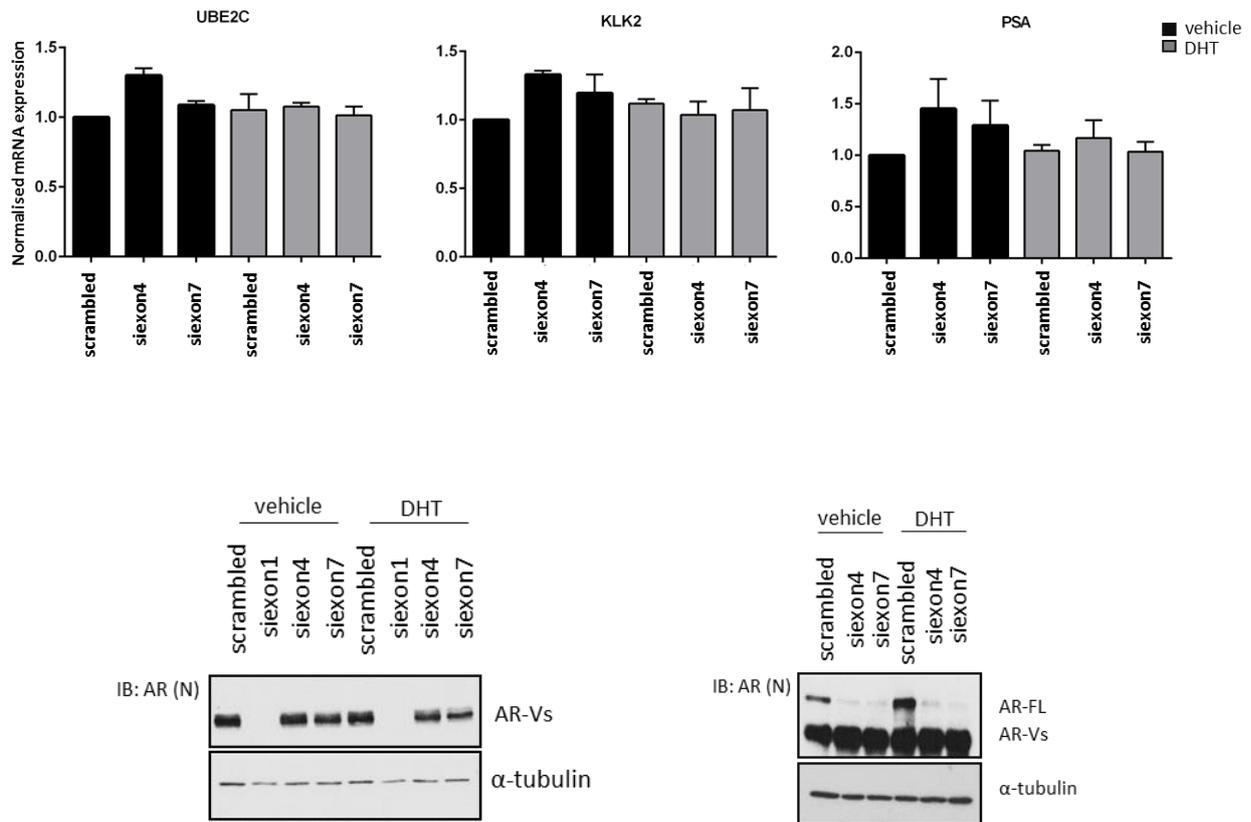


**Figure 44. AR-Vs maintain transactivation of canonical AR target genes in CWR22Rv1-AR-EK cells.**

CWR22Rv1-AR-EK cells grown in steroid-depleted media were reverse transfected with scrambled and AR-V (siAR-V)-targeting siRNAs for 72h. Cells were treated with either vehicle, 1 nM R1881 or 10  $\mu$ M enzalutamide for the final 24h. PSA, TMPRSS2, ATAD2 and UBE2C mRNA expression levels were assessed by qPCR. Data represents the mean of three independent experiments  $\pm$  SEM. Whole cell lysates from a parallel experiment were analysed by SDS-PAGE to confirm AR-V knockdown at protein level using an N-terminal binding AR antibody.

Moreover, AR target gene expression in CWR22Rv1-AR-EK cells was not impacted by siRNAs targeting the C-terminus of the receptor (exon 4: siARex4 and exon7: siARex7) (Figure 45)

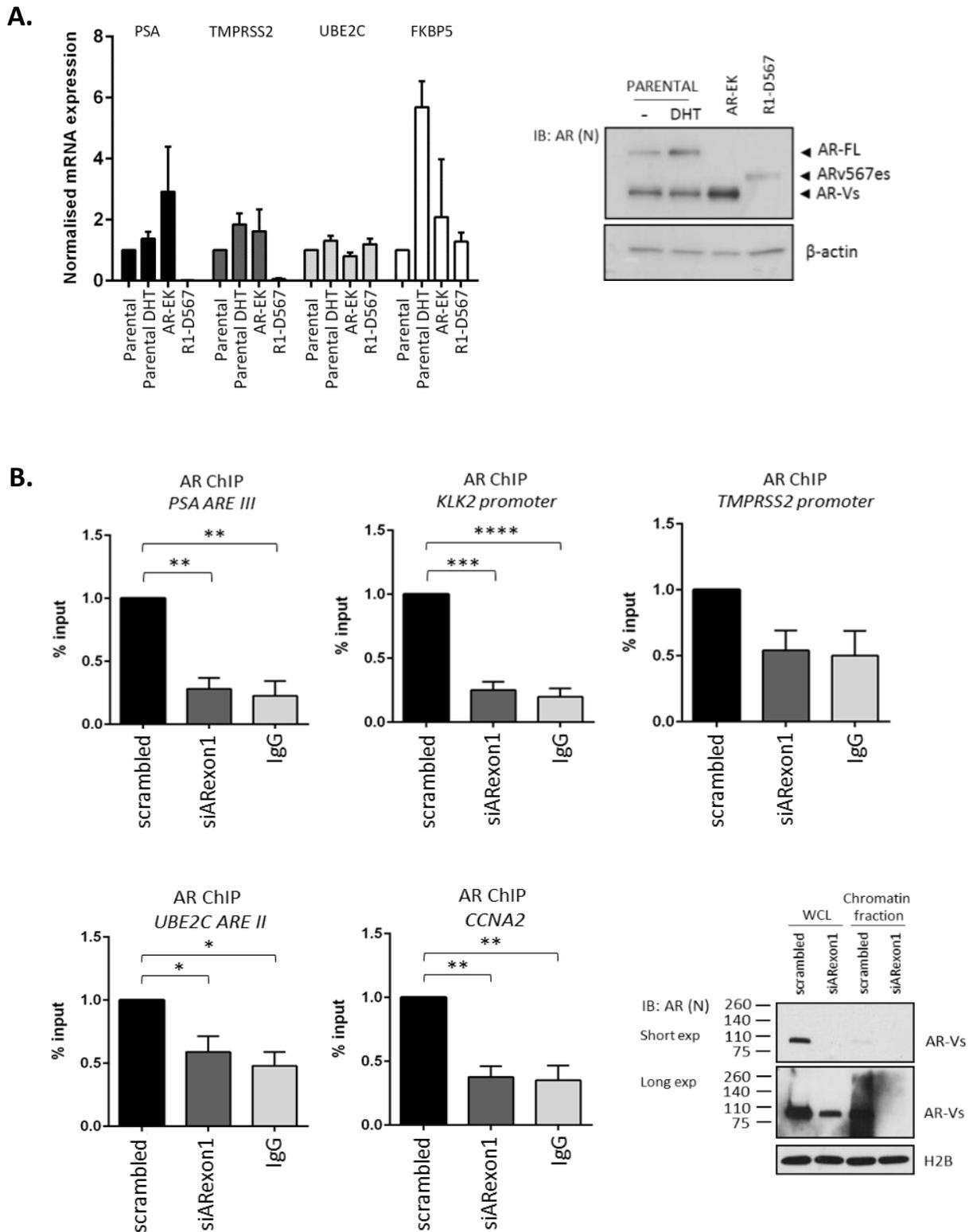
which is again in line with loss of AR-FL and further supports the concept that androgenic gene expression is driven by AR-Vs in this cell line derivative.



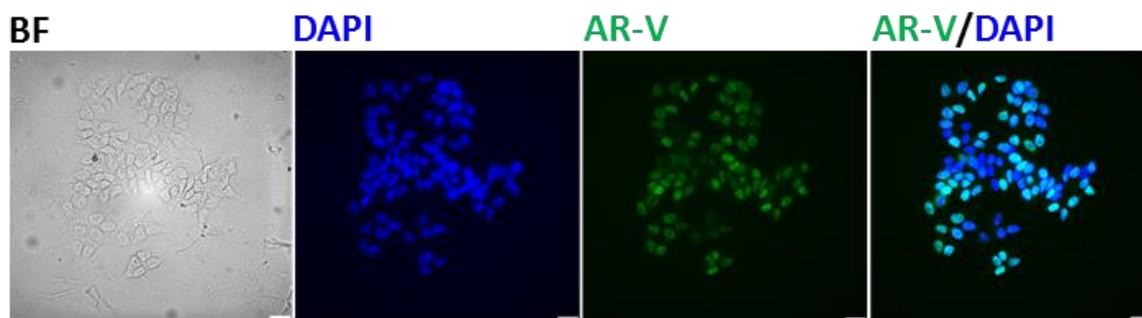
**Figure 45. Androgenic gene expression was maintained by AR-Vs in CWR22Rv1-AR-EK cells.** CWR22Rv1-AR-EK cells grown in steroid-depleted media were subject to scrambled, exon 4 (siexon4) or exon 7 (siexon7) siRNA knockdown for 48h. Cells were treated with either vehicle or 10 nM DHT for the final 24h. UBE2C, KLK2 and PSA mRNA expression levels were assessed by qPCR. Data represents the mean of three independent experiments  $\pm$  SEM. Accompanying immunoblots (bottom panel) were incorporated to verify efficient siRNA-mediated AR depletion (siARex1 in AR-EK cells) and specific targeting of AR-FL (siARex4 and siARex7 in parental cells).

Finally, expression of AR target genes in the CWR22Rv1-AR-EK derivative was largely consistent with parental CWR22Rv1 cells grown in both the presence and absence of DHT, and for *UBE2C* and *FKBP5*, was also comparable to the R1-D567 TALEN-engineered cell line (Figure 46A) confirming the ability of AR-Vs to function as transcriptional regulators without requiring AR-FL. This also suggests the occupancy of previously AR-FL associated genes by AR-Vs, a finding which was backed up by ChIP data which indicated that AR-Vs are recruited to the *cis*-regulatory elements of classical AR genes such as *PSA*, *KLK2* and *TMPRSS2* when AR-FL

is absent, while they retain their original chromatin sites on the *CCNA* and *UBE2C* cis-regulatory regions (Figure 46B). Finally, expected localisation of AR-Vs in the nucleus was confirmed by immunofluorescence (Figure 46C).



C.



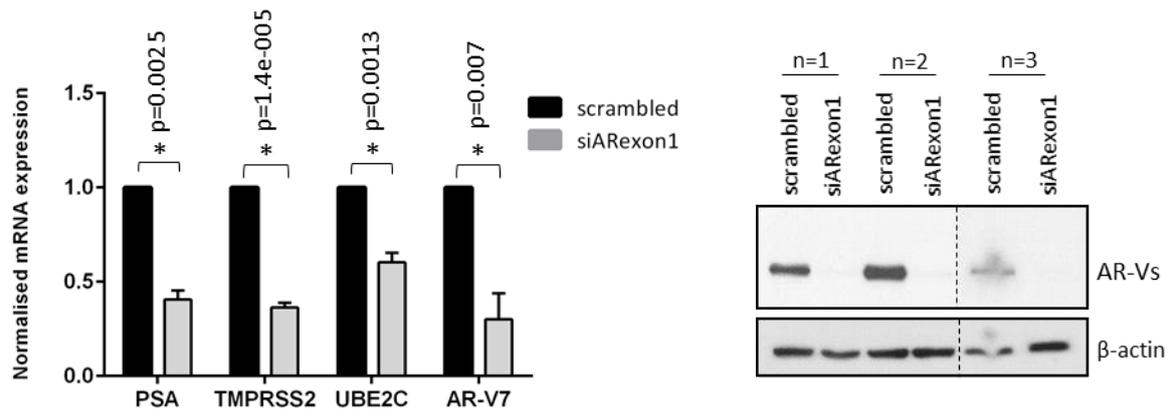
**Figure 46. Occupancy of *cis*-regulatory elements of canonical AR target genes by AR-Vs.** **A.** Comparison of PSA, TMPRSS2, UBE2C and FKBP5 mRNA levels between CWR22Rv1, CWR22Rv1-AR-EK and R1-D567 cells grown in steroid-depleted media supplemented with or without 10 nM DHT, as indicated. Data represents the mean of three independent experiments  $\pm$  SEM. AR protein levels were assessed by immunoblotting using N-terminal binding AR, AR-V7 and  $\beta$ -actin antibodies. **B.** CWR22Rv1-AR-EK cells were subject to control (scrambled) or AR (siARexon1) knockdown for 48 hours prior to ChIP incorporating either N-terminal AR-binding or control (IgG) antibodies. Data represents the mean of three independent experiments  $\pm$  SEM (\*, \*\*, \*\*\*, \*\*\*\* represent  $p < 0.05$ , 0.01, 0.001 and 0.0001, respectively as determined using one-way ANOVA). Accompanying immunoblot of CWR22Rv1-AR-EK whole cell lysates (WCL) and chromatin fractions, incorporating AR and histone H2B antibodies, demonstrates successful depletion of AR-Vs in siARexon1-transfected cells. **C.** Representative bright field (BF) and immunofluorescence images of CWR22Rv1-AR-EK cells at 40x magnification. Scale bars are 25  $\mu$ m.

### 5.3.7. Transcriptomic analysis of CWR22Rv1-AR-EK cells

This is the first study to provide unbiased global AR-V transcriptomic read-outs in a cellular background free of AR-FL upon CRISPR-mediated gene knock-out, not relying on just compromised expression of AR-FL upon partial siRNA-mediated depletion. Hence, the possibility of AR-FL protein traces interfering with the read-outs has been ruled out. Transcriptomic analysis in this unique background was performed in an attempt to highlight networks that AR-Vs regulate autonomously either directly or indirectly. More precisely, RNA sequencing of CWR22Rv1-AR-EK cells was carried out to elucidate the molecular pathways controlled exclusively by AR-Vs to ultimately indicate potential AR-V specific targets which would be validated in *in vitro* and *in vivo* models, revealing yet unknown vulnerabilities of prostate cancer potentially exploited in the clinic.

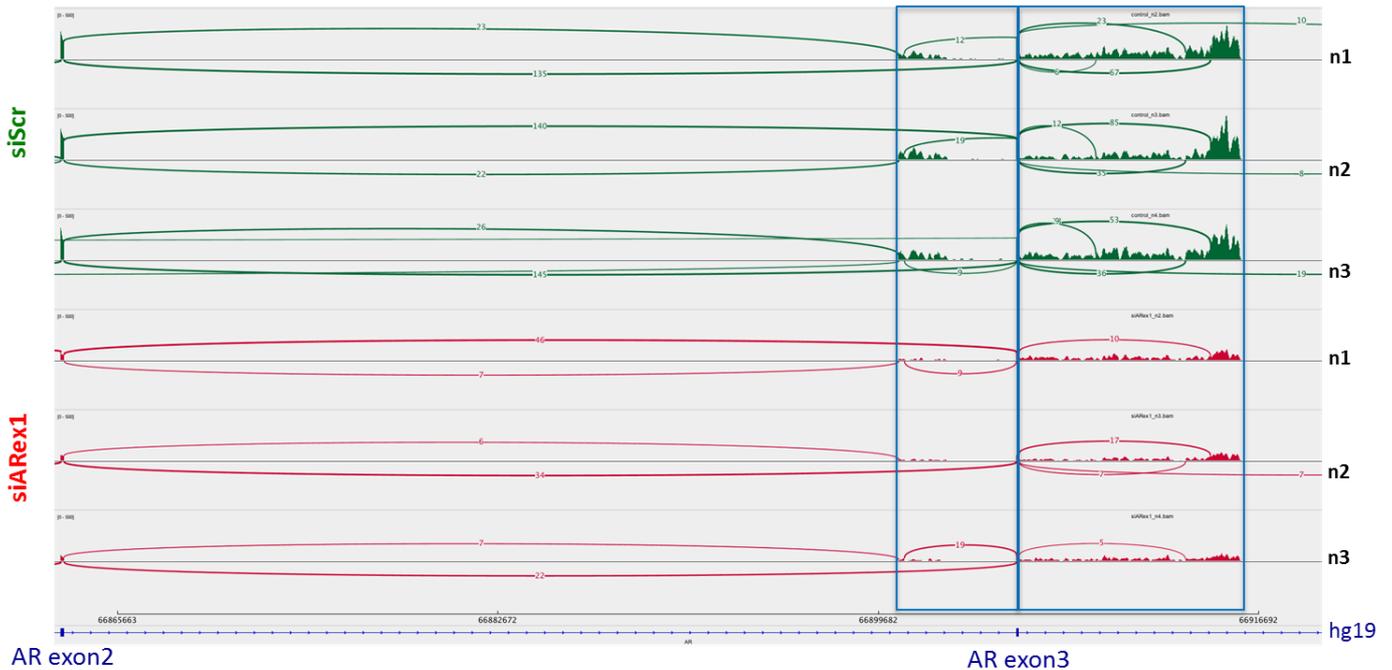
Therefore CWR22Rv1-AR-EK cells were grown in steroid-depleted conditions and were transfected with either an AR exon 1-targeting or scrambled control siRNAs to define the AR-V transcriptome. Extracted RNA from the siARexon1 and non-silencing arms of the

experiment underwent QC evaluation. Knockdown efficiency of AR-Vs was evaluated by qPCR and WB (Figure 47). A number of known AR-target genes were interrogated for the expected drop in mRNA levels using the same RNA samples outsourced to Otogenetics (NGS service provider) for library preparation and RNA sequencing. Sufficient mRNA knockdown of over 50% was achieved and AR-Vs were successfully depleted at protein level in all biological replicates allowing downstream high-throughput analysis.



**Figure 47. Preparation of RNA sequencing samples to define the AR-V transcriptome.** Validation of AR-V knockdown efficiency in CWR22Rv1-AR-EK cells seeded in steroid depleted media and reverse transfected with control (scrambled) and AR-targeting (siARexon1) siRNAs for 48h. Canonical AR-target gene and AR-V7 expression levels were assessed by qPCR (left panel). Data represents the mean of three independent experiments  $\pm$  SEM. AR knockdown efficiency was determined in triplicate (n=1, n=2 and n=3) by immunoblotting using an N-terminal targeting AR antibody. Dashed line indicates cropped immunoblot (right panel).

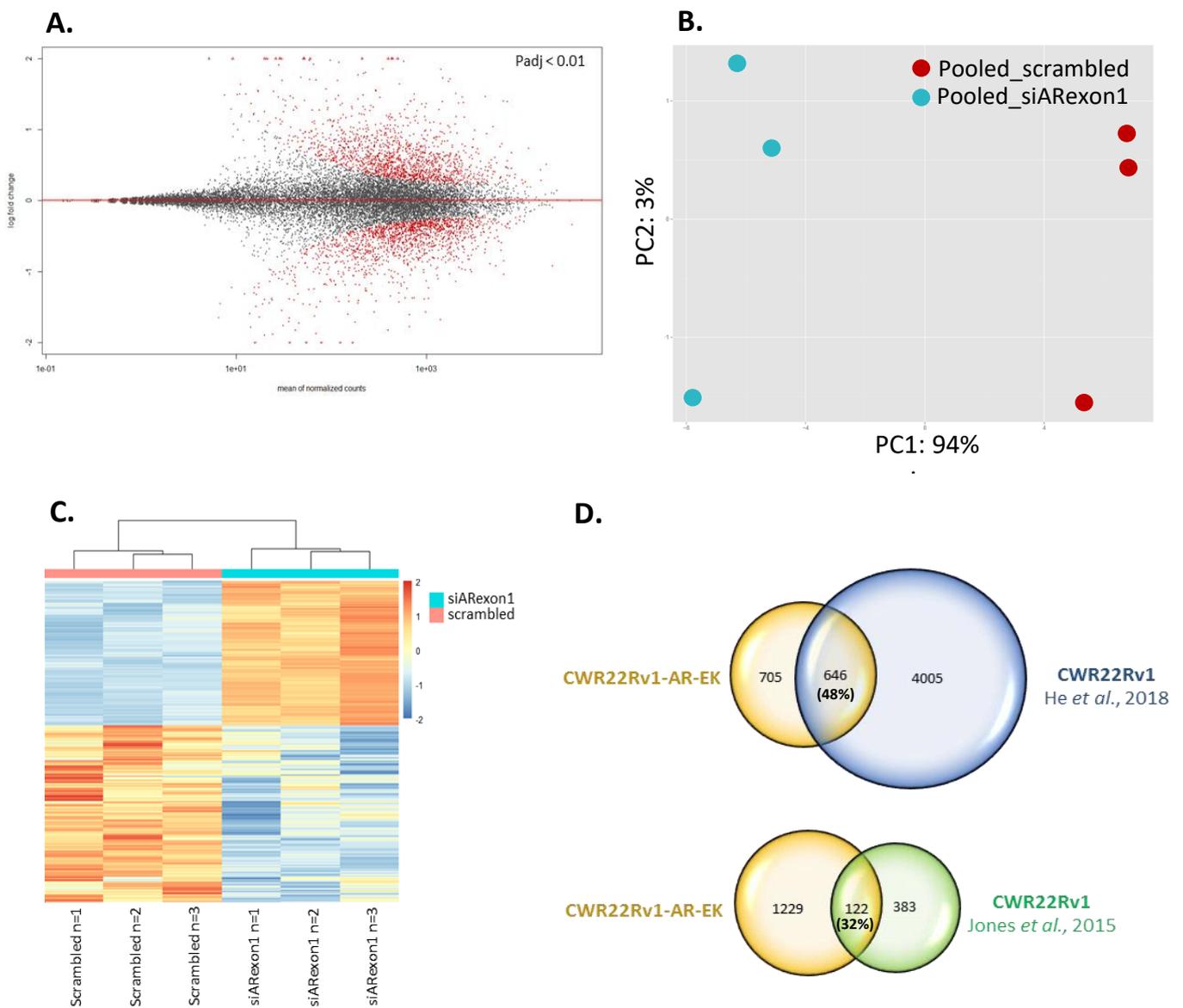
Bioinformatics analysis of the CWR22Rv1-AR-EK RNA-seq data was performed by Otogenetics, Atlanta, USA. AR-V knockdown was confirmed by Integrative Genomics analysis of the alignment data for each sample analysed by RNA-seq. In line with the pre-RNA-seq siRNA efficiency validation (Figure 47), resultant Sashimi plots indicated substantially reduced levels of AR-Vs in all triplicate siAR-V samples (Figure 48).



**Figure 48. Confirming AR-V depletion in siARex1 transfected CWR22Rv1-AR-EK cells.** Sashimi plots of AR gene exon2-3 reads in siARex1 (red) and siScr (green) conditions. Vertical lines indicate read coverage and curved lines indicate splice junctions. Numbers on curve lines denote the number of spliced reads spanning the junctions. The hg19 panel shows the positions of the exons (blue boxes) and introns (blue line) as annotated in the hg19 UCSC human genome. Evidence of reads spanning exon 3 and the downstream cryptic exons located in intron 3 (potentially indicate the presence of AR-V1, V5, V7 and V9 splice variants) were found in siScr but dramatically reduced in siARex1 samples. Similarly, reads spanning exon 2 and the downstream cryptic exons located in intron 2 (potentially indicate the presence of AR-V3) were found in siScr, but reduced in siARex1 samples.

Stringent criteria were applied to determine the number of genes which were significantly altered upon AR-V depletion. More precisely, false discovery rate (FDR) threshold was set at 0.01; hence significantly altered genes were identified as those with  $FDR \leq 0.01$ . Applying this criterion to the CWR22Rv1-AR-EK gene list, 3095 genes were identified as significantly altered upon AR-V depletion (MA plot) and were further sorted to up- and down-regulated genes using a fold change cut-off of 1.5 (Figure 49A). Hence, we identified 607 and 744 genes showing increased and decreased expression, respectively (Appendix E). Altered genes clustered closely across the three biological replicates (Figure 49B) demonstrating opposite

trends between the two experimental arms (scrambled vs siARexon1) (Figure 49C). In order to validate these findings, our data set was compared to two previously published data sets; one from our lab and one from Huang’s lab. Both previous studies revealed the transcriptomic spectrum of AR-Vs in CWR22Rv1 cells depleted of AR-FL either by siRNA-mediated knockdown or enzalutamide treatment, respectively. Of note, overlaps of 32% and 48% were observed between our study and the previous in-house and He et al. studies respectively (Figure 49D), validating the robustness of our findings and indicating substantial retention of AR-V activity in the newly developed cell line whilst highlighting its importance as a valuable model for AR-V studies.

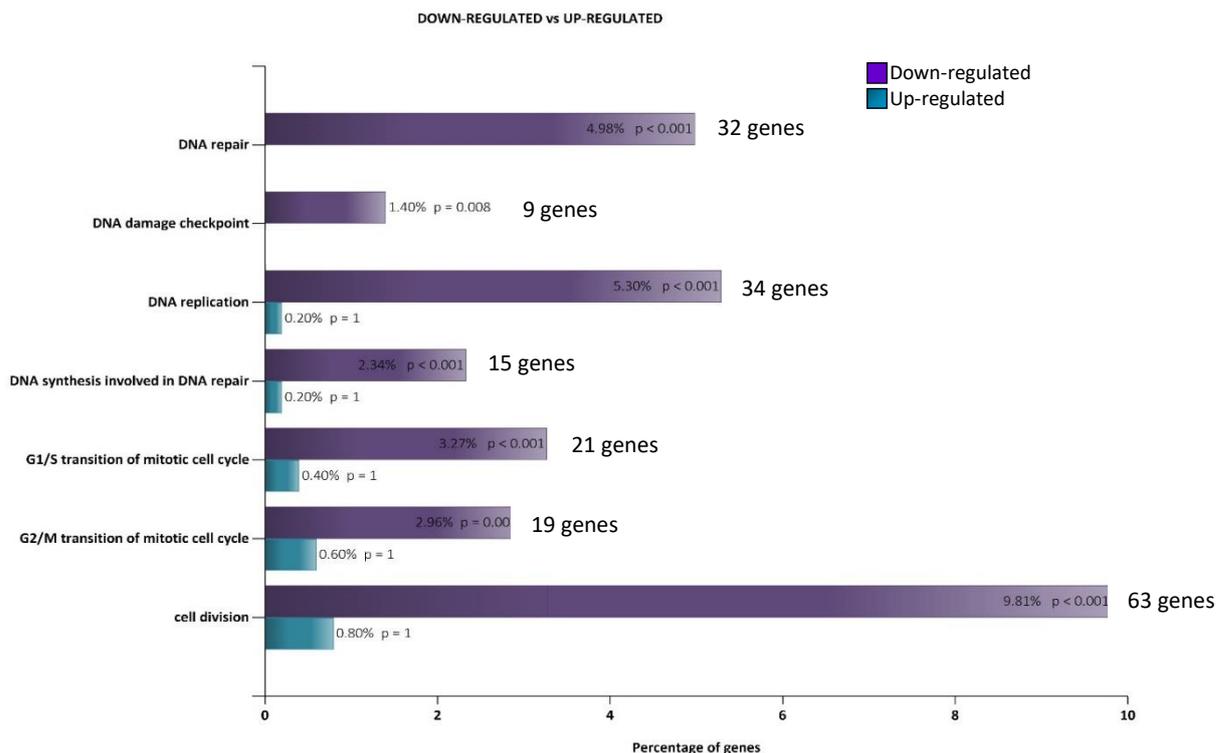


**Figure 49. Deletion of AR-Vs in CWR22Rv1-AR-EK cells impacts global gene expression. A.** MA plot of RNA sequencing data demonstrating statistically significant gene expression changes ( $p < 0.01$ ) above

and below the 0 y-intercept representing up- and down-regulated genes, respectively, highlighted in red. **B.** Principal component analysis (PCA) plot of RNA sequencing data from three biological replicates for each individual sample showing separation of the tested experimental conditions (scrambled vs siARexon1, X axis) and minimal experimental variance across biological replicates (Y axis). **C.** Heatmap of log transformed normalised expression of up- and down-regulated genes in triplicate CWR22Rv1-AR-EK cells subject to either control (scrambled) or AR-V (siARexon1) siRNA depletion. The data is row-scaled with red and blue representing relative higher and lower expression, respectively. **D.** Comparison of the CWR22Rv1-AR-EK RNA-seq dataset to two previously published datasets derived from RNA-seq (top) and microarray (bottom) analysis of CWR22Rv1 parental cells depleted of AR-Vs. Venn diagrams show AR-V regulated gene overlaps of 48% and 32%, respectively.

### 5.3.8. AR-Vs are cell cycle and DNA damage response (DDR) regulators

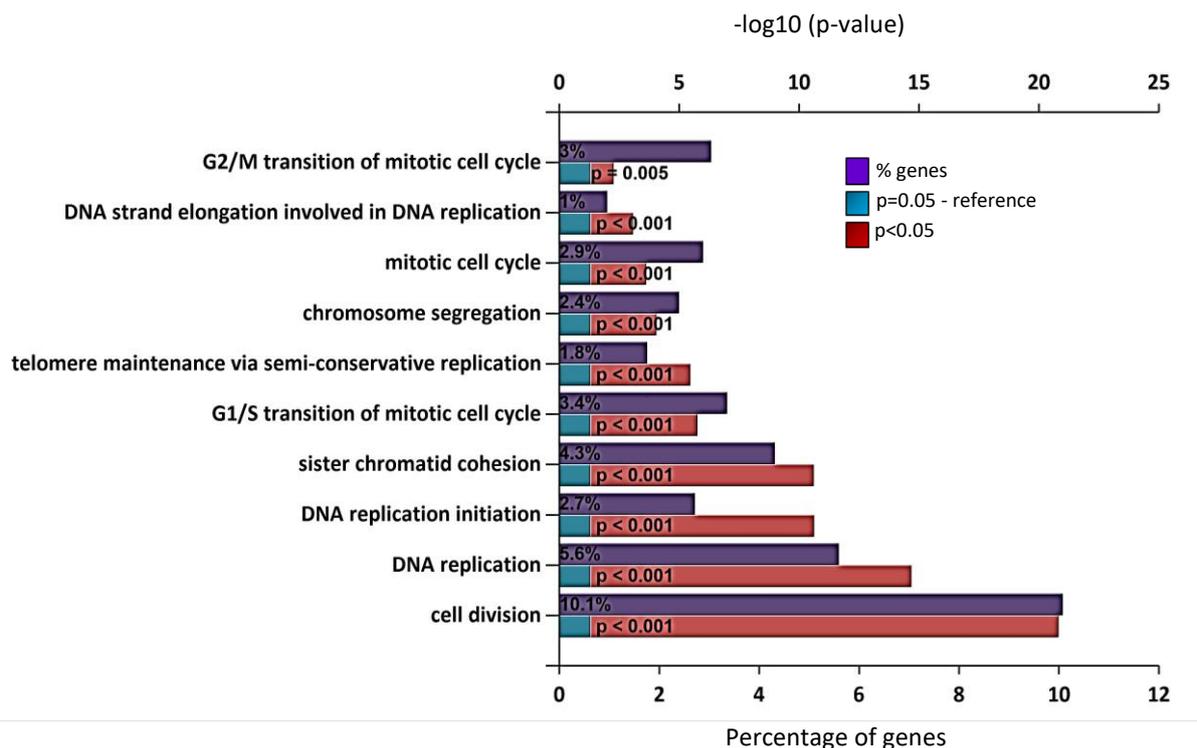
Functional analysis of the significantly altered genes revealed two signalling pathways/biological processes which were significantly impacted upon AR-V loss in CWR22Rv1-AR-EK cells and are hence activated by the truncated receptors in CRPC: 1) the cell cycle and mitosis pathway; 2) the DNA damage response pathway (Figure 50).



**Figure 50. Bar chart of the most significantly altered signalling pathways upon AR-V depletion in CWR22Rv1-AR-EK cells.** Gene ontology analysis was performed using FunRich. Down-regulated pathways are shown in purple, up-regulated pathways are shown in turquoise. The % of genes identified in each pathway are shown alongside absolute gene numbers and statistical significance of the enriched genes featuring in each pathway (Appendices G & H).

### 5.3.9. AR-Vs regulate cell cycle, cell proliferation and cell fat

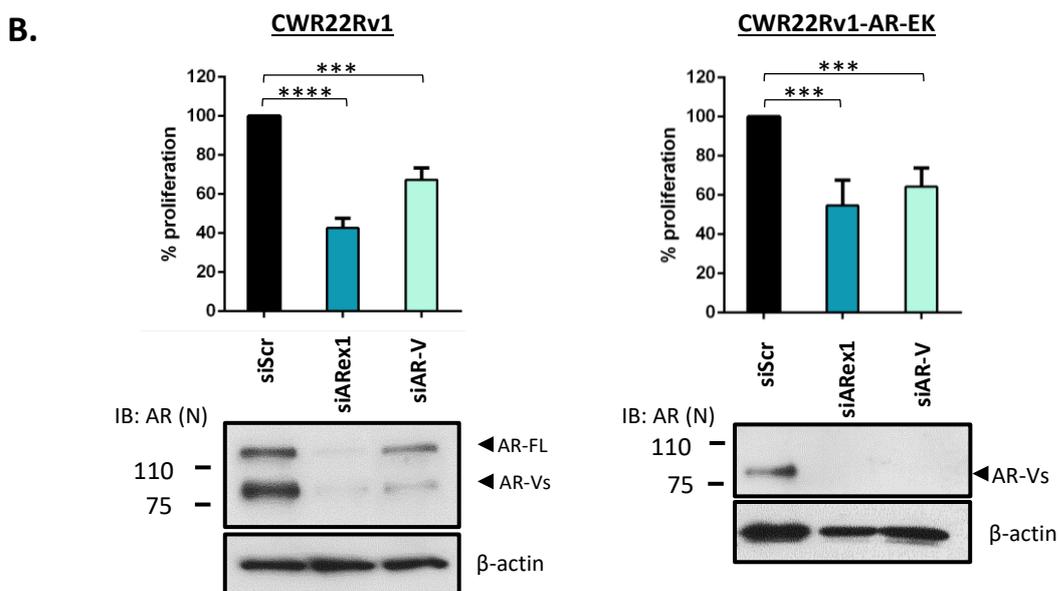
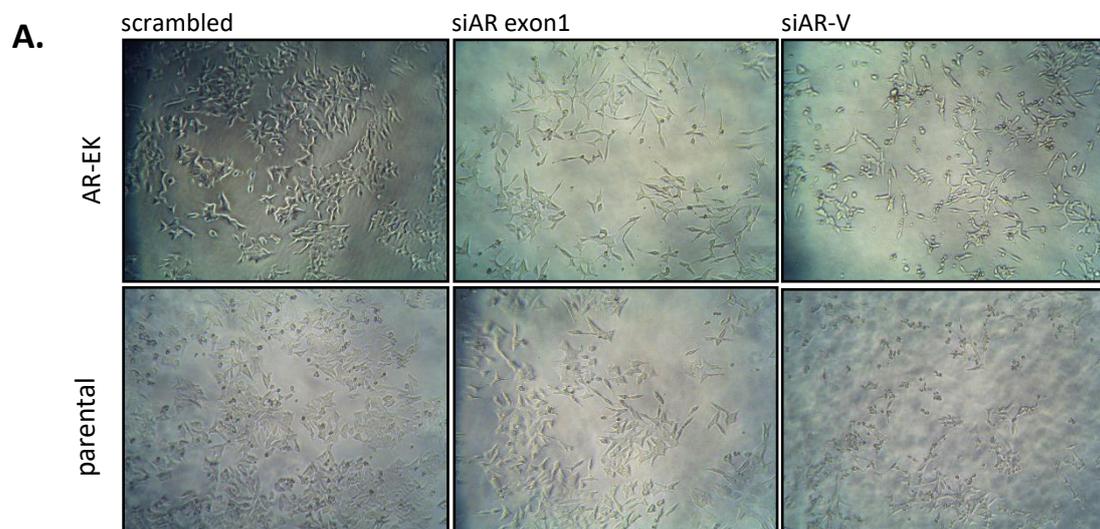
The role of AR-Vs in regulating cell cycle and mitosis has been highlighted in previous prostate cancer studies with He *et al.* being the most recent one. Our findings are consistent with the current literature and strongly support a pro-proliferative role of AR-Vs. More specifically, functional analysis of the significantly altered genes derived from RNA-seq analysis indicated gene clusters significantly associated with namely DNA replication, cell division and cell cycle phase transition (Figure 51 and Appendix G).

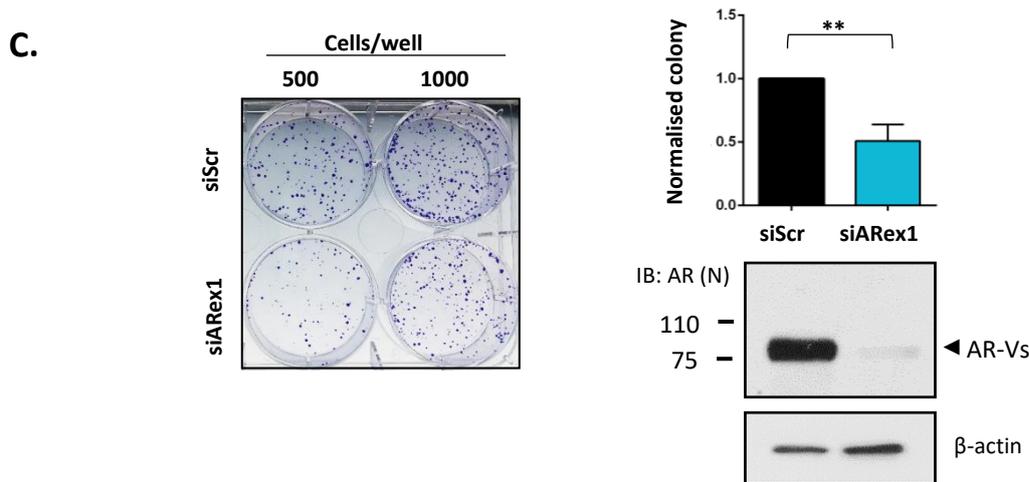


**Figure 51. Bar chart of the most significantly downregulated cell cycle related processes upon AR-V knockdown in CWR22Rv1 AR-EK cells.** Gene ontology analysis was performed using FunRich. The % of genes identified in each pathway are shown alongside statistical significance of the enriched genes featuring in each pathway.

In turn, cell proliferation assays were performed not only to validate the RNA-seq data, but also to evaluate whether CWR22Rv1-AR-EK cells still rely on AR-V activity for growth and expansion upon permanent loss of AR-FL. Therefore, CWR22Rv1 and CWR22Rv1-AR-EK cells were transiently transfected with two siRNA oligos; an ARexon1-targeting siRNA to deplete AR-FL and all AR-Vs and a 3'UTR-targeting siRNA to deplete all variants which share a common polyadenylation signal within the 3'UTR, downstream of cryptic exons. Cell proliferation capacity was assessed by SRB assays whereby cells were allowed to grow for 5 days (which is equivalent to 2 passages) and were siRNA transfected twice during that time-course.

Conventional microscopic image analysis indicated that growth was significantly impacted with cell shrinkage and death being apparent in the AR knockdown arms (Figure 52A). Quantitative analysis using SRB measurement as the end-point read-out demonstrated a significant drop of approximately 50% in CWR22Rv1-AR-EK cell proliferation when cells were depleted of AR-Vs with each siRNA (Figure 52B) indicating that AR-FL knock-out CWR22RV1-AR-EK cells are still highly dependent upon AR-Vs for their survival and proliferation, a finding which was backed up by colony formation assays. In line with proliferation stall, cell propagation was severely impacted by loss of AR-Vs to the extent that cell colony formation was reduced by 50% (Figure 52C), highlighting a critical role of AR-Vs in cell fate.

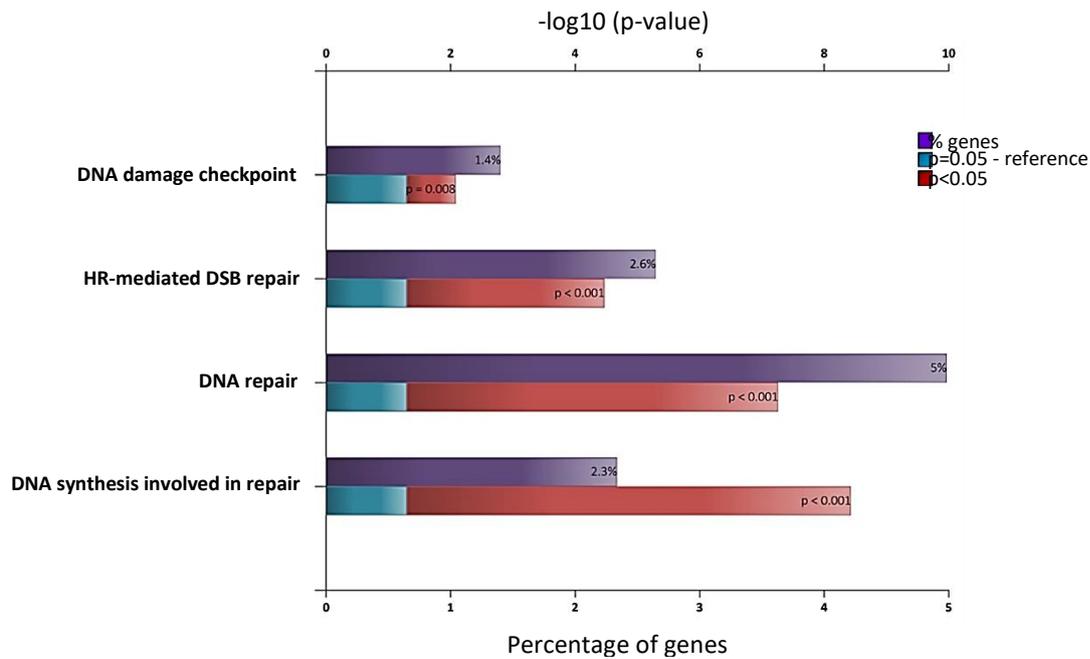




**Figure 52. AR-Vs dictate the proliferation potential of CWR22Rv1 parental and AR-EK cells. A.** Bright field images of CWR22Rv1 parental and AR-EK cells transfected with control (siScr), AR exon1- and AR-V-targeting siRNAs for 72h. **B.** Proliferation of CWR22Rv1 parental and AR-EK cells transfected with control (siScr), AR exon1- and AR-V-targeting siRNAs measured using SRB assays at t=120h post-transfection. Accompanying AR immunoblots are shown to indicate successful AR depletion in both cell lines. **C.** CWR22Rv1 AR-EK cells assayed for their colony forming capacity following transfection with control (siScr) and AR exon1 (siARex1)-targeting siRNAs for 48h. Two different cell densities were examined and colonies arising two weeks post-transfection were fixed, stained (left panel) and quantified (right panel). Accompanying AR immunoblot is shown to indicate successful AR-V depletion.

### **5.3.10. AR-Vs are master regulators of DDR**

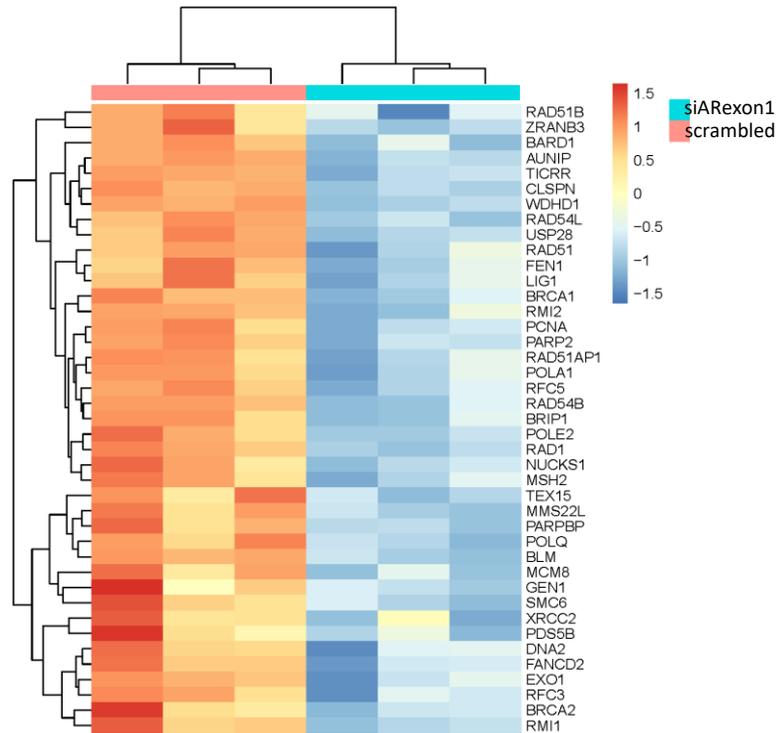
An extremely interesting observation from CWR22Rv1-AR-EK transcriptomic profiling, and subsequent functional analysis of the most down-regulated genes following AR-V depletion, was the apparent enrichment of DNA repair pathways suggesting involvement of the truncated AR receptors in controlling DDR and DNA repair processes (Figure 53 and Appendix H).



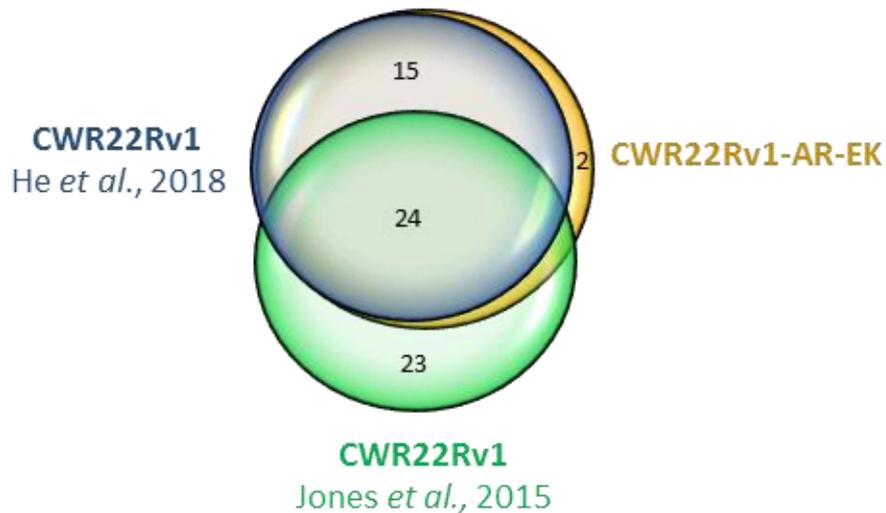
**Figure 53. Bar chart of the most significantly downregulated DDR related processes upon AR-V depletion in CWR22Rv1 AR-EK cells.** The % of genes identified in each pathway are shown alongside statistical significance of the enriched genes featuring in each pathway.

Specifically, a cohort of 41 DDR-associated genes were significantly down-regulated when AR-V levels were compromised indicating that the latter are directly or indirectly involved in DDR (Figure 54A). Our findings were validated by comparing our DDR gene set to an independent and in-house CWR22Rv1-derived AR-V transcriptome (He et al., 2018 GEO:201826 and Jones et al., 2015 GEO:201522) demonstrating respective 95% and 59% overlaps of AR-V-regulated DNA repair genes (Figure 54B).

**A.**

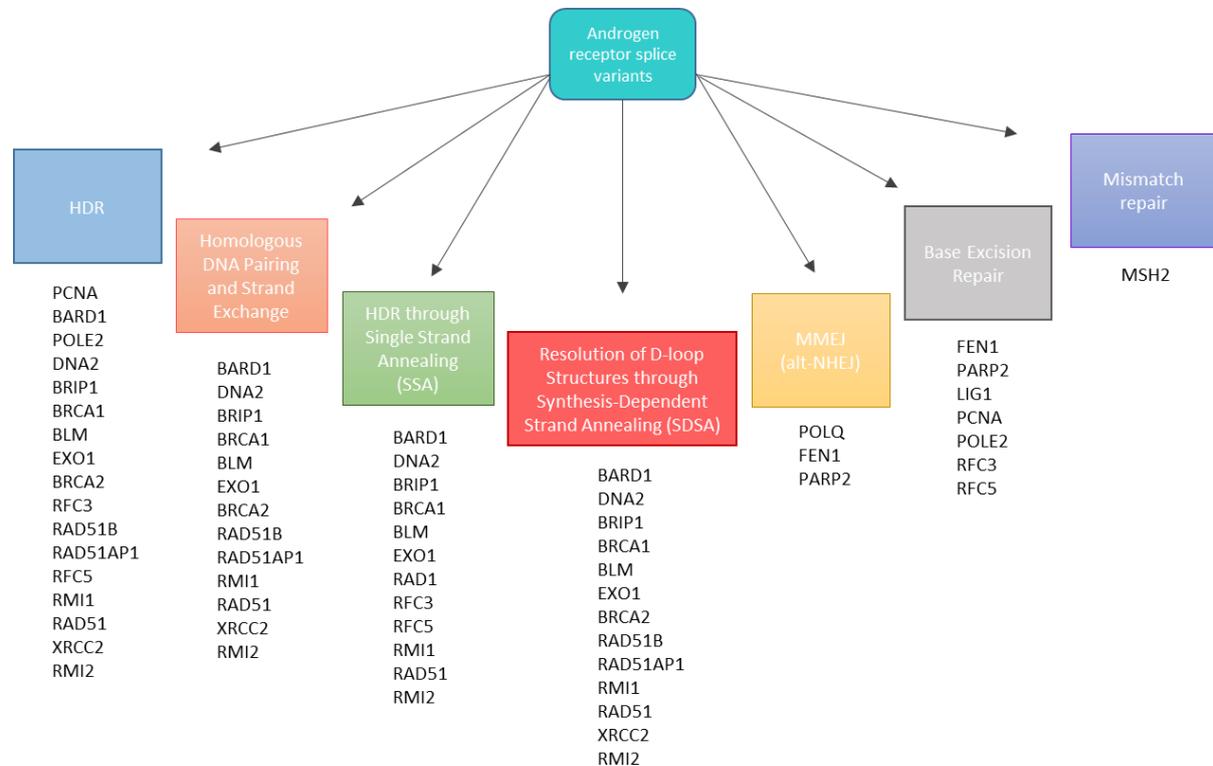


**B.**



**Figure 54. AR-Vs drive expression of a DNA damage response gene signature.** Heatmap showing log transformed normalized expression of the 41 DDR associated genes identified in triplicate CWR22Rv1-AR-EK cells transfected with either control (siScr) or AR (siARex1) siRNAs. The data is row-scaled with red and blue representing relative higher and lower expression, respectively. Venn diagram demonstrating overlap of the 41 DDR-associated genes identified in CWR22Rv1-AR-EK cells and those identified in CWR22Rv1 cells depleted of AR-Vs (He *et al.*, 2018 and Jones *et al.*, 2015).

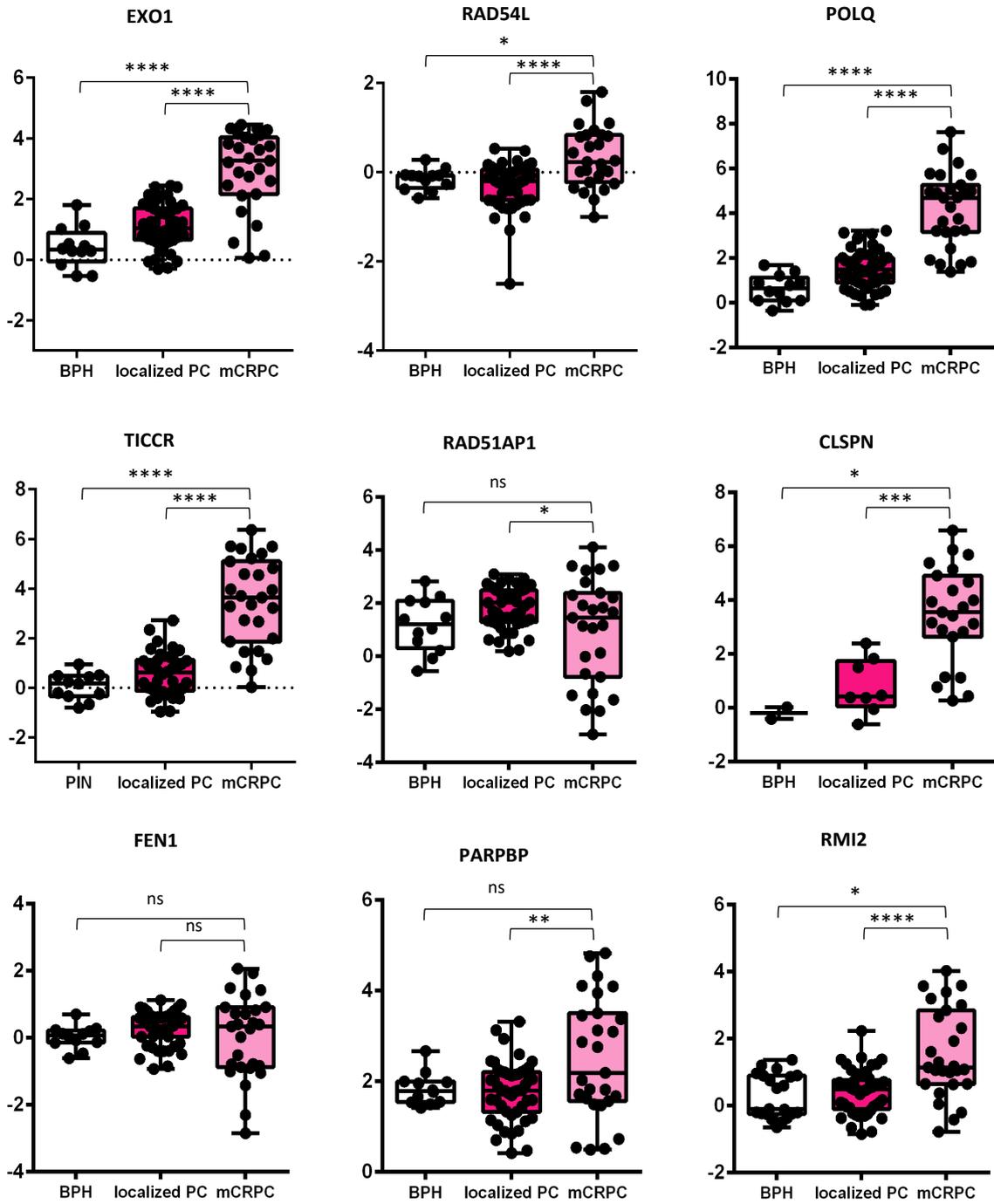
A more detailed analysis of the impacted DDR genes showed the broad range of DDR pathways (detailed in Figure 55) regulated by AR-Vs. There were remarkable gene overlaps across different pathways with a few genes, namely *EXO1* and *BRCA1*, involved in more than one pathways.



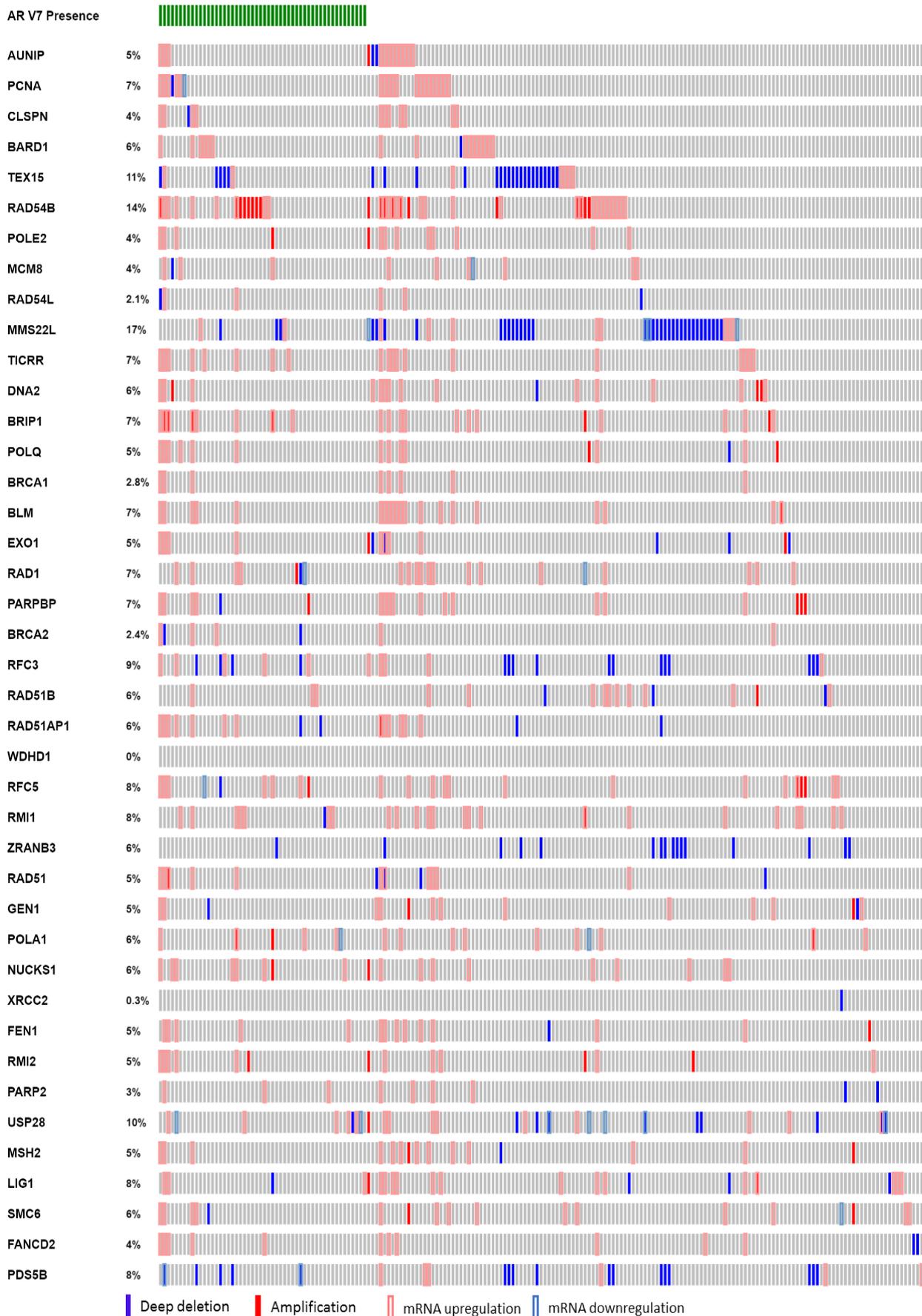
**Figure 55. Spectrum of DDR signalling pathways activated by AR-Vs in CWR22Rv1 AR-EK cells.** Genes featuring in each pathway are shown.

Further *in silico* analysis using the Grasso mRNA data set (GSE35988) indicated a gene set which is significantly upregulated in metastatic CRPC biopsies compared to localised PC and BPH biopsies (Figure 56A). Additional analysis using the TCGA-PRAD data set indicated that all 41 DDR-associated genes identified in our study are upregulated in PC patients and more importantly in a cohort of 84 AR-V7 positive patients (Figure 56B), strengthening the proposed association between AR-V7 status and DDR regulation initially observed in our *in vitro* model. More importantly, a positive correlation between AR-V7 expression and DDR gene regulation was observed for a subset of genes, with significantly elevated mRNA levels in AR-V7 positive vs negative patients (Figure 56C). All the above findings strongly support AR-Vs contribution to elevated DNA repair proficiency in CRPC by essentially acting as potent activators of DDR genes.

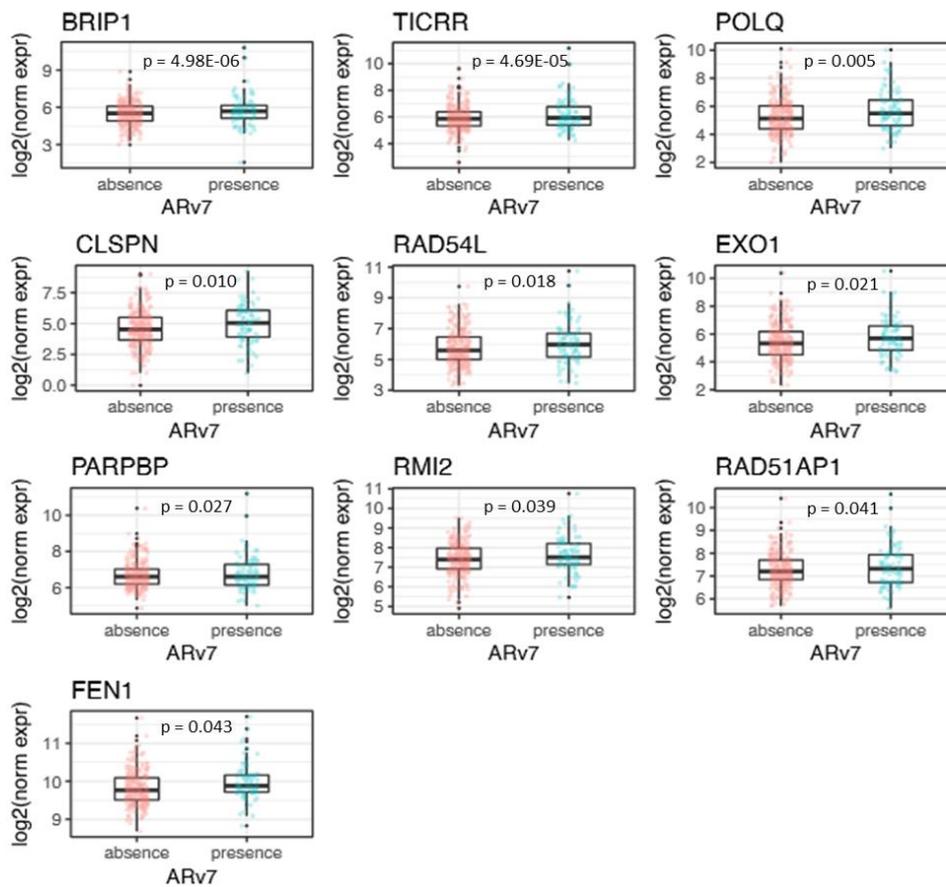
A.



## B.



C.

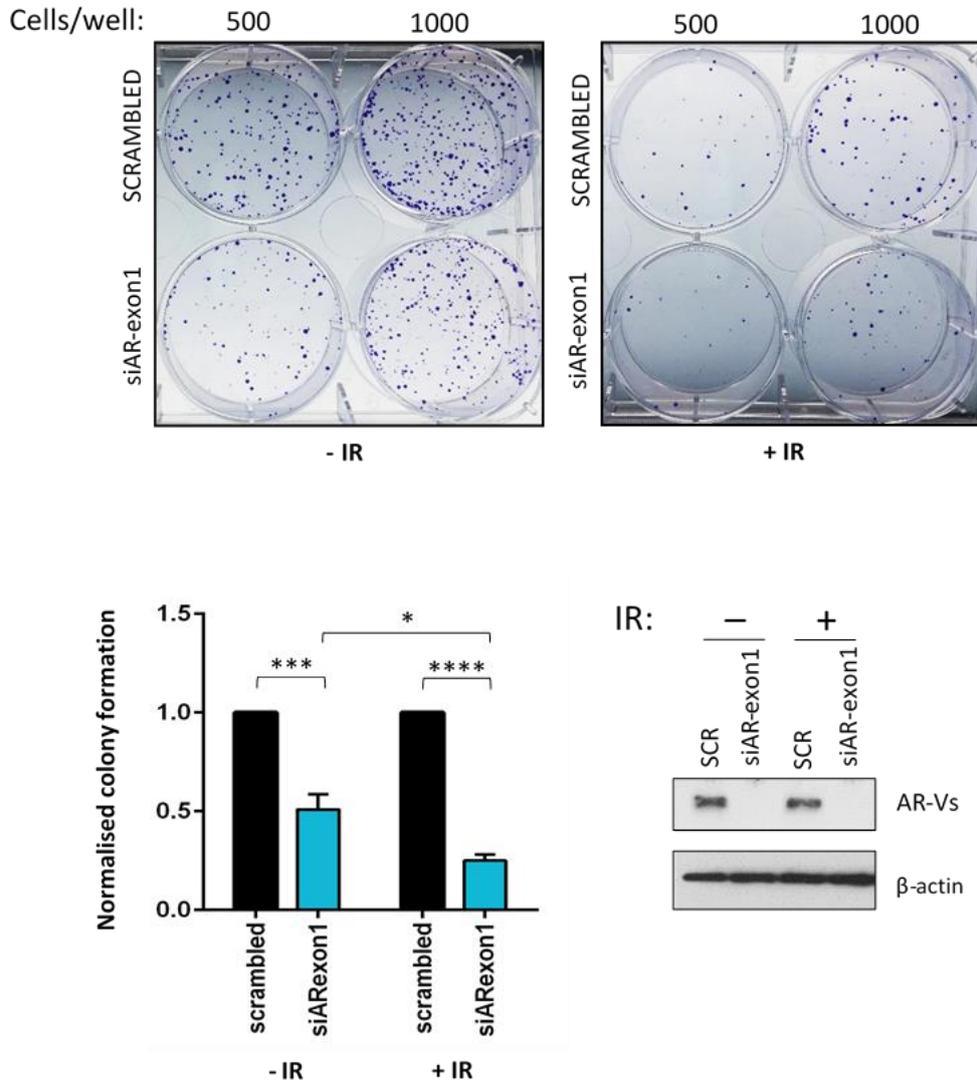


**Figure 56. Expression of AR-V regulated DDR genes is elevated in mCRPC and positively correlates with AR-V7 mRNA expression. A.** Expression of the 41 AR-V activated DDR genes identified in this study was in silico assessed in BPH (n=12), localised PC (n=49) and metastatic CRPC (n=27) biopsies using the Grasso data set. Box plots of the significantly upregulated genes in mCRPC are shown. **B.** The same gene signature was assessed using the TCGA-PRAD data set. AR-V7 status (absence/presence) was applied as an additional clinical parameter/variable. Deep deletion refers to homologous deletion. **C.** Of the 41 DDR-associated genes, 9 demonstrated significantly elevated expression in AR-V7 positive biopsies (n=84) (bottom panel).

### 5.3.11. AR-Vs attenuate sensitisation to ionising radiation

The latter observations suggest that AR-V expression may confer resistance to IR in CRPC via overexpression of multiple DDR associated genes which promote efficient DNA repair to fix the IR-induced damage. To examine this latter hypothesis, CWR22Rv1-AR-EK cells were subject to AR-V depletion followed by exposure to IR and were allowed to expand and form colonies for two weeks. Importantly, AR-V depleted cells treated with 2Gy IR had significantly

reduced survival and colony forming capacity than either depleting AR-Vs or irradiating cells independently (Figure 57).



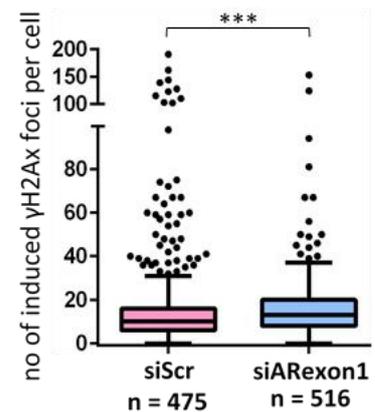
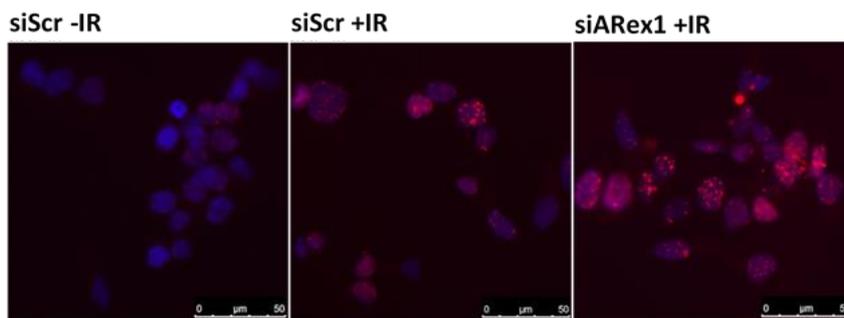
**Figure 57. AR-V depletion sensitises CWR22Rv1-AR-EK cells to IR.** CWR22Rv1-AR-EK cells transfected with either control (siScr) or AR exon1 (siAR-exon1)-targeting siRNAs were subject to 2Gy IR treatment (+IR) and were allowed to grow for two weeks alongside untreated cells (-IR). Individual colonies were fixed, stained with crystal violet (top panel) and quantified (bottom left panel). Cell lysates from a parallel experiment were immunoblotted using an N-terminal-targeting AR antibody to confirm AR-V depletion (bottom right panel).

To measure the extent of DNA damage upon IR in CWR22Rv1-AR-EK cells depleted of AR-Vs,  $\gamma$ H2Ax foci formation at DSB was visualised 24 hours post-IR by immunofluorescence and quantified. Successful recruitment of phosphorylated  $\gamma$ H2Ax was induced in IR treated cells as opposed to untreated cells in which no  $\gamma$ H2Ax foci were observed. Resolution of DSB after

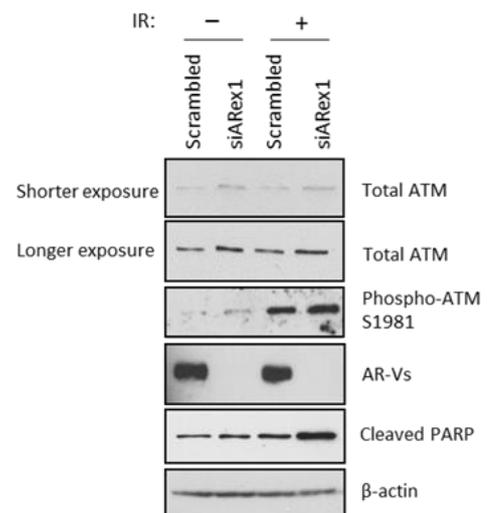
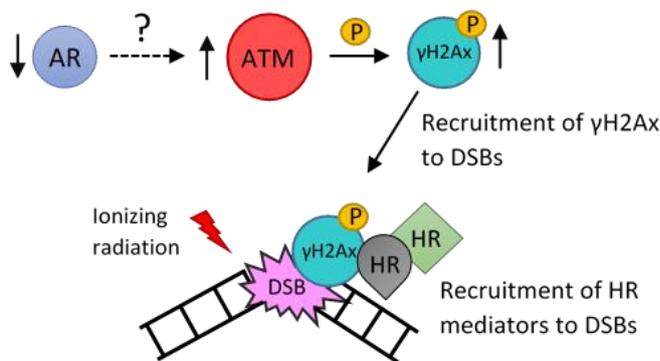
24 hours was severely impacted in AR-V depleted cells as evident by the significantly elevated number of  $\gamma$ H2Ax foci indicative of increased DNA damage or decreased repair capability of these cells compared to those transfected with scrambled siRNA oligo (Figure 58A). Nevertheless, the exact signalling cascade triggered by IR in AR-V depleted cells is unknown. Loss of AR-Vs seems to lead to an increase in ATM protein levels, followed by an increase in phospho-ATM protein levels, stabilisation and phosphorylation of  $\gamma$ H2Ax and recruitment of the latter to DSBs on insulted DNA (Figure 58B). However, further investigation is required to establish the kind of interaction and mediating signals between AR and ATM. Finally, cleaved PARP, a marker of cell death, was increased in AR-V-depleted IR-treated cells compared to scrambled control (Figure 58B), further supporting the notion that loss of AR-Vs coupled with IR might provide a new avenue for managing AR-V positive CRPC patients in the clinic.

### A.

$\gamma$ H2Ax/DAPI



### B.



**Figure 58. Loss of AR-V expression associates with DNA repair deficiency in CWR22Rv1-AR-EK cells.**

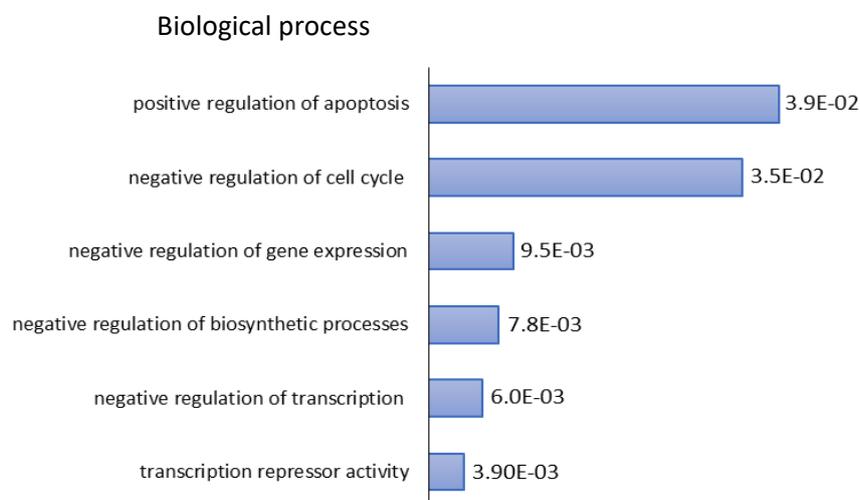
**A.** CWR22Rv1-AR-EK cells subject to either siScr or siARex1 knockdown were treated with and without

2Gy IR and then incubated for 24h before quantifying  $\gamma$ H2Ax foci (red speckles). Representative immunofluorescence images at 20x magnification are shown. Scale bars are 50  $\mu$ m. Box plot in the right panel represents the average number of foci per cell as calculated from two independent experiments  $\pm$  SEM. **B.** Schematic of the signalling cascade triggered by ionising radiation treatment in AR-V depleted CWR22Rv1-AR-EK cells. ATM is rapidly phosphorylated and activated, it then phosphorylates  $\gamma$ H2Ax which is recruited to DSBs alongside other DNA repair mediators to fix the damage. Cell lysates derived from a parallel experiment to (A) were immunoblotted with AR (N), ATM, phospho-ATM (pATM), PARP1/2 and  $\beta$ -actin antibodies.

Taken together, AR-Vs mediate DNA repair and loss of AR-V expression leads to cell sensitisation to IR.

### **5.3.12. AR-Vs may act as transcriptional repressors of tumour suppressor genes**

DAVID functional analysis indicated that genes which are upregulated upon AR-V knockdown and hence normally repressed by AR-Vs in advanced disease clustered in a wide range of functional groups associated with transcription repressor activity, negative regulation of cell cycle and gene expression, induction/positive regulation of apoptosis and response to UV/radiation (Figure 59). It is apparent that all mentioned pathways are linked with a tumour limiting/suppressive phenotype and when they are repressed by AR-Vs tumour progression becomes feasible, indicating that AR-Vs might be involved in repressing tumour suppressor genes. Although it comprises a novel and exciting finding, this avenue has not been deeply explored yet and hence requires more research to understand the contribution of AR-Vs in tumour progression via tumour suppressor gene silencing. Similar findings, supported by cistromic analysis of AR-Vs were recently published and are in line with a tumour limiting role of AR-Vs, particularly AR-V7 in CRPC (Cato et al., 2019).



**Figure 59. Functional analysis of the most up-regulated pathways upon AR-V depletion in CWR22Rv1 cells.**

#### **5.4. Discussion**

The major aims of the study were to: (i) investigate whether AR-Vs are able to exert their oncogenic functions independent of AR-FL, (ii) identify the exact pathways AR-Vs control in an AR-FL free cellular background and (iii) highlight novel therapeutic avenues to eventually potentiate drug discovery and future clinical trials. Therefore, we generated a novel cell line, derivative of the CWR22Rv1 line, CWR22Rv1-AR-EK, which was genetically modified to knock-out AR-FL and study the functional role of the remaining AR-Vs endogenously expressed. This novel cell line model mimics/recapitulates what is seen in the clinic whereby AR-FL is pharmacologically blocked while AR-Vs emerge, enable retention of androgenic signalling and ultimately drive tumour progression.

Firstly, in preparation and subsequent validation of the newly synthesised CWR22Rv1 cell derivative, Cas9 did not affect any predicted off-target loci as no DSBs and subsequent indel formation were observed in those sites, which is consistent with previous reports detailing limited Cas9 target promiscuity (Anderson et al., 2015). However, deep sequencing is recommended and needs to be performed to further support the robustness of our findings.

The CRISPR knock-in strategy presented in this study ended up being more complicated than the initial conception, which relied on a straight-forward integration event of the donor template, containing the desired stop codon, into exon 5 of the *AR* gene via homologous recombination. Our observations, however, indicated that the single-stranded donor template was utilised as a template for *de novo* DNA synthesis with the newly synthesised DNA fragment being knocked-in via HDR. This phenomenon was not limited to our cell line model as it was also observed in *C. elegans* (Paix et al., 2016) and zebrafish (Boel et al., 2018) and more recently in other human cell lines (Paix et al., 2017).

Characterisation of CWR22Rv1-AR-EK cells indicated that AR-V expression remains intact post-CRISPR engineering of the *AR* gene. They are potent activators of classical AR target genes such as *PSA*, *KLK2* and *TMPRSS2*. Conventional ChIP experiments demonstrated that AR-Vs occupy loci otherwise occupied by AR-FL. They hence maintain an AR-FL-like transcriptome, a finding also supported by Chen *et al.* (Chen et al., 2018). This is not a surprise

if one considers that transcriptomic/cistromic profile overlaps between AR-FL and AR-V7 have been shown previously (He et al., 2018). Nevertheless, a more global analysis of the AR-V-specific cistrome in that specific cellular background is required as a considerable degree of cell line specific variation has been observed (Chen et al., 2018).

Importantly, AR-Vs act unhindered by loss of AR-FL to sustain cell proliferation, confirming previous studies which identified AR-Vs as master regulators of cell cycle progression and cell fate (Dehm et al., 2008; Jones et al., 2015). Their pro-proliferative role was in fact strongly supported by the significant drop in cell proliferation and survival rates and the observed inability of cells to recover when all endogenous variants were depleted, indicating that AR-Vs are the main drivers of tumour growth and progression.

Another equally important outcome of our analysis is that AR-Vs comprise potent activators of DDR associated genes. As a rule of thumb, *BRCA*-deficient cancers cannot mediate HR and are therefore vulnerable to IR. Eventually they develop resistance by upregulating and utilising PARP1 as a back-up DNA repair mechanism to fix IR-induced DNA damage. PARP inhibition has now become the gold standard approach to treat these cancers (Ganesan, 2018). In PC, transcriptomic profiling of LNCaP xenografts and prostate cancer cell lines has consistently shown that AR-FL is associated with DDR regulation (Spratt et al., 2015). In fact, AR-FL inhibition reduces DDR gene expression, conferring a BRCAness-like phenotype which eventually leads to apoptosis and tumour recession due to unresolved DNA damage induced by radiation (Asim et al., 2017). Similar to the *BRCA* deficiency scenario, AR-FL-depleted prostate tumours eventually acquire properties which allow them to overcome this obstacle. In particular, they upregulate PARP1 which can mediate repair and rescue IR insulted cells. In this context, ADT combined with PARP inhibition, an approach known as synthetic lethality has demonstrated efficacy in the clinic (Lord and Ashworth, 2017). From our findings, we hypothesised that tumours with AR-V expression will sustain DDR, abrogating response of AR-V positive CRPC patients to IR. This latter assumption was confirmed by our observation that AR-Vs regulate expression of DNA repair genes in an AR-FL-deprived cellular background, such as the CWR22Rv1-AR-EK cell line. In all, AR-Vs define a gene signature which is consistent with a BRCAness-like gene signature identified by previous studies (Li et al., 2017), strongly suggesting that AR-V positive tumours display enhanced DNA repair proficiency which is, in

turn linked to the non-responsiveness to IR observed following AR-FL inhibition by enzalutamide.

However, it remains unknown as to whether AR-Vs regulate this gene cohort directly. Polkinghorn *et al.* (Polkinghorn et al., 2013) performed RNA-seq analysis and identified a DNA repair gene signature (n=74) which is activated by androgens in LNCaP derived xenografts. Further analysis by CHIP-seq revealed the presence of canonical AR-FL binding sites at *cis*-regulatory regions (enhancers) of almost half of the androgen regulated DNA repair genes, defining a subset of DNA repair genes (n=32) which comprise *bona fide* AR targets. Based on previous observations that AR-Vs can occupy classical AR-FL binding sites (including PSA, TMPRSS2 etc.) in the absence of AR-FL (Chen et al., 2018), it is conceivable that AR-Vs can similarly bind to enhancers of DNA repair genes and promote transcription ruling out an indirect mechanism of regulation. However, global cistromic analysis in CWR22Rv1 AR-EK cells is required to confirm this hypothesis.

We have shown that tumour promoting processes such as cell proliferation and DDR are indeed downregulated upon AR-V knockdown and hence normally activated by AR-Vs in CaP. Interestingly we have concluded that AR-Vs may act as suppressors of genes which are linked with tumour suppression and recession. In fact, our analysis indicated that genes which are upregulated upon AR-V knockdown and hence normally repressed by AR-Vs clustered in a wide range of functional groups associated with a tumour limiting/suppressive phenotype and when they are repressed by AR-Vs tumour progression becomes feasible, indicating that AR-Vs might be involved in repressing tumour suppressor genes. Critically, this data fits with a very recent report that AR-Vs, particularly AR-V7 mediates trans-repressive activities in CRPC (Cato et al., 2019) and are capable of down-regulating a distinct tumour suppressor gene-set.

Our results rely on a cell line model derived from a single cell. It would be great to be able to test our hypotheses using more than one AR-FL knock-out (KO) clones in parallel. The study would be remarkably strengthened if our findings from CWR22Rv1 AR-EK cells (clone c34) could be replicated in a different CWR22Rv1 clonal cell population with the same CRISPR induced genetic modifications (AR-FL KO).

To sum up, we have shown that AR-Vs can retain their full functional capacity when AR-FL is lost, and act as dual regulators of gene expression by enforcing tumour promoting mechanisms while suppressing tumour elimination processes, indicating that AR-Vs remain the most clinically relevant disease drivers to be targeted.

## Chapter 6. Repurposing CRISPR to identify AR-V7 specific splicing factors

### 6.1. Introduction

Targeted agents to inactivate AR-Vs is a current clinical unmet need. Lack of the druggable LBD in concert with the disordered N-terminal structure of AR-Vs makes drug design against these receptors extremely challenging. Considerable efforts to develop alternative therapeutic opportunities to tackle persistent AR-V signalling is currently underway. One attractive option involves disruption of spliceosome assembly to prevent formation of AR-V mRNA and subsequent protein expression (Paschalis et al., 2018).

Aberrant alternative splicing has been reported in prostate cancer (Munkley et al., 2017). Despite its contribution to disease progression it simultaneously offers the opportunity of novel therapeutic avenues to be explored and targeting the spliceosome might reveal new vulnerabilities in prostate cancer. A genome wide CRISPR knock-out screen in prostate cancer cell lines revealed a number of splicing factors responsible for the production of alternatively spliced mRNA transcripts which demonstrate tumour-promoting properties (Fei et al., 2017). Over the past few years, more refined and target-specific studies claimed to have identified splicing factors responsible for the generation of AR-V7 namely ASF/SF2 (Liu et al., 2014), Hsp90 (Ferraldeschi et al., 2016) and Sam68 (Stockley et al., 2015).

Despite the efforts to develop specific inhibitors against components of the spliceosome, no promising clinical trials have been reported yet and although early stage clinical trials are currently ongoing, all existing evidence to date is derived exclusively from preclinical models (reviewed by Paschalis et al, 2018). Nevertheless, the field remains elusive, a fact that necessitates more robust research strategies.

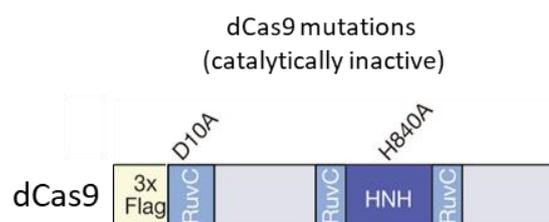
Splicing is highly complex and is characterised by precise and tightly regulated processes principally controlled by *cis*-regulatory sequences present in the intron-exon junctions of pre-mRNA substrates. Canonical sequences have been described and are frequently found in splice sites surrounding introns. They comprise strong splice sites which compete against weak splice sites for selection by the spliceosome (Paschalis et al., 2018). Importantly, the splice sites that surround AR cryptic exon 3 show high resemblance to the consensus 5' GAG/GUAAGA and 3' NYAG/A splice sites, suggesting they may be selected over weaker splice

sites to drive AR-V synthesis; and certain mechanisms or treatments may favour this selection in CRPC.

Identification of the AR-V7 spliceosome will fill a critical knowledge gap in our understanding of how AR-Vs are generated and provide key targets for future therapeutic intervention in advanced PC patients. A high precision antibody affinity-based approach to identify AR-V7 specific splicers was developed: what causes the inclusion of CE3? What are the factors which assemble the splicing machinery and carry out mRNA processing? Although a genome-wide CRISPR screen would possibly provide similar insight, it would additionally highlight factors which contribute to AR-V7 mRNA processing via secondary events and this would limit our ability to identify AR-V7 splicers in direct association with the AR pre-mRNA.

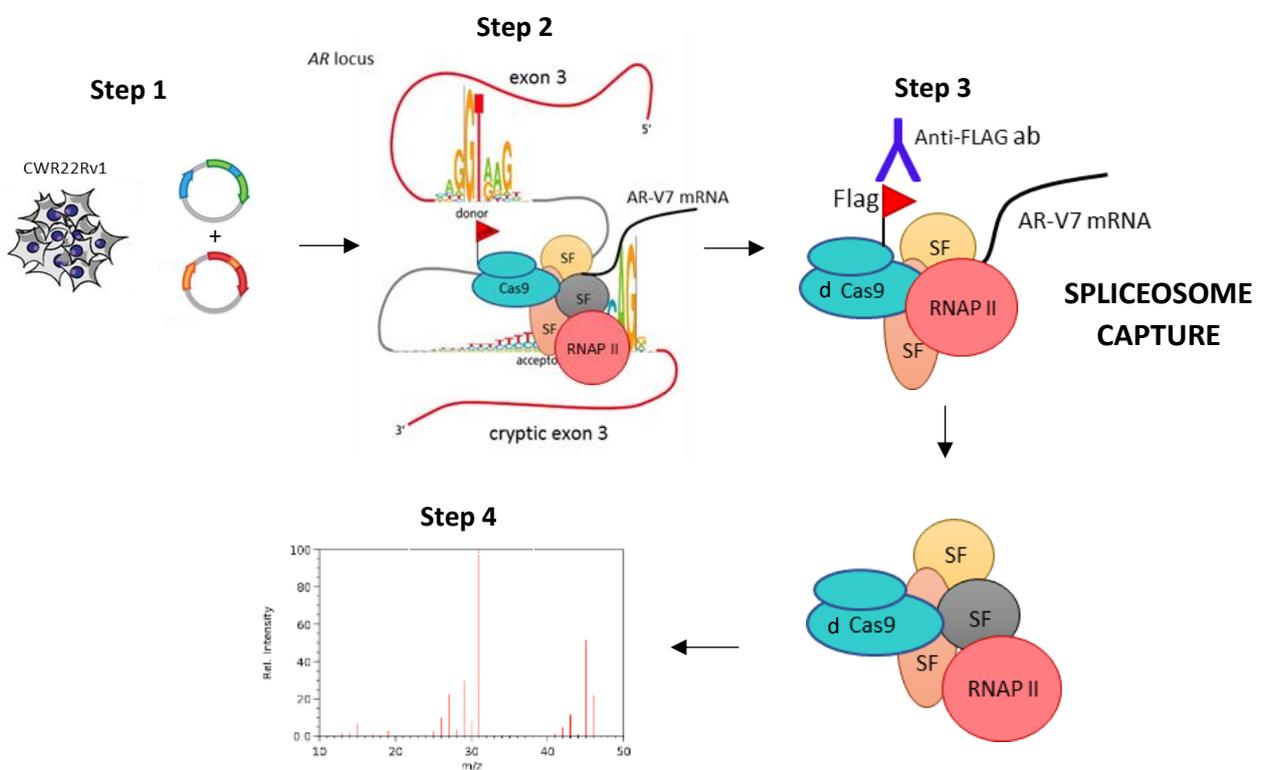
A recently developed method by Carroll's group, named rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) utilises antibody-based affinity purification of endogenous protein complexes cross-linked with chromatin to establish the transient interactome of the protein of interest by mass spectrometry (MS) (Mohammed et al., 2016). RIME has been previously used to identify interacting partners of estrogen receptor in breast cancer cell lines and xenografts (Mohammed et al., 2013). My host lab have previously applied the method to interrogate the interactome of AR-FL and AR-V7 in the CWR22Rv1 prostate cancer cell line (unpublished data).

A deviation of the conventional RIME approach called Cas9-directed Rapid Immunoprecipitation Mass spectrometry of Endogenous proteins (CRIME) was developed in our lab (described in this chapter) to study loci-specific assembly of the AR-V7 specific spliceosome. This CRISPR-based approach allows selective targeting of a catalytically inactive form of Cas9 (dead Cas9, dCas9) (Figure 61) to native chromatin regions and subsequent isolation of protein/RNA complexes associated with these sites.



**Figure 60. Schematic of the nuclease deficient Cas9 (dead Cas9, dCas9).** The point mutations D10A and H840A causing DNA cleavage deficiency are located in the RuvC (*E. coli* protein involved in DNA repair) and HNH (Histidine & Asparagine) domains respectively.

Two gRNAs were designed against the AR cryptic exon 3 genomic region. Targeting of FLAG-tagged dCas9 to that site was exploited for *in situ* capture, isolation and characterization of splicing factors specific to AR-V7 (Figure 61). dCas9 pull-down coupled with mass spectrometry to identify the components of the protein complexes associated with dCas9 at AR-V7-encoding CE3 of the *AR* gene; specifically AR-V7 candidate splicing factors. The advantage of this approach is that low-abundant proteins can be picked-up relatively easy.



**Figure 61. CRIME outline.** CWR22Rv1 cells were transfected with the appropriate amount of dCas9- and gRNA-expressing vectors [step 1]. RNA-guided binding of dCas9 to the target locus and recruitment of SFs to that site [step 2] was followed by formaldehyde-induced cross-linking of the dCas9 associated SFs and splicing regulators to chromatin. Immunoprecipitation of FLAG-dCas9 was then performed and a snapshot of the AR-V7 specific spliceosome was captured [step 3]. Immunoprecipitated dCas9 associated protein complexes were finally analysed by mass spectrometry [step 4].

The DNA-binding dCas9 nuclease was preferred over the RNA-binding dCas13 nuclease given the spatiotemporal regulation of splicing. Remarkably, >80% of spliceosomes were associated

with chromatin in HeLa cells (Bentley, 2014). Hence, mRNA processing and maturation of nascent mRNA transcripts occur co-transcriptionally, as the newly synthesised pre-mRNA emerges from the RNA Pol II exit channel rather than just following transcription. This prevents potential degradation of the pre-mRNA by endogenous nucleases. More importantly, coupling of transcription and splicing in space and time allows the creation of novel protein-protein interaction networks; making spliceosome assembly a much more sophisticated system to study. Trafficking of splicing factors on and off their target pre-mRNA substrate is regulated by RNA Pol II which is also associated with and elongates that same mRNA transcript. More specifically, its carboxy-terminal domain, when phosphorylated, serves as a dock for splicing factors to land and exert their function on the 5' and 3' splice sites of the target mRNA. Cycling of different splicing factors and formation of transient protein-RNA complexes subsequently occurs (Pandya-Jones, 2011). A snapshot of this process can be captured using CRIME.

## 6.2. Specific Materials and Methods

### 6.2.1. Construction of gRNA expressing plentiCRISPR-iCer vectors

Cryptic exon 3 targeting gRNA sequences containing 5' *BsmBI* sites, for directional cloning, were purchased from Sigma (Table 17). *BsmBI* digestion (Invitrogen) of 2 µg plentiCRISPR-iCer vector was performed overnight according to manufacturer's instructions. The digested vector was gel purified and ligated with the different gRNA inserts at a 1:3 ratio using the Anza T4 DNA Ligase Master Mix (Thermo Fischer Scientific). Successfully recombinant vectors were defined by conventional Sanger sequencing.

**Table 17.** Sequences of gRNAs used for CRIME. Part of the *BsmBI* recognition site is highlighted in red.

| Oligo Name     | Sequence (5'→3')          | 5' modification |
|----------------|---------------------------|-----------------|
| gRNA_D forward | CACCGATACTAGAAAAATTCCGGGT | phosphorylated  |
| gRNA_E forward | CACCGGATGCTTGCAATTGCCAACC | phosphorylated  |

### 6.2.2. CRIME, Silver staining and LC-MS

4x10<sup>6</sup> CWR22Rv1 cells were seeded in 150 mm dishes in steroid-depleted media for 24 h. Cells were then transfected with 10 µg pCMV7.1-3xFLAG-dCas9 and plentiCRISPR-iCer vectors and were allowed to grow for an additional 48 h. At 72 h, cells were fixed and harvested as described in Chapter x (ChIP protocol). Pull-down of the dCas9-spliceosome complexes was performed using 10 µg of FLAG M2 and control IgG antibodies (Sigma and Diagenode, respectively). DNase I (Invitrogen) treatment was performed at 37°C for 1 h to remove non-specific proteins bound to the immunoprecipitated chromatin fragments.

Cell lysates were analysed on a pre-cast 4-12% SDS-PAGE gel (BioRad) as described in Chapter 3. The gel was then stained using the ProteoSilver Plus Silver Stain kit (Sigma) as per manufacturer's protocol and was finally stored in 5% acetic acid solution at 4°C until LC-MS analysis. LC-MS and data analysis were performed by members of the Newcastle University Protein and Proteome Analysis (NUPPA) facility. In brief, the gel was cut into thin slices and in-gel digestion was performed. The gel slices were air dried and then incubated in 50 mM

ammonium bicarbonate/trypsin solution at 37°C overnight. Peptides were extracted from the slices upon incubation with increasing concentrations of acetonitrile. They were then concentrated using a vacuum evaporator and finally injected into and analysed using the LTC OrbiTrap XL mass spectrometer (Thermo Fischer Scientific).

### **6.2.3. Flow Cytometry**

4x10<sup>6</sup> CWR22Rv1 cells transfected with 10 µg pCMV7.1-3xFLAG-dCas9 and either 10 µg of lentiCRISPR-iCer (gRNA\_D) or empty vector were cultured for 72h in DCC media. Cells were trypsinised and sorted for Cerulean expression (fluorescent marker expressed by the gRNA expressing lentiCRISPR-iCer vector). Cells transfected with a GFP expressing vector of similar size to the lentiCRISPR-iCer vector were also analysed.

### **6.2.4. siRNA design**

Broad Institute's GPP web portal was utilised to choose the siRNA sequences which were specific to the desired mRNA targets. Alternatively, the Sigma website was exploited and validated siRNA sequences with limited off-target sites and high knockdown scores (>85%) were purchased (Table 18). Cells were transfected as described in Chapter 3.12 and grown in DCC media to enforce AR-V specific splicing events. 48h post-transfection knockdown validation was performed by qPCR using the primers outlined in Table 19.

**Table 18.** Sequences of siRNAs targeting splicing factor hits identified using CRIME.

| <b>Name</b>      | <b>Sequence (5'→3')</b> |
|------------------|-------------------------|
| siSRSF3_sense    | UGGAACUGUCGAAUGGUGAAA   |
| siSRSF9_sense    | GAUGACACCAAUUCGCUCU     |
| siSam68_sense    | ACGAAGGCUACGAAGGCUAUU   |
| siRBM42_sense    | CGUGAGAUGAAUGGGAAGUAU   |
| siKIAA0368_sense | UGCAAUUUGUGCAUCAUUAUUU  |
| siZDHC6_sense    | GGUUUACGAUACUGGUUAUAU   |

|                |                       |
|----------------|-----------------------|
| siLUC72L_sense | GCAGAGGAAGUUUAUCGGAAU |
|----------------|-----------------------|

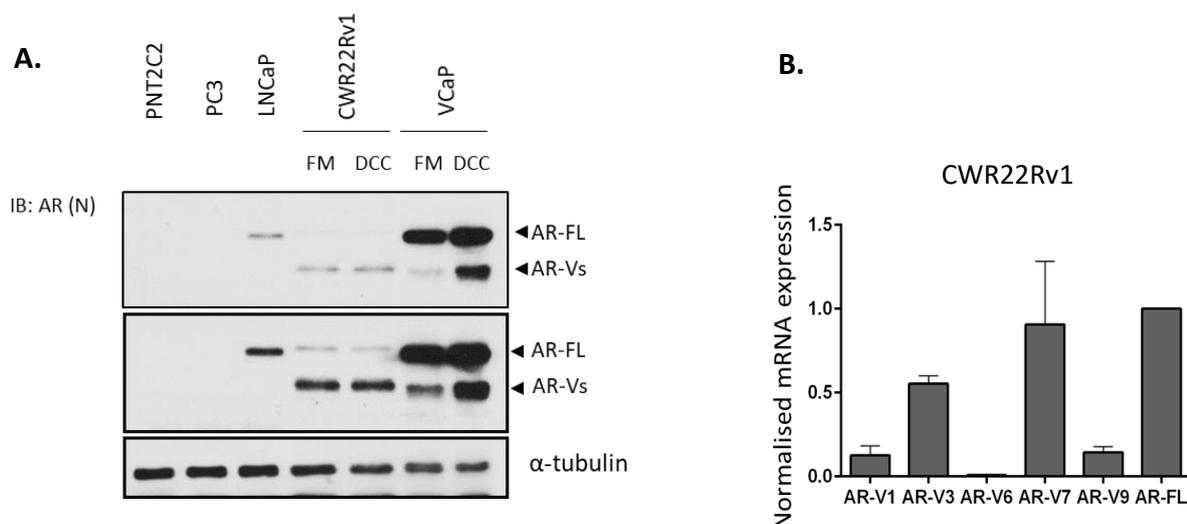
**Table 19.** Sequences of primers used for siRNA knockdown validation.

| Name     | Forward sequence (5'→3')   | Reverse sequence (5'→3')    |
|----------|----------------------------|-----------------------------|
| SRSF3    | CTATGTGGCTGCCGTGTAAG       | TTGGAGATCTGCGACGAG          |
| SRSF9    | ATGGTTATGATTATGGCCAGT<br>G | TGACGGAGGAAGTCCTGA          |
| Sam68    | TGCTGACGGCAGAAATTGAG       | TTGGCTTTGTCTCTCATTGAGC      |
| RBM42    | GAATCCCAACTGCTGTGC         | CTCCAGAGTCTGCTGGAC          |
| KIAA0368 | GACAGATCAGCTTGAACGG        | GGACCAGCAGTTCATTAC          |
| ZDHHC6   | ACTTCAATGCCATGTTTGTCTG     | ATGGTCCATCTTCATCACACA       |
| LUC72L   | ATTCGCCAAGCCGAGATAC        | TGGATGCAATTTTCATAATCCG<br>C |

### 6.3. Results

#### 6.3.1. CWR22Rv1 cell line suitability to apply CRIME

The CWR22Rv1 cell line was deemed the most relevant CRPC model to apply CRIME for interrogation the AR-V7 specific spliceosome: (1) high levels of AR-Vs are expressed in CWR22Rv1 cells and drive proliferation and survival pathways; (2) AR-V7 comprises the most prominent AR-V isoform in this cell line (Figure 62), making it the most suitable *in vitro* model to study AR-V7 specific splicing events.



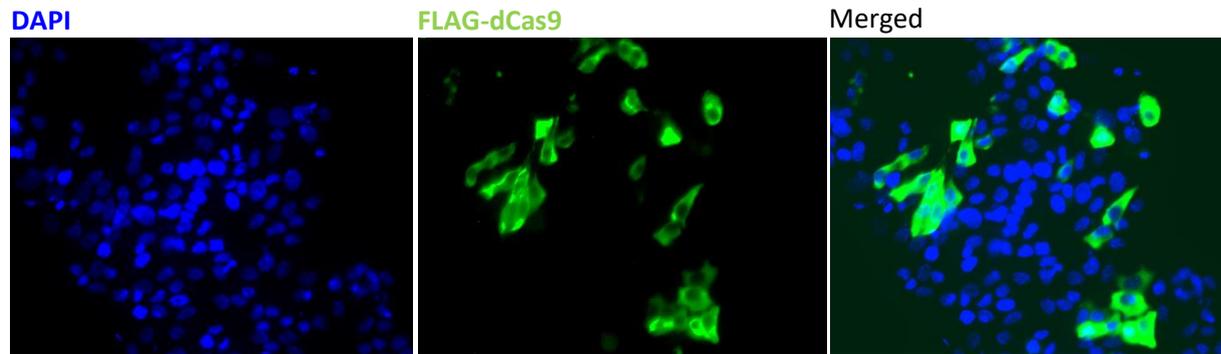
**Figure 62. AR-Vs constitute the most prevalent form of AR in CWR22Rv1 cells.** **A.** AR protein expression was profiled in PNT2C2, PC3, LNCaP, CWR22Rv1 and VCaP cells cultured in full (FM) or steroid-depleted (DCC) media as indicated. An N-terminal AR targeting antibody was used for immunoblotting. **B.** AR-V mRNA expression levels were profiled in CWR22Rv1 cells grown in full media for 48 h prior to qPCR analysis using primers specific to each AR isoform.

#### 6.3.2. Successful RNA-guided recruitment of dCas9 to the AR cryptic exon 3 chromatin region in CWR22Rv1 cells.

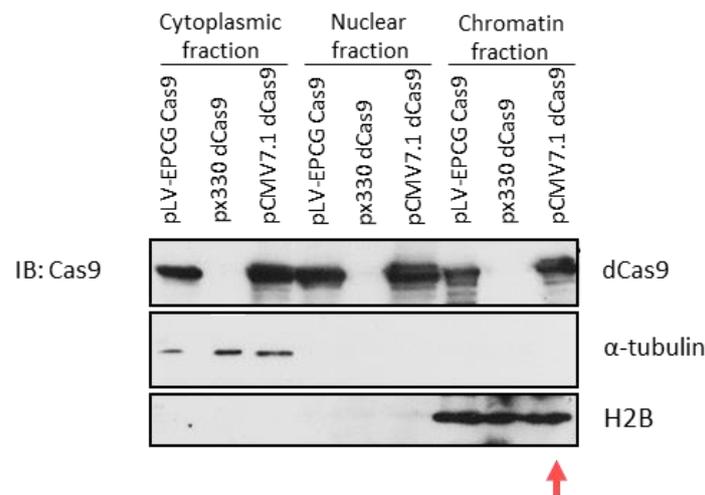
Localisation of dCas9 in the nucleus was critical for chromatin binding and subsequent application of the CRIME procedure. Despite the presence of 3 consecutive NLS signals at the N-terminal of dCas9 in the pCMV7.1 vector sequence, its localisation in the nucleus was assessed by immunofluorescence, incorporating a FLAG antibody, to detect ectopically-expressed FLAG-tagged dCas9 in CWR22Rv1 cells. (Figure 63A). In addition to IF, CWR22Rv1 cellular fractionation also indicated probably more robustly the presence of dCas9 in the

chromatin fraction (Figure 63B). Different Cas9-expressing vectors, such as the px330 and pLV-EPCG, were transfected into CWR22Rv1 cells in parallel experiments and were utilised as negative and positive controls (Figure 63B), expecting no or successful import of Cas9 in the nucleus, respectively.

**A.**



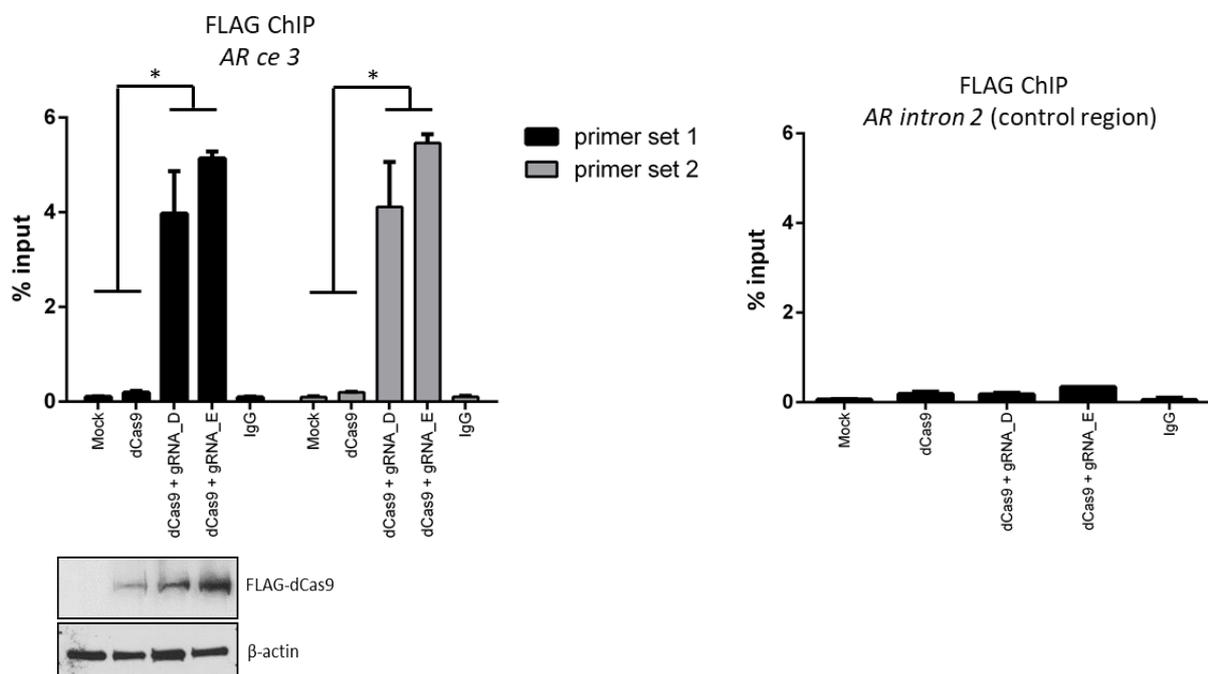
**B.**



**Figure 63. dCas9 successfully migrates to the nucleus and is present in the chromatin fraction of CWR22Rv1 cells. A.** Representative immunofluorescence images of CWR22Rv1 cells transfected with 3  $\mu$ g of pCMV7.1-3XFLAG-dCas9. A FLAG specific antibody was used for FLAG-dCas9 detection (top panel). **B.** Subcellular fractions of CWR22Rv1 cells transfected with 3  $\mu$ g of different Cas9/dCas9 expressing vectors were immunoblotted using a Cas9 specific antibody and  $\alpha$ -tubulin and H2B antibodies as nuclear and chromatin loading controls, respectively (bottom panel).

PC3 cells were first used to examine whether dCas9 can be successfully recruited to the AR gene CE3 chromatin region by two individual gRNAs. PC3 cells were chosen as they show high transfection efficiency and comprise a semi-relevant cell line model as they express detectable levels of AR pre-mRNA, meaning that the AR gene may conceivably reside in a loose chromatin region and access of dCas9 to the target site would not be blocked by tight

nucleosome complexes. A FLAG specific antibody was used for pulling down FLAG-tagged dCas9. This particular FLAG antibody has been used in similar experiments in the lab and has shown high specificity to FLAG-tagged proteins extracted from formaldehyde fixed cells. FLAG ChIP indicated that both gRNAs were capable of positioning dCas9 specifically to CE3 with high and selective efficiency as evidenced by negligible Flag immunoprecipitation in the IgG and dCas9 minus gRNA controls. Non-specific recruitment of dCas9 to an independent intron 2 region of the *AR* gene was also examined and as expected, no dCas9 was detected at that site (Figure 64).

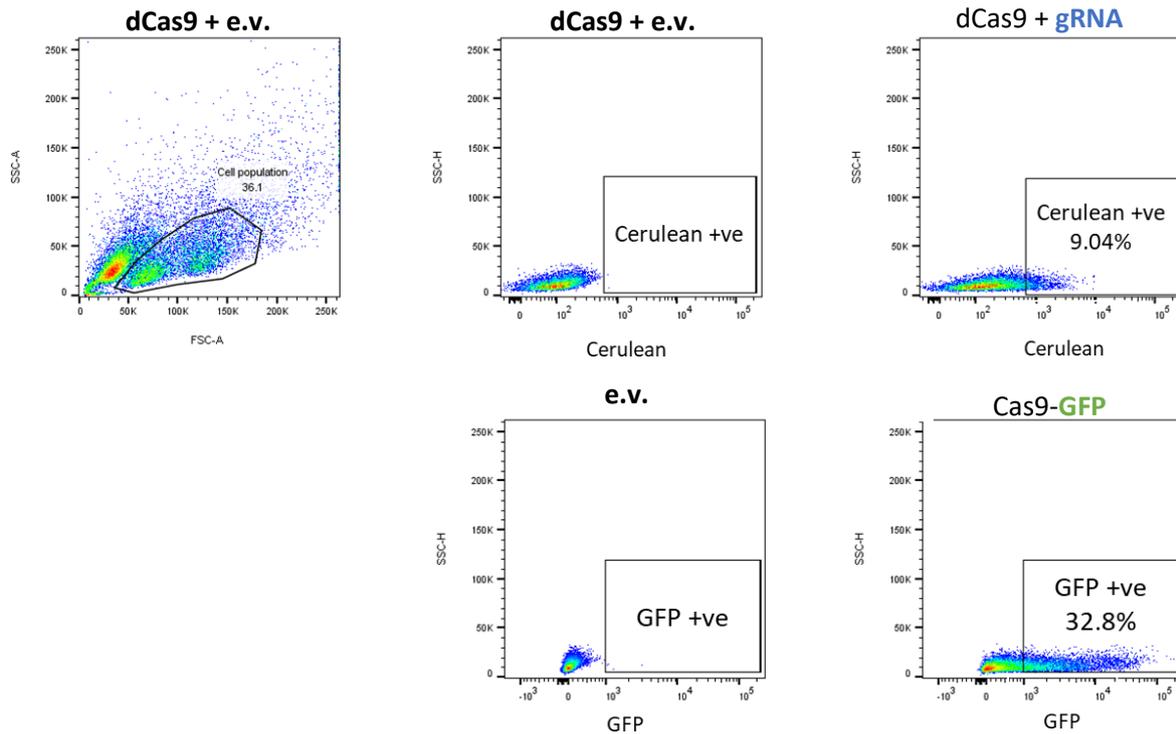


**Figure 64. dCas9 is successfully RNA-guided to the chromatin region of CE3 in PC3 cells.** PC3 cells transfected with 10  $\mu$ g of pCMV7.1-3xFlag-dCas9 and lentiCRISPR-iCer vectors were subject to FLAG ChIP analysis to assess FLAG-dCas9 recruitment to AR CE3 by two individual CE3-targeting gRNAs (gRNA\_D and gRNA\_E). Non-specific recruitment of the deficient nuclease was examined at intron 2 of the AR gene. dCas9 protein expression levels were assessed across the different experimental arms (mock, dCas9, dCas9+gRNA\_D, dCas9+gRNA\_E) using a FLAG specific antibody.

Having seen successful gRNA-driven recruitment of dCas9 to CE3 in PC3 cells, our efforts focused on getting similar read-outs using CWR22Rv1 cells which would be more challenging based on their low transfection efficiency. In a previous chapter (Chapter 4.3.5), nucleofection was applied to a small population of CWR22Rv1 cells ( $\sim 2 \times 10^6$ ) to perform genome editing and reasonable transfection efficiency was observed, but was accompanied by high rates of cell

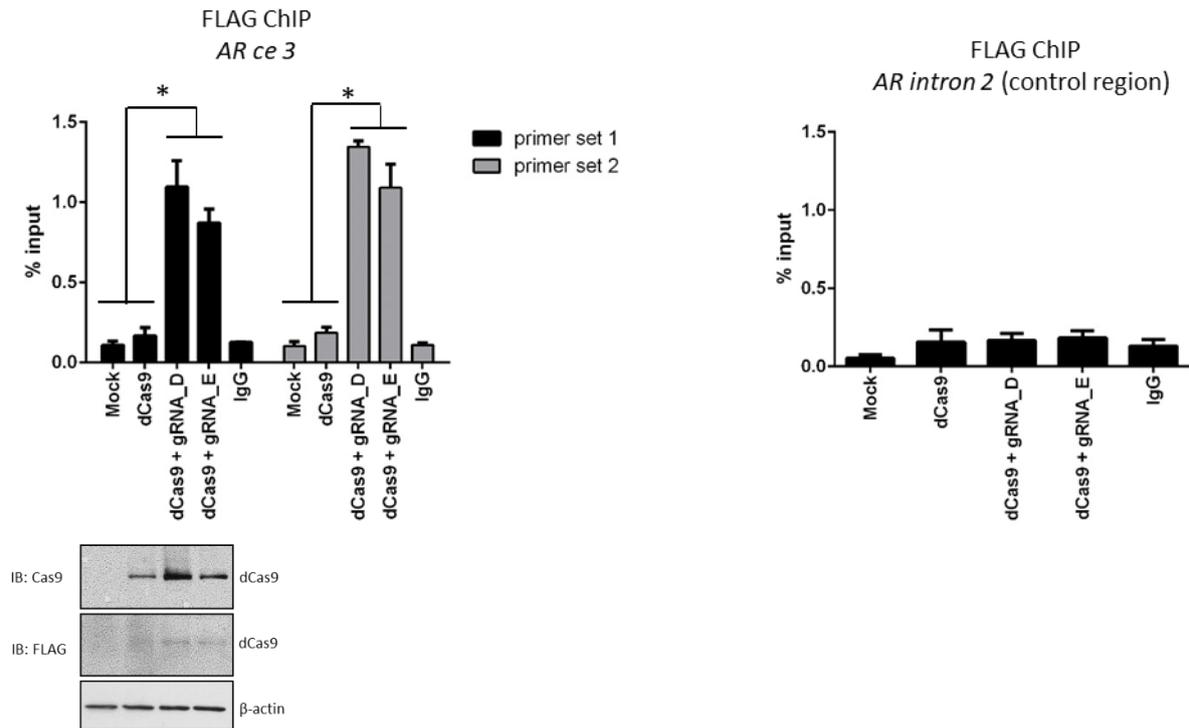
death (20-30%). For CRIME, however, a higher number of cells was required to obtain adequate material for downstream analysis by MS. Hence,  $15 \times 10^6$  cells were nucleofected with the dCas9 and gRNA\_D expressing vectors to assess recruitment of dCas9 to CE3. To our surprise, cell viability following nucleofection was low with 64% of the starting population dying which prevented further analysis (data not shown). Therefore, transfection using lipophilic reagents (lipofection) remained the only other available option within the timeframe.

To measure the transfection efficiency using this lipofection approach,  $4 \times 10^6$  cells were transfected with 10  $\mu\text{g}$  of dCas9 expressing vector and 10  $\mu\text{g}$  of gRNA\_D expressing vector. Transfection efficiency was measured at 48h by flow cytometry based on cerulean expression (blue fluorescence) derived from the gRNA expressing lenti-iCer vector, on the assumption that transfected cells will take up both the dCas9 and gRNA expressing vectors. An independent GFP expressing vector of similar size to lentiCRISPR-iCer was also transfected in CWR22Rv1 cells alongside the dCas9 expressing vector in a parallel experiment to serve as an extra control in which a strong fluorophore such as GFP was used. Despite the similar vector size, transfection efficiency varied between the GFP and cerulean positive cell populations, with respective positivity of 32.8% and 9.04% (Figure 65). This discrepancy may be explained by the fact that cerulean is known to be a weak fluorophore and low expressors may not have been detected, dropping significantly the observed transfection efficiency when in reality a larger cell population was transfected with exogenous DNA.



**Figure 65. CWR22Rv1 cells demonstrate relatively low transfection rates using conventional lipofection for dCas9/gRNA vector delivery.** CWR22Rv1 cells transfected with 10  $\mu$ g of pCMV7.1-3xFlag-dCas9 and lentiCRISPR-iCer gRNA vectors, or GFP-expressing vector were subject to flow cytometry analysis to sort for blue (cerulean) and green (GFP) positive cells respectively. 10,000 events were analysed.

Keeping the relatively low transfection rate in mind, dCas9 recruitment to CE3 was assessed by FLAG ChIP. Similar to PC3 cells, both gRNAs performed as expected and selectively guided dCas9 to CE3 (Figure 66). However, dCas9 enrichment was profoundly poorer in CWR22Rv1 cells compared to PC3 cells when the same conditions were applied across the entire experimental workflow. The observed difference however can be attributed to the lower transfection rates of CWR22Rv1 cells as discussed above.

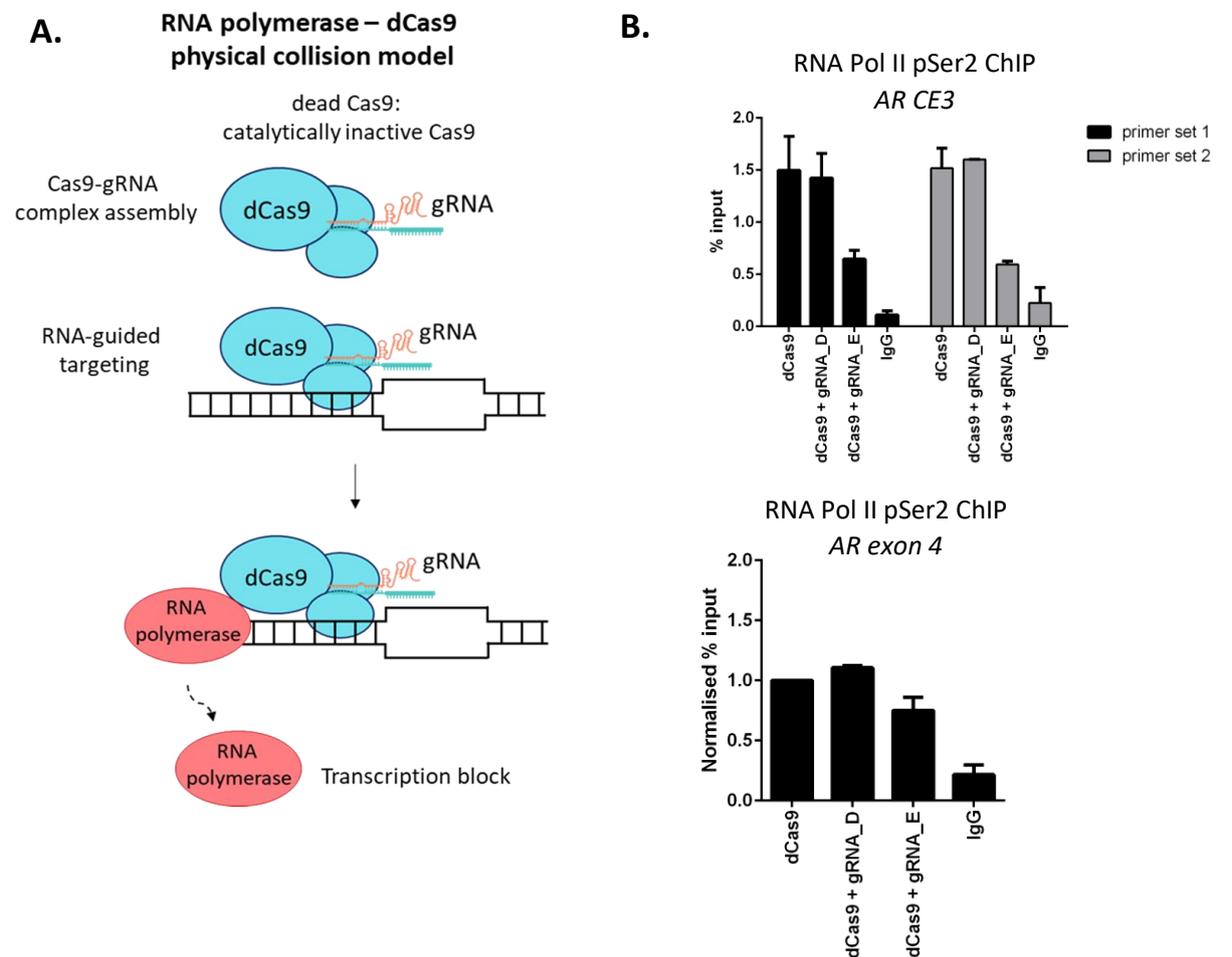


**Figure 66. dCas9 is successfully RNA-guided to the chromatin region of ce3 in CWR22Rv1 cells.** CWR22Rv1 cells transfected with 10  $\mu$ g of pCMV7.1 and lentiCRISPR-iCer vectors were subject to FLAG ChIP analysis to assess FLAG-dCas9 recruitment to AR CE3 by two individual CE3-targeting gRNAs (gRNA\_D and gRNA\_E). Non-specific recruitment of dCas9 was also examined at AR gene intron 2 (AR intron 2). dCas9 protein expression levels were assessed across the different experimental arms (mock, dCas9, dCas9+gRNA\_D, dCas9+gRNA\_E) by immunoblotting using a FLAG specific antibody.

### **6.3.3. The nuclease-deficient Cas9 blocks transcription in a strand-specific manner**

Cas9 is a large protein of  $\sim$ 160 kDa. It is conceivable to consider that its presence on chromatin may cause a roadblock for RNA Pol II elongation and compromise subsequent spliceosome formation. In this context, RNA Pol II would pause upstream of dCas9, unable to continue along the DNA template, and subsequently dissociate from chromatin due to physical collision with dCas9 (Figure 67A). That would therefore compromise massively the CRIME protocol. To test whether this is the case in our system, ChIP using an antibody to the actively elongating form of RNA Pol II (phospho-serine 2 (pSer2) RNA Pol II) was performed and the regions surrounding CE3 were qPCR amplified in order to assess potential RNA Pol II elongation blocking by dCas9. Surprisingly, one of the two gRNAs (gRNA\_D & gRNA\_E) used for recruitment of dCas9 to CE3, gRNA\_E, had a profound impact on the levels of RNA Pol II at the specific CE3 locus; diminishing enrichment by almost 50% compared to the dCas9-only control arm of the experiment. In contrast, RNA Pol II was not dissociated from chromatin

when gRNA\_D was used, as no drop in RNA Pol II levels at CE3 was observed (Figure 67B), indicating that different gRNAs may variably affect the transcription fate of a gene. To also rule out that RNA Pol II might be simply paused on chromatin but not dissociated, its presence was checked at sites downstream of the dCas9 binding site within exon 4 which would indicate maintained processivity of the enzyme (Figure 67B). No or minimal impact was observed at recruitment of RNA Pol II at the AR exon 4 region using gRNA\_D and gRNA\_E, respectively. This confirms that RNA Pol II is paused upstream of CE3 rather than dissociated from chromatin. Regardless, since RNA Pol II elongation was impacted by the gRNA\_E-containing dCas9 complex, gRNA\_E was dropped from the analysis and only gRNA\_D was used for downstream experimentation.



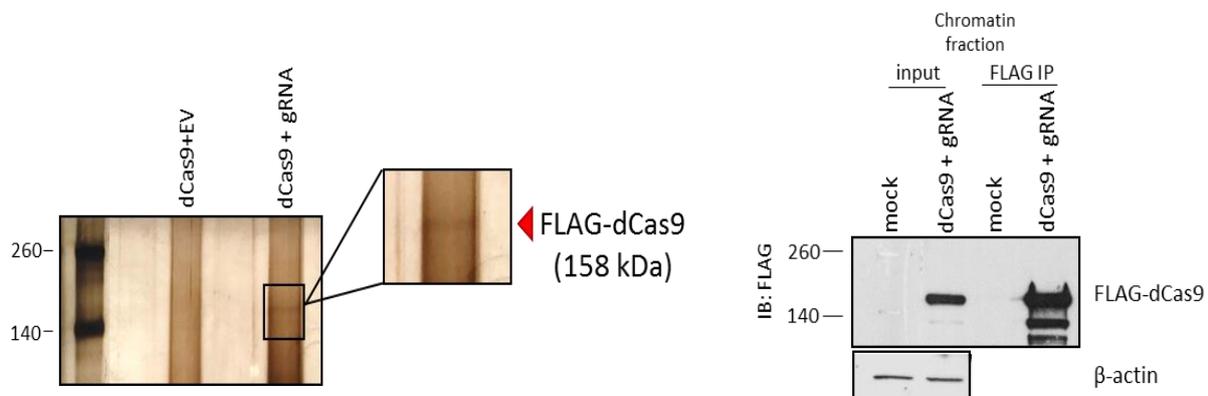
**Figure 67. RNA polymerase II dissociates from chromatin or engages with dCas9 to mediate transcription in a gRNA dependent fashion. A.** Graphic outline of the RNA Pol II-dCas9 physical collision model. The paused RNA Pol II occasionally aborts transcription elongation upon encountering the dCas9-gRNA roadblock. **B.** Actively elongating RNA Pol II recruitment to AR CE3 and

exon 4 chromatin regions was assessed by ChIP using a phospho-Ser2 RNA Pol II-specific antibody (RNA Pol II pSer2) and CE3- or exon 4-specific primer sets for post-ChIP qPCR analysis.

#### 6.3.4. Immunoprecipitation of FLAG-dCas9 yielded sufficient material for CRIME analysis

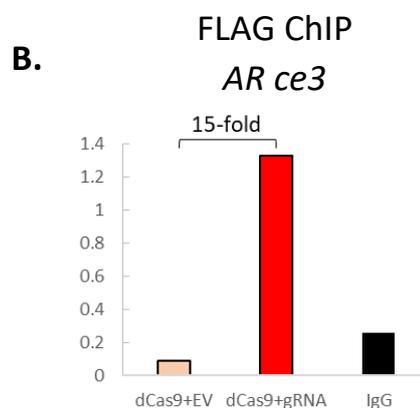
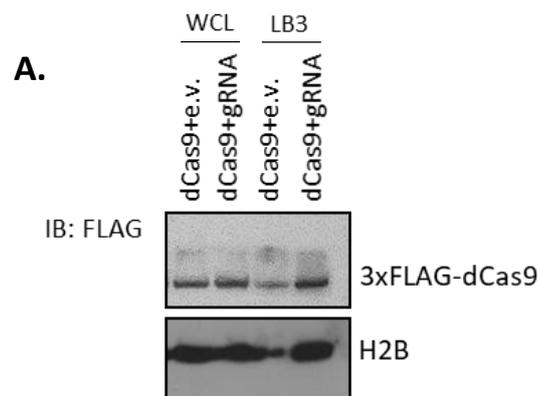
In light of our findings above, we subsequently performed CRIME using ectopically-expressed dCas9-gRNA\_D complexes in  $5 \times 10^6$  CWR22Rv1 cells to facilitate detection of bound proteins in downstream MS analysis. Cells were cultured in DCC media for 24h prior to transfection in an attempt to enforce AR-V splicing events and so enhanced recruitment of splicing factors a hypothesis based on the observation that enzalutamide induces recruitment of ASF/SF2 to AR-V7 splice sites (Liu et al., 2014). Chromatin was then extracted and sonicated, FLAG-dCas9 was immunoprecipitated and resulting IP samples from the dCas9+EV. and dCas9+gRNA arms were analysed on SDS-PAGE and assessed for dCas9 presence by Coomassie and silver staining.

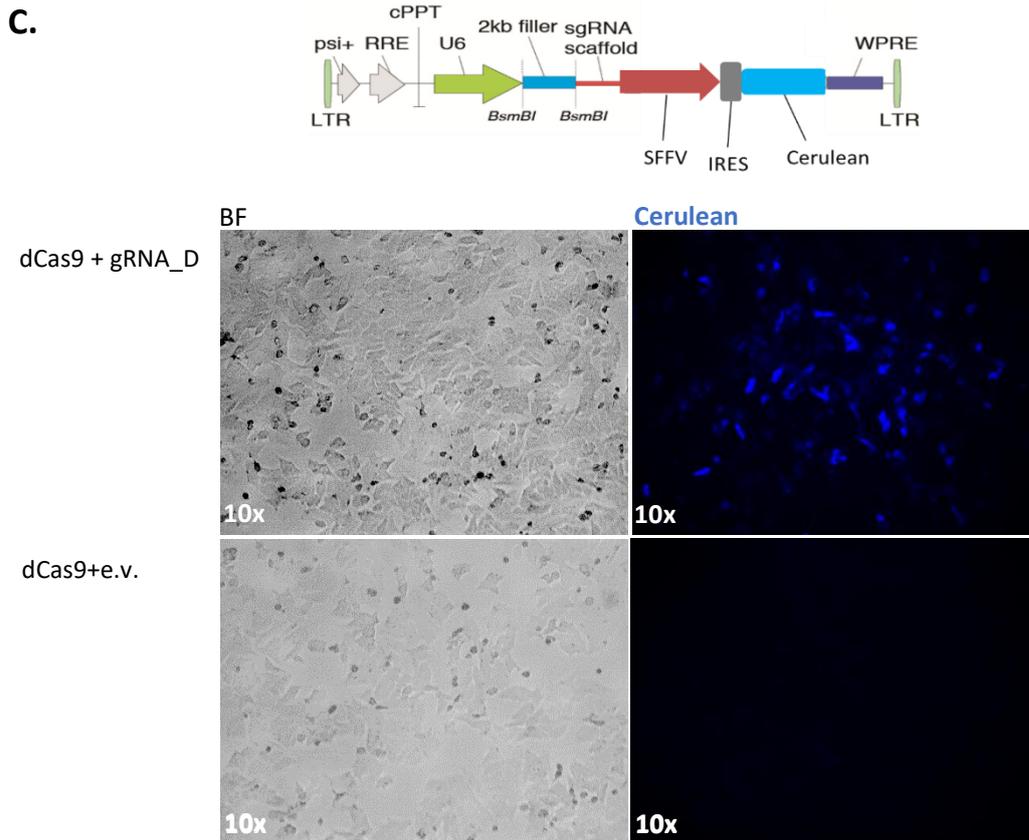
Coomassie staining of the gel was attempted but no bands were observed at the end of incubation (~3h) (data not shown). Due to its limited sensitivity, Coomassie staining was replaced by silver staining. As a rule of thumb, band detection by silver staining indicates sufficient amount of the pulled-down protein of interest which in turn permits downstream handling. Part of the lysate was immunoblotted in parallel to silver staining (Figure 68) to examine whether there is a match between the suspected dCas9 band detected by silver staining and the dCas9 band detected more reliably by a Cas9-specific antibody. Three dishes with  $5 \times 10^6$  cells per dish were set up for each condition (dCas9+EV, dCas9+gRNA\_D). In addition, chromatin input was 450  $\mu$ g per condition to ensure sufficient amount of immunoprecipitated dCas9.



**Figure 68. Immunoprecipitation of dCas9 from formaldehyde fixed cells.** CWR22Rv1 cells transfected with pCMV7.1-3xFLAG-dCas9 and gRNA\_D expressing vectors (dCas9+gRNA) or empty vector (dCas9+EV) were formaldehyde fixed, lysed, sonicated and subjected to IP using an anti-FLAG antibody. Resultant immunoprecipitates were subjected to SDS-PAGE followed by silver staining of the gel to confirm presence of dCas9 (left panel). Part of the dCas9+gRNA derived lysate was WB analysed alongside a control lysate (mock) derived from non-transfected cells to assess Cas9 antibody specificity (right panel).

Successful detection of dCas9 by silver staining was achieved (Figure 68). Following this, all the appropriate checks were performed prior to MS using the same cell population used for extracting the material for MS. More precisely, part of the cell lysate was used to confirm detection of dCas9 in the chromatin fraction by WB (Figure 69A) as well as sufficient dCas9 recruitment to CE3 by FLAG ChIP (Figure 69B). Finally, a small cell population was seeded on chamber slides to confirm gRNA expression by IF (Figure 69C). Having seen a 15-fold increase in dCas9 recruitment at CE3 in presence of gRNA\_D (Figure 69B), samples were approved for downstream MS analysis.





**Figure 69. Validation of dCas9 and gRNA expression and recruitment to AR CE3 prior to MS. A.** Whole cell lysates (WCL) and chromatin fractions derived from CWR22Rv1 cells transfected with the dCas9-expressing vector and either the gRNA\_D-expressing vector or an empty vector were immunoblotted with a FLAG antibody for FLAG-dCas9 protein detection and H2B was used as a loading control (top left). **B.** FLAG ChIP analysis was performed using the same chromatin fractions to assess RNA-guided dCas9 recruitment to AR CE3 over a dCas9 only (dCas9+e.v.) control and an independent IgG control (top right). **C.** A Fraction of transfected CWR22Rv1 cells seeded in chamber slides were subject to fluorescence imaging for cerulean detection (blue) and gRNA expression verification. Representative images at 10x magnification are shown (bottom).

### 6.3.5. dCas9 was successfully detected by LC-MS/MS

Prior to slicing the gel and moving on to LC-MS/MS analysis of all the individual gel slices (n=9 per sample), the dCas9 band was excised from the gel and analysed by MS. Peptides corresponding to dCas9 were identified with high confidence (low log(e) value) in the dCas9+gRNA\_D and dCas9+e.v. arms (Table 20), confirming successful and specific pull-down of dCas9. The rest of the gel was then sliced and each slice was analysed by MS. The protein species identified in both experimental arms (dCas9+gRNA\_D and dCas9+e.v.) were excluded as they may have been artefacts of non-specific binding of dCas9 to chromatin, possibly due to its high expression levels which may have led to excessive stickiness even in the absence

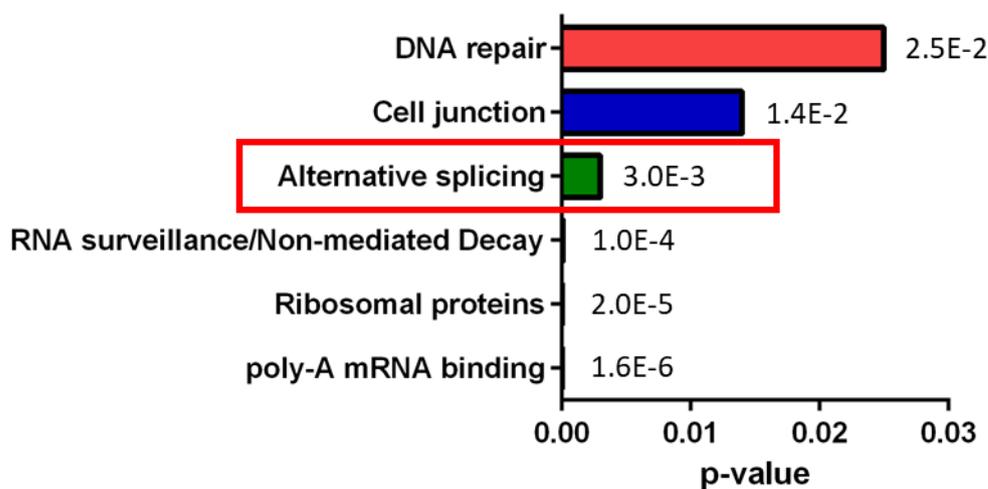
of gRNA. Additionally, a cut-off of -2 was set to omit any peptides detected by chance. The detailed protein list can be found in Appendix F. To our surprise, a very limited number of proteins met the selection criteria in the dCas9+gRNA\_D arm and this could be attributed to a number of factors discussed in section 6.4.

**Table 20.** Identification of dCas9 peptides by LC-MS. (I) – sum of raw spectrum intensities, rI – number of peptides detected, log(e) – expectation of detecting the protein stochastically, pI – isoelectric point of the intact gene product, Mr – mass of intact gene product.

| Sample         | log(I) | rI | log(e) | Mr  |
|----------------|--------|----|--------|-----|
| dCas9 + gRNA_D | 3.84   | 9  | -57.5  | 158 |
| dCas9 + e.v.   | 3.76   | 9  | -61.1  | 158 |

### 6.3.6. RNA-associated proteins were enriched using CRIME

The protein list derived from the dCas9+gRNA\_D arm, after applying the exclusion criteria mentioned above, was subject to gene ontology analysis to identify the pathways and subsequently the processes that the enriched proteins mediate. Distinct RNA associated processes were highlighted in the analysis such as alternative splicing, RNA surveillance via non-sense mediated decay and poly-A mRNA binding (Figure 70), confirming the validity of the model and its capacity to be used as an RNA splicing interrogation tool.



**Figure 70.** Pathway enrichment analysis of CRIME derived proteins.

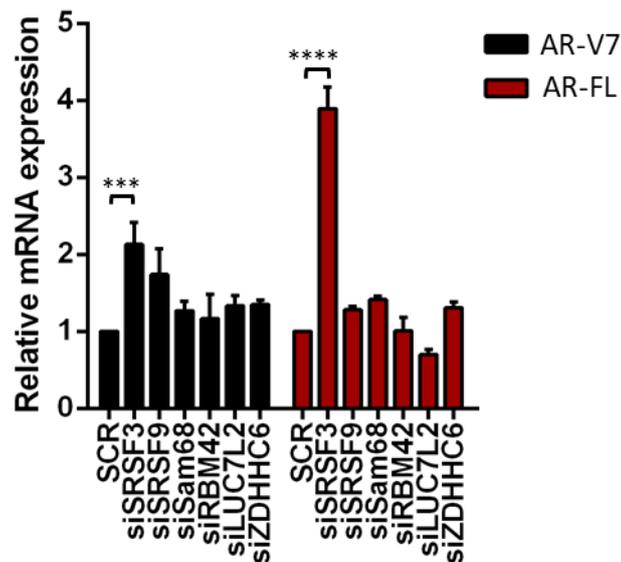
### 6.3.7. Splicing factor hit validation

Several CRIME-derived candidates were chosen to be further examined for their potential role in modulating AR-V7 mRNA splicing as they indicated apparent links to RNA processing and pre-mRNA splicing events. More precisely:

- Sam68 (UniProt: Q07666) is an RNA-binding protein which contains a KH domain and is implicated in alternative splicing of numerous pre-mRNA substrates (Bielli et al., 2011). Stockley *et al.* have demonstrated that Sam68 mediates AR splicing and potentiates AR-V7 mRNA generation (Stockley et al., 2015). Hence, isolation of Sam68 using CRIME provided great confidence that our proposed experimental approach to study AR splicing is valid.
- SRSF9 (UniProt: Q13242) is a member of the Serine/Arginine (SR)-rich protein family of splicing factors. It's an important component of the constitutive splicing machinery and can modulate the selection of alternative splice sites. Similarly to SRSF3, SRSF9 regulates GR splicing in response to cortisol and DHEA (Buoso et al., 2017).
- RBM42 (UniProt: Q9BTD8) is an RNA-binding protein which may be required for assembly of an active spliceosome alongside U4/U6 & U5 snRNPs (Suvorova et al., 2013).
- LUC7L2 (putative RNA-binding protein Luc7-like 2) (UniProt: Q9Y383, GeneCards ID: GC07P139344), similarly to SRSFs, can bind RNA via its SR-rich motifs, contains C2H2 zinc-finger motifs which are required for DNA binding and is associated with the generation of multiple alternatively spliced isoforms of mRNA transcripts. It may be implicated in the recognition of non-consensus splice sites in association with the U1 snRNP spliceosomal subunit.
- ZDHHC6 (zinc finger DHHC-type containing 6) (UniProt: Q9H6R6) is not strongly linked to splicing. However, it is a CCCH-type zinc-finger protein and such proteins such as U2AF have been involved in regulating pre-mRNA splicing (Lai et al., 2000).

To address the contribution of each of the aforementioned splicing mediators to AR-V7 mRNA production, siRNA mediated knockdown experiments were performed in CWR22Rv1 cells

grown in DCC media. DHT stimulation for 24h was also performed to examine whether SRSF3 regulates AR-FL mRNA levels. Each candidate was depleted by a specific siRNA for 48h and levels of AR-V7 and AR-FL mRNA were measured by qPCR. Out of 6 candidates interrogated, only one, SRSF3 (UniProt: P84103) appeared to significantly alter AR-V7 mRNA levels as almost double the amount of AR-V7 mRNA was detected when SRSF3 was depleted. Interestingly, SRSF3 knockdown resulted in a remarkable increase in AR-FL mRNA levels (Figure 71).

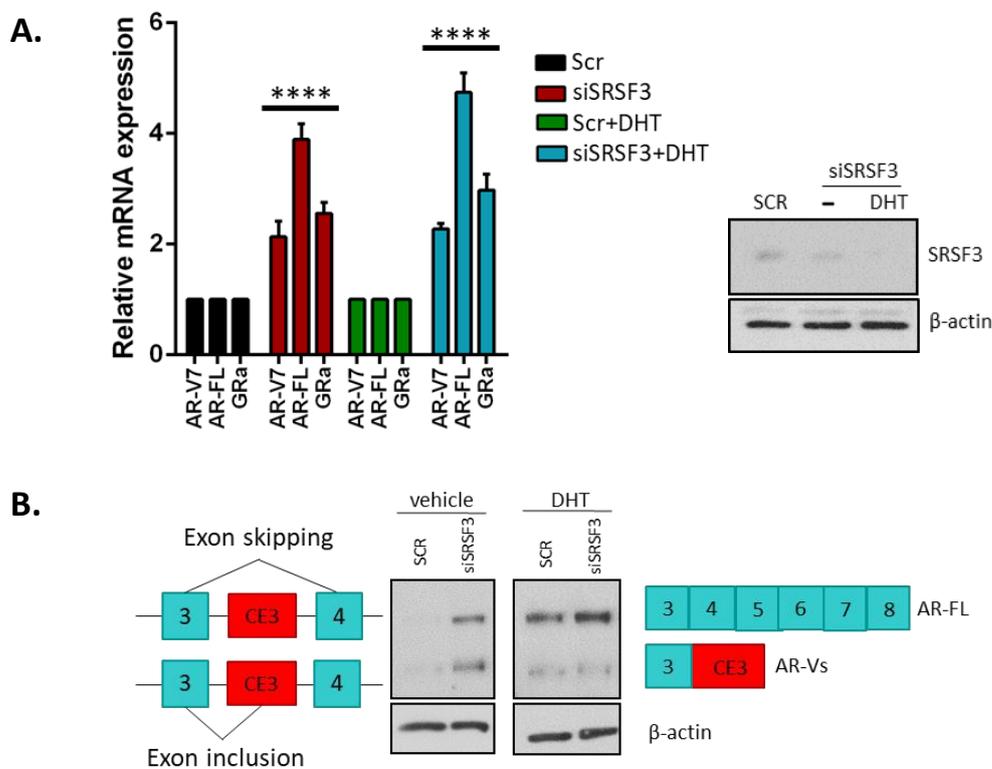


**Figure 71. CRIME hit validation.** CWR22Rv1 cells grown in DCC media were transfected for 48h with individual siRNAs against each CRIME-derived hit. AR-V7 and AR-FL mRNA levels were measured by qPCR.

### 6.3.8. SRSF3 modulates AR-V7 mRNA levels

As discussed in the previous section, SRSF3 depletion led to a significant increase in AR-V7 mRNA levels in CWR22Rv1 cells (Figure 72A), suggesting that SRSF3 and AR-V7 are mutually exclusive in a way that when SRSF3 is expressed, it prevents incorporation of CE3. Irrespective of their relationship, SRSF3 seems to be a novel AR-V7 splicing factor with no evidence of direct or indirect implication in AR biology reported before. To further investigate this novel avenue of AR-V7 modulation, the AR genomic sequence was interrogated for the presence of SRSF3 binding sites. The canonical SRSF3 binding motif CCAGCC was found several times at positions -628, +527 and +614 in the intronic regions downstream and upstream of CE3 respectively, an observation which further validates our findings and supports our assumption that SRSF3 is a genuine splicer of CE3.

SRSF3 is also implicated in regulating splicing of other hormone receptors. GR pre-mRNA is a known substrate of SRSF3. In particular, SRSF3 levels increase in response to cortisol and this leads to selection of isoform alpha over beta (Buoso et al., 2017). Therefore, GRalpha mRNA levels were also profiled following depletion of SRSF3 in CWR22Rv1 cells, which express endogenous levels of GR. Consistent with reports, SRSF3 depletion resulted in a significant increase in GRalpha mRNA levels (Figure 72A). Moreover, SRSF3 seemed to also modulate AR-FL levels in a similar fashion to AR-V7 indicating that SRSF3 is not exclusive to AR-V7 splicing. As expected, AR-V7 and AR-FL protein levels were also increased in response to SRSF3 knockdown, irrespective of androgen stimulation (Figure 72B).



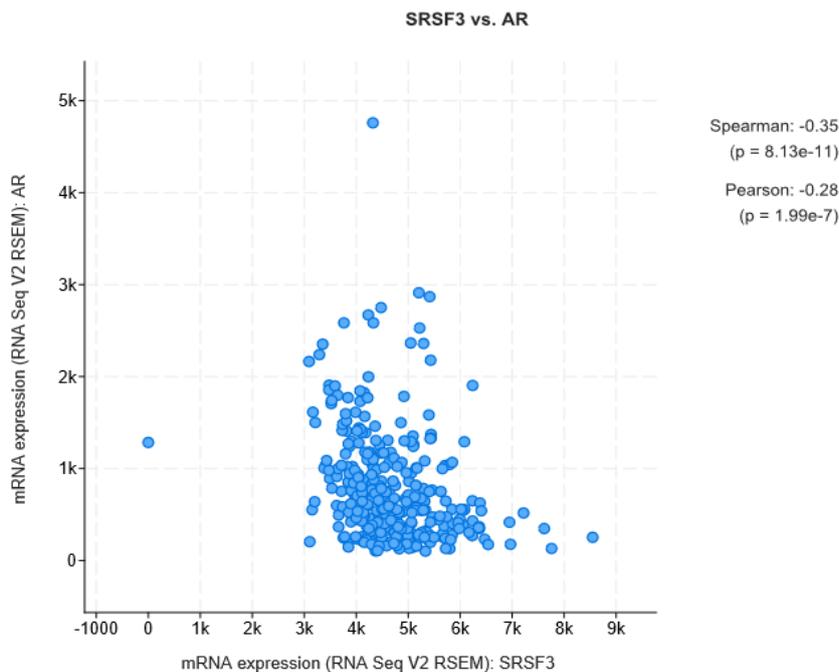
**Figure 72. SRSF3 modulates AR-V7 mRNA levels.** CWR22Rv1 cells grown in DCC media were transfected for 48h with control (Scr) or SRSF3-targeting siRNAs (siSRSF3). Cells were then lysed and AR-V7, AR-FL and GRa mRNA levels were measured by qPCR. Knockdown efficiency of SRSF3 was assessed at protein level using a SRSF3-specific antibody. AR-FL and AR-V protein levels were assessed by immunoblotting using an N-terminal targeting AR antibody.

### 6.3.9. SRSF3 expression is downregulated in mCRPC and negatively correlates with AR expression

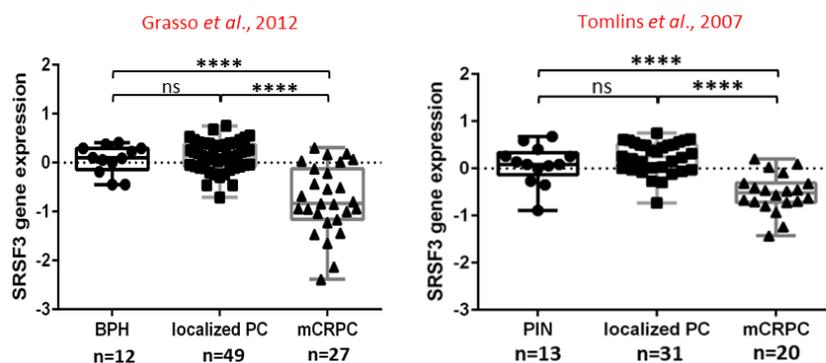
To better understand how SRSF3 controls AR pre-mRNA fate, *in silico* analysis was performed using the PRAD-TCGA dataset, as well as two well-accepted transcriptomic datasets described

by Grasso *et al.* and Tomlins *et al.* to assess expression of SRSF3 and AR. Opposing trends were observed between the mRNA levels of total AR and SRSF3 in 333 prostate cancer biopsies (Figure 73A) confirming our *in vitro* observations that when SRSF3 levels are depleted, total AR levels elevate. Finally, SRSF3 levels are significantly downregulated in mCRPC compared to localised prostate cancer and BPH in two independent datasets (Figure 73B). However, no information involving patient survival rates were found.

**A.**



**B.**



**Figure 73. SRSF3 expression is downregulated in CRPC and negatively correlates with AR expression.** **A.** Regression analysis of AR and SRSF3 mRNA levels was extracted from RNA seq analysis of 333 prostate cancer biopsies within the PRAD-TCGA dataset. **B.** Box plots of SRSF3 mRNA expression levels in BPH, localised PC and mCRPC biopsies. Data was derived from two publicly available datasets GSE35988 (Grasso *et al.*, 2012) and GSE6099 (Tomlins *et al.*, 2007).

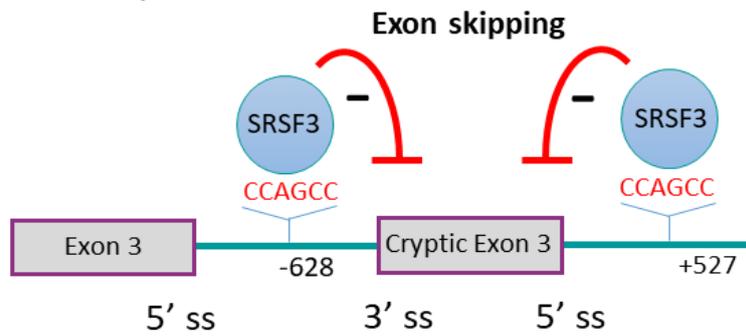
### ***6.3.10.SRSF3 is a potential tumour suppressor in CRPC***

Having seen that SRSF3 loss resulted in higher levels of AR-V7 mRNA, we can assume that SRSF3 would normally act as a suppressor of AR CE3 inclusion in the mature AR-V7 mRNA transcript. Given that AR-V7 is driving tumour progression, factors that disrupt its production could be considered potential tumour suppressors. This assumption is in line with the SRSF3 expression pattern in prostate cancer, as SRSF3 mRNA levels are significantly compromised in patients with mCRPC like many other tumour suppressor genes.

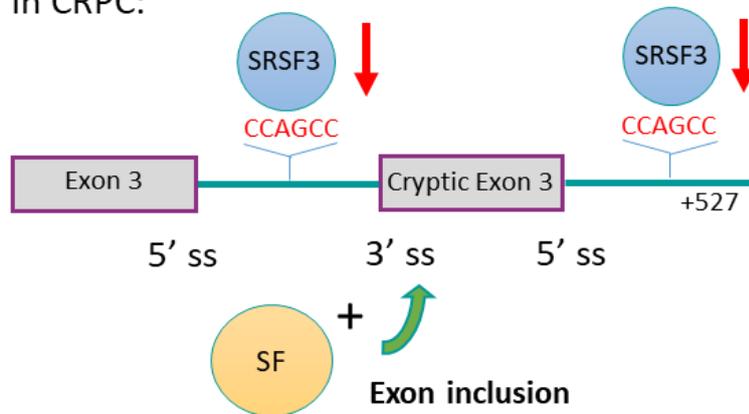
We hence propose that in normal prostate, SRSF3 regulates AR splicing and mediates CE3 skipping upon binding to specific sites in the intronic regions surrounding CE3; hence blocking AR variant synthesis. In mCRPC, SRSF3 levels drop, hence cryptic exon 3 inclusion is permitted and gives rise to AR-V7 (Figure 74). Despite our focus on AR-V7 mRNA, the effect of SRSF3 depletion on AR-FL mRNA levels was equally remarkable, suggesting that SRSF3 may exert multiple functions on the AR pre-mRNA by also regulating AR pre-mRNA stability.

## AR pre-mRNA

Normal prostate:



In CRPC:



**Figure 74. Schematic of the proposed model of SRSF3's role in AR-V7 pre-mRNA processing.** SRSF3 binds to the CCAGCC sequences upstream and downstream of ce3 (highlighted in red) and promotes skipping in normal prostate. Expression of SRSF3 is downregulated in mCRPC and hence ce3 inclusion can occur, potentially promoted by other SFs, and gives rise to AR-V7. ss: splice site, SF: splicing factor

#### 6.4. Discussion

The mechanism by which AR-V7 mRNA is generated might be a novel avenue to eliminate such a challenging target. However, very little is known about the precise processes that regulate the emergence of this particular transcript. There has been a number of splicing factors associated with generation of AR-V7 mRNA in CRPC. ASF/SF2 seemed to be a promising candidate for drug targeting. ASF/SF2 was found to lie in *cis*-regulatory regions around AR CE3 and its abundance at this locus was elevated upon treatment with anti-androgens (Liu et al., 2014). Additionally, transient knockdown of ASF/SF2 in VCaP cells remarkably reduced AR-V7 mRNA copies, albeit suggesting that ASF/SF2 binds to and regulates AR CE3 inclusion. However, the findings of this particular study were disputed by Dehm's group who demonstrated that ASF/SF2 has no impact on AR-V7 mRNA levels in CWR22Rv1 cells (Van Etten et al., 2017) which comprise a more physiological model to study AR-V7 compared to VCaP cells. On the other hand, cell line as well as locus specificity should certainly be taken into account, as VCaP cells display a 20-fold amplification of the AR gene which may impact on the AR pre-mRNA splicing requirements and contribute to the observed discrepancy between VCaP and CWR22Rv1 cells. Despite the controversy, ASF/SF2 remains an attractive target for therapeutic intervention. According to a recent study (Denichenko et al., 2019), sense RNA oligonucleotides known as decoy oligonucleotides were developed to target directly ASF/SF2 and exclusively block its splicing activity without interfering with other functions of the SF, nor with its RNA binding site.

It was acknowledged that identifying splicing mediators of AR-V7 pre-mRNA using an unbiased and more high-throughput approach to individual siRNA knockdowns would fill a huge knowledge gap in the AR-V7 field. Therefore, we developed a workflow which exploited a more delicate version of the conventional CRISPR technology for genome editing. This kind of approach has been applied before (Fujita and Fujii, 2016) to identify proteins which interact with specific loci on chromatin. However, it is the first time that this concept was applied to identify splicing regulators. It relies on the fundamental concept of coupled transcription and pre-mRNA splicing which enables SFs to be in the vicinity of genomic intronic/exonic boundaries and hence would be detected by CRIME.

Its development required several optimisation steps and raised precautions with regards to physical contact between dCas9 and RNA Pol II, an essential component of the coupled process of transcription and splicing. A striking observation led us to the conclusion that it's the gRNA that dictates the fate of this interaction. In particular, the position of the gRNA on DNA determines whether the RNA Pol II will elongate through and transcribe the gene of interest or whether it will be blocked by the CRISPR complex. We observed that the gRNA that targeted the template stand of DNA, which is copied during transcription, forced RNA Pol II to abort transcription elongation, whereas the gRNA that targeted the non-template strand did not impact on transcriptional elongation by the polymerase. This finding contrasts with a previous study which showed that it is the gRNA on the non-template strand that blocks transcription elongation by RNA Pol II (Qi et al., 2013). This study, however, was conducted in *Escherichia coli* cells using an RFP reporter assay to measure transcriptional activity in the presence of dCas9. The discrepancy therefore might be caused by the different transcription mechanisms between prokaryotes and eukaryotes.

AR gene transcription is enhanced upon ADT and this offers the opportunity for robust contacts between the AR pre-mRNA and splicing factors (Liu et al., 2014). We applied this logic to the CRIME technique by growing CWR22Rv1 cells in DCC media prior to transfection with the catalytically-deficient CRISPR complex and for 48h after to allow for that time and interaction window to occur.

The CRIME-derived candidates were subject to a mini knockdown screen to validate whether they comprise genuine AR-V7 splicers. Despite the apparent enrichment of splicing proteins in the isolated complexes, only one candidate had a robust effect on AR-V7 pre-mRNA levels and that was SRSF3. However, the unaltered AR-V7 mRNA levels observed following individual depletion of the other CRIME-identified candidates does not rule out that these proteins are genuine components of the AR-V7 spliceosome. The observed failure to modulate AR-V7 mRNA production in response to their knockdown could be attributed to the fact that these proteins may be less crucial in mediating AR splicing and their loss can be replenished by functional compensation and adaptation events such as upregulation of other factors with similar function.

SRSF3 or SRp20 is a member of the SR-rich protein family. It comprises an essential splicing factor which mediates alternative splicing events by promoting exon inclusion/exclusion or

selection of alternative splice sites. Depending on the mRNA isoform being produced, SRSF3 can exert oncogenic or tumour suppressive functions. For example, generation of ILF-3 isoforms 5 and 7 leads to tumour cell apoptosis and is suppressed by SRSF3 in cancer cells. On the contrary, generation of ILF-3 isoforms 1 and 2 is associated with elevated tumour cell proliferation and is mediated by SRSF3 (Jia et al., 2019).

SRSF3 is overexpressed in cancers including breast, lung and colon. SRSF3 acts as a proto-oncogene in these tumours by promoting the production of mRNA isoforms with oncogenic properties including CD44 and ASF/SF2 (Anczukow and Krainer, 2016). For instance, SRSF3 is responsible for alternative splicing of PKM (pyruvate kinase muscle). PKM isoform 2 (PK-M2) is generated upon exon 10 inclusion in colon cancer cells maintaining a metabolic environment which favours cancer cell growth (Kuranaga et al., 2018). However, this is not the case in CRPC. Our observations indicate that SRSF3 may act as a silencer of *AR* CE3 by promoting skipping from the mature mRNA transcript and hence functions as a tumour suppressor. No further characterisation of SRSF3 in PC has been attempted before though.

SRSF3 has been reported to exert multiple functions, other than mediating alternative splicing. Early observations have indicated that SRSF3 is responsible for mRNA 3' end formation. According to these reports SRSF3 binds to its pre-mRNA substrate, recruits polyadenylation factors to alternative polyA sequences proximal to 3' terminal exons to promote production of truncated transcripts (Shen et al., 2019). This resembles the CPSF1 scenario in CRPC. CPSF1 is a polyadenylation factor which binds to the alternative polyadenylation signal present in *AR* intron 3 and causes early transcription termination resulting in production of AR-Vs including AR-V7. It is conceivable to speculate that SRSF3 which also binds to *AR* intron 3 may prevent interaction of CPSF1 with the polyA signal in intron 3 when it is expressed in non-malignant prostate. However, loss of SRSF3 in CRPC may allow free access of such factors to this alternative polyA leading ultimately to higher AR-V7 transcript production (Van Etten et al., 2017). Interestingly, PTEN mRNA polyadenylation demonstrates the proposed polyadenylation pattern. In fact, it has been shown that SRSF3 represses proximal polyA sites of PTEN mRNA in murine and human cells, whilst its loss potentiates selection of polyA sites that give rise to shorter mRNA transcripts (Shen et al., 2019).

Further validation experiments are required to understand the exact role of SRSF3 in regulating AR-V7 mRNA processing. The potential binding sites identified in AR introns 2 and 3 could be mutated in order to establish whether they comprise genuine SRSF3 specific motifs or whether SRSF3 is part of a broader protein complex not necessarily making direct contact with the mRNA substrate. An AR-V7 minigene experiment in response to SRSF3 knockdown would also indicate whether the latter promotes CE3 skipping and would hence validate our previous observations of elevated AR-V7 mRNA copies in CWR22Rv1 cells depleted of SRSF3. This would allow interrogation of AR-V7 splicing in other cellular backgrounds as well.

From a translational stand-point, if SRSF3 activity could be restored, AR CE3 inclusion would be blocked and subsequently AR-V7 generation would be prevented. Unfortunately, there are no known SRSF3 activators which could be used to treat cells in culture and see whether this hypothesis can be applied *in vitro* and *in vivo*. The only known activator so far is cortisol which triggers SRSF3 expression in MDA-MB-231 breast cancer cells (Buoso et al., 2019). However cortisol also activates GR which has been shown to have at least some degree of overlap with AR in relation to gene targets (Arora et al., 2013).

A higher number of MS-derived hits would be ideal so more candidates could be explored and validated for their potential role in mediating AR CE3 splicing. Despite the fact that CRIME provided proof-of-principle observations and turned out to be a valid method to study spliceosome assembly, it was only performed once limiting the listing of candidates alongside the tight MS stringency scores and cut-offs that were applied. Low confidence in the MS run could be attributed in part to the low transfection efficiency of CWR22Rv1 cells, as well as to the low abundance of the AR CE3 site per cell. Less dCas9 molecules per  $\mu\text{g}$  of chromatin were expected compared to a nuclear receptor such as AR which is normally distributed at different sites across the genome. dCas9 is only recruited to its single target site guided by the corresponding gRNA. Hence theoretically, the higher the chromatin input the more dCas9 will be pulled-down. Therefore, the amount of chromatin input per IP was maximised, but it was also limited by the amount of starting material (i.e. cell number), antibody affinity and immunoprecipitation efficiency.

A promising approach which could be applied in order to overcome such weaknesses and improve IP yield is the positioning of more than one gRNAs in and around AR CE3 to create overlapping gRNA tiles so more than one dCas9 molecules can simultaneously be recruited to

the same CE3 site. A similar approach exploiting paired gRNAs has been applied previously (Ahmed and He, 2017).

In addition, splicing is a very dynamic process with constant assembly of complexes around the splice site of interest and only a snapshot of this is captured by CRIME. Sequential IPs should be performed in order to timely monitor the assembly of the various protein complexes which regulate the splicing events that take place in and around *AR* CE3.

Altogether, this piece of work comprises a proof-of-concept approach to study spliceosome formation at *AR* CE3 and subsequent *AR*-V7 mRNA formation. It was an initial attempt to find out whether this kind of approach could be applied to our cell line model, to also highlight the appropriate experimental conditions to maximise efficiency and output and to finally generate preliminary read-outs which could be followed up with more biological repeats and robust validation of identified targets.

It is therefore apparent that this is an ongoing piece of work with great potential to identify robust preclinical targets to be pursued.

## Chapter 7. Key conclusions and future work

Restored AR activity is a hallmark of CRPC. To date, AR-Vs remain refractory to ADT and next-generation antiandrogens and drive androgen-independent tumour progression.

Lack of disease-relevant preclinical models which reflect AR-V specific co-regulator dependencies as well as the AR-V specific transcriptome and cistrome has made it challenging to define the precise role of AR-Vs in CRPC. This is, in part, a consequence of not being able to fully distinguish between the activities of AR-Vs and AR-FL which are co-expressed in cell lines such as CWR22Rv1 and VCaP (Liu et al., 2014). The development of the TALEN-engineered R1-D567 cell line which expresses the single receptor splice variant ARv567es has been a valuable addition to the PC model toolbox (Nyquist et al., 2013); providing an important insight into co-regulator requirements of AR-Vs, such as dependency on the BET family of bromodomain-containing proteins (Chan et al., 2015). There remains, however, a need to have additional cell lines that recapitulate the clinical scenario. Given that CWR22Rv1 cells express multiple AR-Vs, a phenomenon observed in CTCs (De Laere et al., 2017), it is important that new models express several clinically-relevant AR-Vs to mimic CRPC.

To address this, a CRISPR-engineered CRPC cell line derivative, modelled in CWR22Rv1 cells, that has lost AR-FL expression, but retains expression of all endogenous AR-Vs was generated. This new cell line, termed CWR22Rv1-AR-EK (AR-exon knockout) has an edited *AR* gene containing a knock-in stop codon to prevent synthesis of AR-FL protein, and wild-type exons encoding the N-terminal transactivation and DNA-binding domains, to enable expression of all AR-Vs nascent to the parental cell line.

There has been much controversy regarding the importance of AR-FL for AR-V activity. The work presented here indicates that AR-Vs do not require AR-FL for transcriptional activity, a finding which is consistent with Hu et al. (Hu et al., 2012). AR-Vs are sufficient to drive gene transcription in the absence of AR-FL. Importantly, functional annotation of the CWR22Rv1-AR-EK AR-V-driven gene signature provided evidence for a role of AR-Vs in cell cycle regulation and mitotic pathways which is consistent with both Jones *et al.* (Jones et al., 2015) and He *et al.* (He et al., 2018) suggesting that the core overlapping genes from the distinct AR-V transcriptomes play key roles in regulating cell fate. However, the different experimental approaches utilised to assess AR-V activity between the studies, particularly with respect to

how AR-FL was inactivated, the degree of AR-V knockdown and RNA sequencing versus microarray platforms, is likely to contribute to a considerable degree of variation in the overall numbers of differentially expressed genes reported. Nevertheless, it is apparent that a well-conserved core of AR-V-regulated genes exists and both cell proliferation and clonogenic assays validated AR-Vs as key regulators of cell growth and viability in CWR22Rv1-AR-EK cells.

Outside of cell cycle regulation, one of the other highly ranked AR-V-regulated pathways identified by functional annotation was DNA repair. This was an exciting observation given the number of recent reports describing a role for AR-FL as a regulator of the DDR. Namely, Polkinghorn *et al.* and Asim *et al.* have both highlighted DDR associated gene signatures defined by AR in LNCaP cells (Asim *et al.*, 2017; Polkinghorn *et al.*, 2013). AR-Vs mediate DNA repair upon IR and hence confer resistance to radiation therapy, highlighting that AR-Vs drive tumour progression. Interestingly, no DDR associated genes were regulated by AR-V binding in two recent studies (Ling Cai *et al.*, 2018; Chen *et al.*, 2018). Both studies focused on solely investigating AR-V7 transcriptomes and cistromes using an AR-V7 specific antibody and siRNA oligos. However, it is now well-established that multiple AR-Vs are co-expressed in cell lines, tissue specimens and CTCs (De Laere *et al.*, 2017). Co-expression of AR-Vs in the same patient may indicate that AR splice variants form a more complex interaction network (with one another) to what was initially thought and mediate signalling in certain combinations. In other words, more than one AR splice variants may be required for the same function to be executed.

Taking this into account, our study did not solely focus on AR-V7, but instead aimed to look at all AR-Vs expressed in CWR22Rv1 AR-EK cells and how they function in concert (we siRNA depleted all AR-Vs exploiting the common poly-A sequence they share). The transcriptomic profile presented in this study reflects the activity of all variants (either functioning individually or in combination) whereas the two aforementioned studies look specifically at AR-V7 which may not regulate DDR genes itself or it may require other AR-V partners to do so. Development of antibodies which would discriminate the different AR-Vs by binding to their distinct cryptic exons would ultimately answer this question.

The CWR22Rv1-AR-EK cell line is the first CRISPR edited PC cell line that has been developed to date. More CRISPR edited cell line models are now emerging. Namely, parental CWR22Rv1

cells have been engineered to express an AR-V7-GFP fusion to precisely assess how AR-V7 expression levels alter in response to SF3B2 depletion (Kawamura et al., 2019).

Expansion of the AR-FL knock-out pipeline presented in this study to other PC cell lines such as the androgen-independent LNCaP95 cell line, which expresses endogenous AR-Vs will be useful in order to assess whether AR-Vs behave differently or similarly in different PC cell lines to ultimately establish their precise role in CRPC relevant backgrounds. This is currently an ongoing project in the lab.

Development of PC cell lines which harbour certain point mutations in the LBD of AR using CRISPR was of keen interest. Previously, lentiviral or adenoviral delivery of the gene of interest was the gold standard approach for generating cell lines that express the desired protein (O'Neill et al., 2015; Sun et al., 2006). However, lentiviral delivery is associated with constant overexpression of the gene of interest which can sometimes be toxic to the host cell line. Random integration of the lentiviral expression cassette is also possible to disrupt other genes and their function, interfering with phenotypic read-outs. CRISPR allows researchers to apply more physiological approaches to overcome such weaknesses by permanently engineering the locus of interest while maintaining expression of the gene of interest at physiological endogenous levels.

A CRISPR knock-in pipeline was developed in order to introduce the W741L point mutation in AR exon 5. *In silico* analysis revealed that this is a recurrent mutation observed in PC patients and position 741 comprises a mini hotspot within the AR genomic sequence (COSMIC, 2019).

CRISPR knock-in experiments comprise a challenging task which requires extensive optimisation of more than one steps throughout the process including cell transfection, rationale gRNA and donor template design as well as enforcing high endogenous HDR rates in the host cell line. The latter is the most unpredictable variable when one performs knock-in experiments. Incorporation of the desired point mutation, in our case this was the W741L mutation relies on the exchange of an exogenous DNA donor template which encompasses the W741L mutation with the homologous fragment of endogenous DNA, a phenomenon called homologous recombination or HDR (Ran et al., 2013). However, HDR does not occur frequently making this exchange extremely challenging.

This study focused on optimising the steps to maximise HDR in order to generate a CWR22Rv1 cell line derivative which would express the AR<sub>W741L</sub> mutant and ultimately understand its function and contribution to PC progression.

Previously low HDR rates were ultimately enhanced using a streptavidin-tagged Cas9 approach in conjugation with biotinylated donor templates (Ma et al., 2017). Applying this approach in parental CWR22Rv1 cells resulted in the generation of an AR<sub>W741L</sub> mutant cell line derivative. The engineered CWR22Rv1 cells (clone 11) harbour the H874Y mutation in addition to the knock-in W741L mutation in the *AR* locus. In order to discriminate the function of the two mutant receptors an additional round of CRISPR knock-in should be performed to correct the endogenous H874Y mutation to wild type.

More recent advances in the CRISPR field to improve knock-in rates include the fusion of Cas9 to certain HDR mediators such as Rad51 in order to guarantee their constant presence in close proximity to the cut site and immediate activation of HDR at the expense of the error-prone NHEJ which would lead to imprecise editing and most likely knock-out of the gRNA targeted gene (Rees et al., 2019). Finally, the emergence of base editors has revolutionised the application of CRISPR to achieve precise genome editing. Base editors offer the advantage of donor template free, seamless editing which relies on spontaneous deamination of adenine to inosine which is recognised as guanine by DNA polymerase. A catalytically impaired Cas9 is fused to an adenine deaminase which operates on one strand of target DNA to convert A residues to G residues. Repair mechanisms are then triggered to fix the mismatch in the next round of DNA replication. This is important as most mutations associated with human genetic diseases result from C-G to T-A transitions and hence adenosine base editors could be exploited to correct pathogenic mutations in the clinic (Gaudelli et al., 2017).

Divergent application of CRISPR to study the AR-V7 specific spliceosome was also in the scope of this study. Development of a novel strategy to identify the SFs recruited at a certain genomic locus was attempted using a catalytically impaired Cas9 against CE3. A series of proof-of-principle experiments were conducted to assess whether the strategy was worth to be applied, further optimised and expanded. The CWR22Rv1 cell line was used to investigate spliceosome assembly around CE3. Targeting of dCas9 at CE3 and subsequent MS analysis of the isolated protein complexes yielded a number of candidate SFs and splicing regulators which were enriched around CE3 for further investigation and validation. Importantly, SFs

previously shown to regulate AR-V7 mRNA levels such as Sam68 were identified, confirming that CRIME is a valid method for assessing spliceosome assembly.

Preliminary experiments were performed to validate each CRIME-MS protein hit. SRSF3 was identified as a genuine modulator of AR pre-mRNA splicing. Despite the observation that SRSF3 might be responsible for CE3 skipping from the mature AR mRNA transcript, additional experiments are required to establish its precise function in CRPC. RNA IP should be performed using an SRSF3 specific antibody followed by PCR amplification of AR-V7 mRNA using the immunoprecipitated RNA as template. Although this was attempted, there was no robust evidence of AR-V7 mRNA presence in association with SRSF3. Failure to detect AR-V7 mRNA could be attributed to the fact that formaldehyde fixation limited antibody performance in IP experiments. Additionally, spliceosome assembly involves the formation of various protein complexes associated with the AR-V7 mRNA. Hence, it is a matter of finding the right timing to capture the desired SF.

A CRISPR screen is currently underway. A custom gRNA library against all known SFs has been designed and synthesised in collaboration with Astra Zeneca. This will provide a broader insight into AR pre-mRNA maturation processes and could provide extra validation if read-outs overlap with CRIME hits. However, CRISPR induced knock-out of SFs may not necessarily indicate AR-V7 specific SFs. It is known that SFs can exert other functions associated with posttranslational processing of their interacting partners (Ratnadiwakara et al., 2018). This fits with the observation that a significant number of proteins involved in splicing are associated with AR-V7 protein (unpublished data).

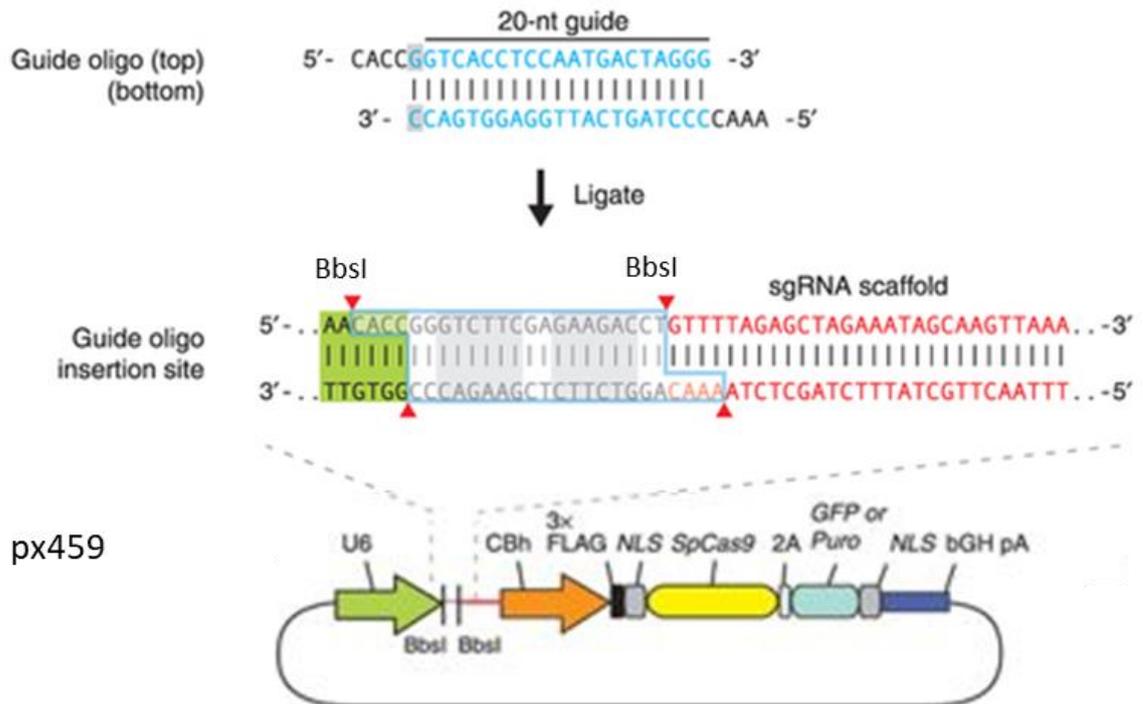
The application of CRIME to clinical specimens is certainly a promising provision to expand CRIME to more translationally relevant models. Difficulties such as maintaining *ex vivo* cultures must be overcome prior to its application to such models of disease. However, various strategies to establish viable *ex vivo* CRPC cultures are currently optimised in the lab.

In summary, this study generated novel CRISPR-engineered CWR22Rv1-derived cell line models to facilitate improved modelling of CRPC. CWR22Rv1-AR<sub>W741L</sub> cells permanently express the CRPC relevant AR<sub>W741L</sub> mutant were established and now require further validation so they can be used to elucidate the mechanisms of regulation of this particular mutant receptor. Generation of AR-FL knock-out CWR22Rv1 cells (CWR22Rv1-AR-EK)

provided unequivocal insight in AR-V function in a CRPC relevant cellular background devoid of interfering AR-FL activity. AR-Vs regulate DDR and can confer resistance to radiation therapy, contributing to PC progression. This observation highlighted new modes of AR-V regulation that suggest novel pharmacological sensitivities in AR-V positive CRPC patients. Outside of CRISPR editing, the technology was repurposed to study the AR-V7 spliceosome. A pioneering method called CRIME which couples a catalytically inactive Cas9 targeting at CE3 with MS was developed to identify the RNA-protein complexes assembled at CE3 during AR pre-mRNA splicing. Proof-of concept experiments were conducted to assess the validity of the method and highlighted SRSF3 as a potential AR-V7 splicer in CRPC models. These preliminary read-outs are currently validated and backed up in the lab using other genome-wide approaches, namely CRISPR knock-out screens.

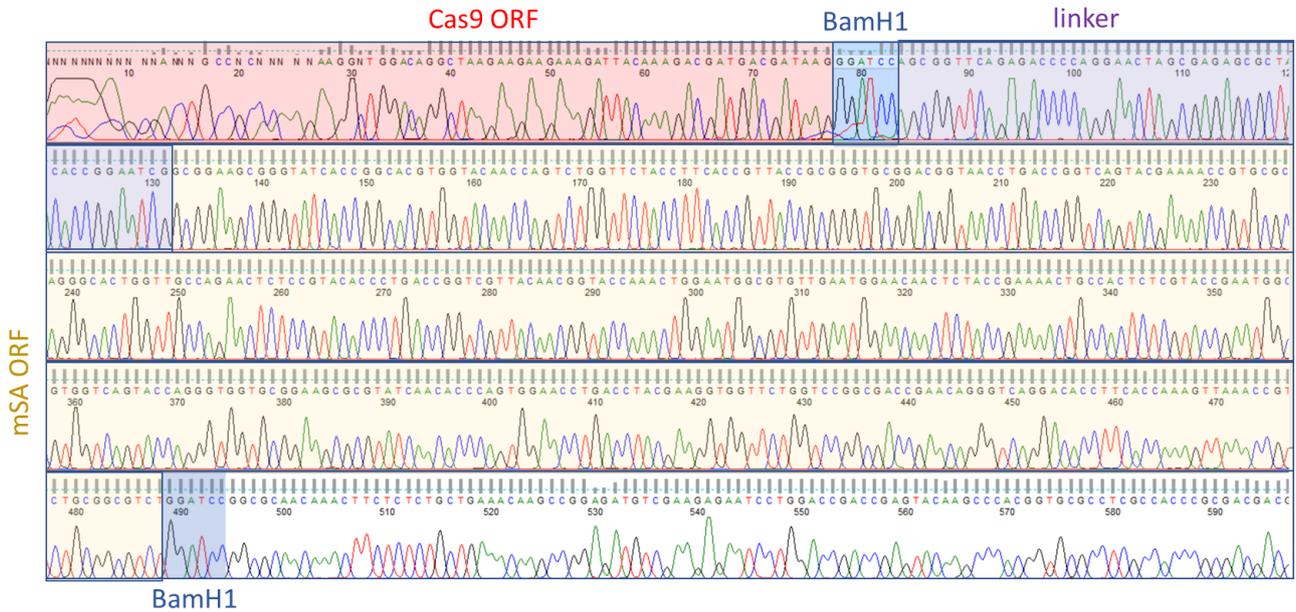
## Appendices

### Appendix A. Cloning of gRNA into the px459 vector.

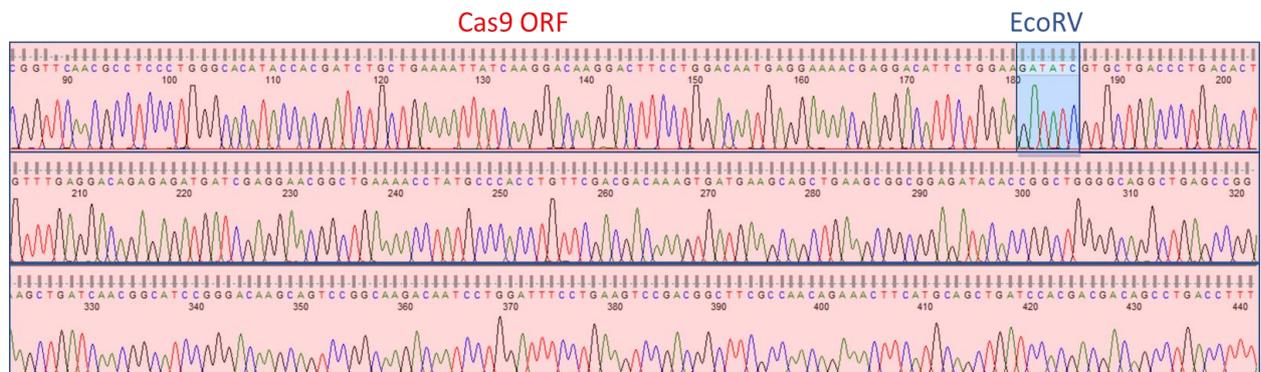


**Appendix B.** In-frame cloning of linker-mSA insert into lentiCRISPR V2.0 (intermediate vector). In-frame cloning of partial Cas9-linker-mSA into px459 (recipient vector) (top and middle panel). Validation of mSA cassette insertion into px459 by *BstEII* digestion (bottom panel).

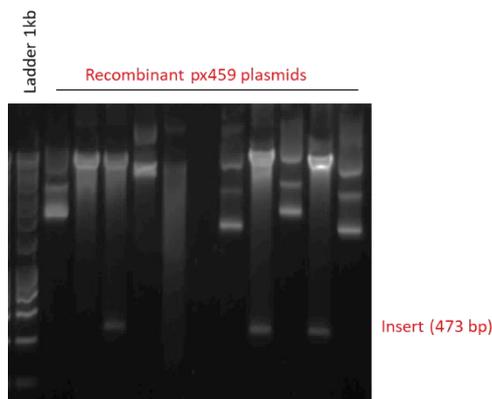
**A.**  
**Recombinant lentiCRISPR V2.0**



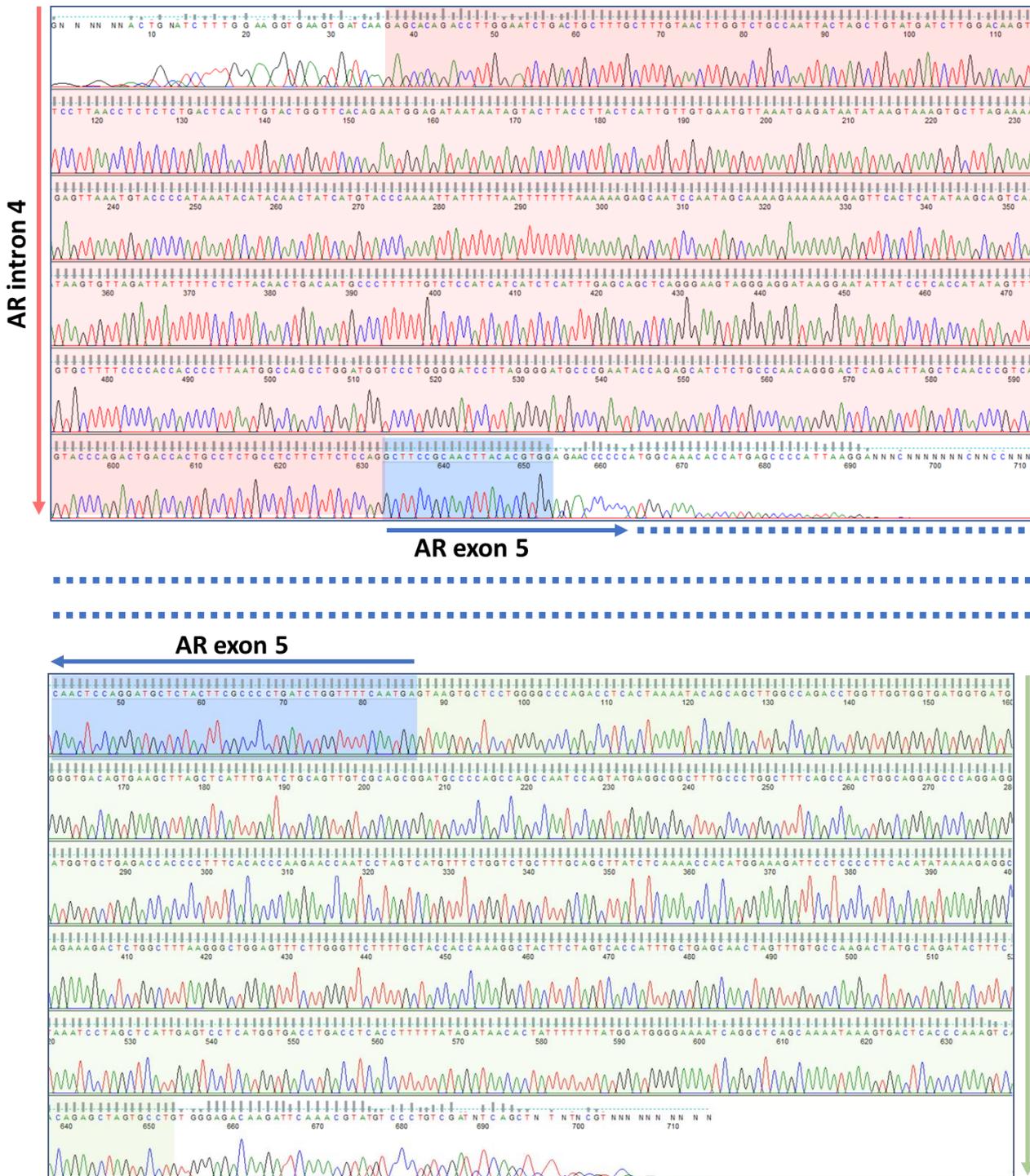
**B.**  
**Recombinant px459**



**C.**



**Appendix C. Representative examples of CWR22Rv1-AR-EK AR intron 4 and 5 sequencing chromatograms showing no off-target CRISPR induced mutations.**



**Appendix D.** List of the top 45 predicted gRNA\_2 off-target sites. Genes highlighted in red are not expressed in the prostate. #MM, number of mismatches

| location | sequence                 | # MM | gene         | gene name  |
|----------|--------------------------|------|--------------|--|
| chrX     | CTTACACGTGGACGACCAGATGG  | 0    | NM_001011645 | AR   |
| chr1     | TATTCACGTGGATGACCAGATAG  | 4    | None         |  |
| chr21    | TTTTACATGGATGACCAGATGG   | 4    | None         |  |
| chr15    | ATTTTCATGTGCACGACCAGACAG | 4    | None         |  |
| chr1     | CTTGCTCTTGGAGGACCAGACAG  | 4    | None         |  |
| chr10    | CTTAATCATGGAAGACCAGACAG  | 4    | None         |  |
| chr5     | ATTGCACGTGGGGGACCAGAGAG  | 4    | None         |  |
| chr10    | CCTACACGTGGGCGATCAGATGG  | 3    | None         |  |
| chr5     | CTGACAAGAGGACGACAAGAAGG  | 4    | None         |  |
| chr9     | CATACACTTGAAGACTAGATGG   | 4    | None         |  |
| chr7     | TTTAGACTTGGACAACCAGAAAG  | 4    | None         |  |
| chr3     | CTCACACTTGGATGACCACACAG  | 4    | None         |  |
| chr14    | CCTACATGTGGGAGACCAGAGGG  | 4    | NM_004131    | granzyme B (GZMB)  |
| chr11    | CCTCCACGTGGACGCCAGCTGG   | 4    | None         |  |
| chr4     | CATACAAGTGGATGACTAGAAAG  | 4    | None         |  |
| chr15    | ATAACATGTGGACCACCAGAAAG  | 4    | None         |  |
| chrY     | CTTCCACTTGGAGGCCAGAAAG   | 4    | None         |  |
| chr20    | CTTCCACGAGGAAGACGAGAAAG  | 4    | None         |  |
| chr5     | CTTGCAAGTGGAGGACCAAAAAG  | 4    | None         |  |
| chr17    | CTTACACTGGGAAGACAAGACAG  | 4    | None         |  |
| chr19    | CCTCCACGTGGATGACCCGAGAG  | 4    | NM_014727    | lysine methyltransferase 2B (KMT2B)                          |
| chr2     | CTTACCCTTGGAGGACCAATAG   | 4    | None         |  |
| chr8     | ATTACACTTGAAGAACAGAAAG   | 4    | None         |  |
| chr17    | CTTACACCTGGACACCCAGAAGG  | 3    | None         |  |
| chr14    | CTAACACCTGGACAACCAGGTGG  | 4    | None         |  |
| chr2     | CTCACACGTGTACGTCCAGCTGG  | 4    | None         |  |
| chr14    | CTCACAGGAGGACCACCAGAGGG  | 4    | NM_012244    | solute carrier family 7 member 8 (SLC7A8)                    |
| chr7     | CTTACACCTGGAGGTCCAGGAGG  | 4    | NM_000089    | collagen type I alpha 2 chain (COL1A2)                       |
| chrX     | CTAACTCGTGGAGGAGCAGAAGG  | 4    | None         |  |
| chr1     | CTACCACGTGGACCTCCAGAGAG  | 4    | None         |  |
| chr2     | CTTACACCTGGAAGGCAAGAAGG  | 4    | None         |  |
| chr8     | CTTCCACGTGCAGGAACAGAAGG  | 4    | None         |  |
| chr12    | CTTACACTTGGGCCACCAGGAAG  | 4    | None         |  |
| chr11    | CTTCCACGTGGAGGATCAGTCAG  | 4    | None         |  |
| chrX     | CTTCCACGTGGAAGAGCAGCAAG  | 4    | None         |  |
| chr21    | CTTACACGTTGACGGCTACACAG  | 4    | None         |  |
| chr5     | CTTACAAGGGGAAGAGCAGATGG  | 4    | None         |  |
| chr13    | CTTAAACTTGGACAATCAGATAG  | 4    | None         |  |
| chr2     | CTTACAAGTGAAAGAACAGAAAG  | 4    | None         |  |
| chr16    | CTTACATGTGGATGAGCAGGCAG  | 4    | None         |  |
| chr13    | CTTACACGAGGAGGATGAGAAGG  | 4    | None         |  |
| chr19    | GTTACACGTGGAGGAGCTGAGGG  | 4    | None         |  |
| chr13    | CTTACACGTGGTTGAACAGGCAG  | 4    | None         |  |
| chr13    | CTTACACGTGGTTGAACAGGCAG  | 4    | None         |  |
| chr5     | CTTACAGGTGGAATACCTGAGAG  | 4    | NR_046680    | ARHGAP26 antisense RNA 1 (ARHGAP26-AS1), long non-coding RNA |

**Appendix E. Up-regulated (green table) and down-regulated (red table) genes in response to AR-V depletion.**

|           |          |           |           |           |           |           |              |              |           |
|-----------|----------|-----------|-----------|-----------|-----------|-----------|--------------|--------------|-----------|
| AATK      | AKR1A1   | ABCD1     | ACE       | ACTL10    | ABHD14B   | ACBD4     | ABCA1        | APCDD1       | AHR       |
| ADAT3     | ATP7A    | AMOTL1    | ANKRD16   | ALX4      | ABTB1     | ACPP      | ADM          | ATP1B1       | ARHGAP32  |
| ATG2A     | BDH2     | AMPD2     | ASTN2     | ANKRD52   | ADCK5     | ADAM15    | ALOX15       | BAMBI        | BASP1     |
| BAHD1     | C4orf48  | ARAP2     | BAIAP2L2  | ANXA9     | AMIGO2    | ADAMTS1   | AMER3        | BCRP2        | C11orf95  |
| BID       | CA11     | ARNTL2    | BPNT1     | APPL2     | ANG       | AES       | ANXA1        | C19orf21     | C8orf4    |
| C17orf103 | CALCOCO1 | BICD2     | C1orf115  | ARSJ      | ASS1      | ARHGEF3   | AQP3         | C2orf15      | C9orf152  |
| C19orf24  | CAMK2N1  | BTBD9     | CDH7      | ATXN7L1   | BOK       | ATP2B1    | B3GNT3       | CCDC159      | CACNA1D   |
| C6orf132  | CBX6     | C20orf118 | CFD       | BSPRY     | CACFD1    | ATP8A1    | CDC42EP3     | CDH18        | CCDC64B   |
| C7orf43   | CLN8     | C4orf32   | CYP1A1    | CASP7     | CDKN1A    | C15orf57  | CNTN3        | CLDN11       | CCDC80    |
| CARD14    | CLTB     | CAMK1     | DBN1      | CDIPT     | CHRM3     | CCDC120   | CREB3L1      | CLDN4        | CDH12     |
| CARNS1    | CRB3     | CAMLG     | DDN       | CDK8      | CLSTN3    | CPEB3     | DSCAM-AS1    | CYP4F35P     | CHST2     |
| CBLC      | CREBL2   | CTC1      | DNAJC18   | CHFR      | CYB561D1  | CYSTM1    | ELF3         | DCUN1D3      | COL14A1   |
| CHRNE     | DNALI1   | CTXN1     | DSC2      | COL5A2    | DDHD1     | DUSP1     | ENPP4        | DUSP4        | CTTNBP2   |
| COBL      | DUSP2    | CUEDC1    | EFNA1     | DNASE1    | DUSP16    | DYRK1B    | FBLIM1       | FAM110C      | CYP1A2    |
| DBC1      | EEF1A2   | DACH1     | EML6      | DOPEY2    | EFNB2     | EFNA3     | FLJ38109     | FLJ22184     | DDIT4L    |
| DCAF12L2  | EID1     | DEGS2     | ERBB3     | ELMO3     | FAM102A   | EGR1      | GAB2         | FLJ23867     | DGKA      |
| DEAF1     | EPAS1    | DHRS3     | ERGIC1    | EPHB3     | GLYCTK    | FAM195B   | GATA2        | FOXO4        | ELF5      |
| DENND5A   | EPN3     | DOCK4     | ERO1L     | FAM109A   | GMIP      | FAM214B   | GPR3         | FOXP4        | ENPP5     |
| DLG5      | FGFR3    | ENTPD6    | FAM63A    | FAM111A   | HIST3H2A  | FLJ20021  | HABP4        | GDPD1        | EVX1      |
| ETV4      | FKRP     | ERBB4     | FAM86HP   | FGD3      | HOXC10    | GABARAPL1 | HGD          | GPR35        | FOS       |
| EVPL      | FOSL2    | FAM173A   | FNIP2     | FXYD3     | HSPG2     | HIST1H3H  | HID1         | GUSBP1       | GALNT3    |
| FAM162A   | G3BP2    | FERMT3    | GADD45G   | GRAMD1A   | IKZF2     | HOXA13    | HIST1H2AG    | HGF          | GSN       |
| FAM84A    | HEXIM1   | GPR160    | GSTT1     | HIST1H2AC | IL17RE    | IL17RC    | HIST1H2BK    | HIST1H1C     | HMGCS2    |
| FKBP8     | IFI35    | GYS1      | HES7      | HOXC13    | ITGA3     | IMMP2L    | KCND2        | HIST1H3E     | HOXC12    |
| FZD4      | IFT27    | KLC3      | ICA1      | IRAK2     | IZUMO4    | IRF7      | KIAA1467     | HS6ST2       | IER2      |
| GGT1      | KDM6B    | LOC646862 | ID1       | ITGA5     | JOSD2     | LAD1      | KIAA1522     | IL36RN       | KIAA1199  |
| GREB1L    | KIAA0513 | MIR600HG  | JMJD7     | KHNYN     | JUN       | LOC113230 | KRT8         | INPP5A       | KIAA1324  |
| HIST2H2BE | KLF4     | MTHFR     | JUP       | LOC284578 | KDM2A     | LOC338758 | LANCL3       | JUNB         | LCOR      |
| IGFBP3    | KLHL28   | MVP       | KIF9      | LXN       | KIAA0922  | MAFK      | LOC100862671 | KCNJ11       | LNX1      |
| LAMB2     | LDHD     | MYOF      | LATS2     | MAPK15    | LLGL2     | MAP1S     | LOC388692    | KCNJ3        | LOC286367 |
| LITAF     | LRFN4    | NCOA3     | LOC283335 | MEX3D     | LOC729737 | MAPRE2    | LRP10        | KIF13B       | MAFF      |
| LPPR2     | LYPLA2   | NGEF      | LRSAM1    | MICAL1    | MAP1LC3A  | MNT       | MAPK13       | KLHL1        | MAPK4     |
| MROH6     | LZTR1    | NRP1      | MAN2A2    | MXRA8     | MMP24     | MTSS1L    | MST4         | LHX9         | MFSD4     |
| MXD4      | MANSC1   | OSGIN1    | MAN2B1    | MYH14     | NACC2     | MYRIP     | MUC1         | LIMA1        | NCMAP     |
| MZF1      | MAPKAPK3 | PAN3      | MFSD6     | NKAIN1    | NFKBIZ    | NCAM2     | MYO7A        | LIPH         | NOTCH3    |
| NADK      | MARK1    | PANX2     | NCK2      | NOXA1     | PLEKHA2   | NR1H2     | PCED1B       | LOC100128770 | NOV       |
| NUDT22    | MC1R     | PCDHA10   | PCBP4     | NPDC1     | POMGNT1   | PCDHA4    | PEG10        | LRRCS56      | OGFR      |
| PLCD3     | MFSD10   | PERP      | PSMB10    | NUDT14    | PRRG2     | PEX11A    | PEL1         | MAL2         | OSBPL5    |
| PLXNA1    | MPZL3    | PKP2      | PTPRK     | PCDH19    | RAB30     | PLEKHA7   | PHLDB3       | MESDC2       | PAN3-AS1  |

|           |         |          |          |          |          |           |         |           |          |
|-----------|---------|----------|----------|----------|----------|-----------|---------|-----------|----------|
| PODXL2    | NAP1L3  | PLEKHA6  | RBPM5    | PCDH9    | RUNDC3B  | PLEKHB1   | PIK3AP1 | MESP1     | PART1    |
| POLD4     | NAPRT1  | PNPLA6   | REEP2    | PCNXL3   | SDC4     | PLXNA2    | PLXNB3  | NKPD1     | PLA2G4A  |
| PRAC      | NDUFB4  | PPP2R2A  | RHOC     | PDE4B    | SDA3C    | PTGS2     | PRPH    | PBXIP1    | PLA2G4F  |
| PRRT3     | PDF     | PRMT6    | S100P    | PDLIM2   | SGSM3    | RAB17     | PTPRCAP | PLA2G16   | PPFIA2   |
| PVT1      | PHF12   | PROM2    | SERTAD1  | PSD3     | SH3GLB2  | RENBP     | PTRH1   | PNCK      | RALGAPA2 |
| QDPR      | PITPNM1 | RAB24    | SGSM1    | PTGFR    | SIX2     | RHPN1     | RALYL   | PNPLA7    | REG4     |
| RAB11FIP4 | PPP2R5A | RNF11    | SH2B1    | PVRL4    | SLC26A11 | SATB1     | RDH5    | PPP1R3B   | SCUBE2   |
| RASEF     | PTPN21  | SAMD10   | SH3BGRL2 | RAB3B    | SLC7A8   | SH3RF1    | RELB    | PTHLH     | SLC5A8   |
| RFX3      | RILP    | SAT1     | SIDT2    | RBMS2    | SLIT1    | SLC30A10  | RNF208  | QSOX1     | SPEF1    |
| ROBO1     | SLC29A4 | SDA3F    | SLC12A6  | RIT1     | SOX9     | SLC40A1   | SGK2    | RAB25     | STAP2    |
| RSPH1     | SLC41A2 | SIGIRR   | SLC25A29 | RNPEPL1  | SPHK2    | SPRYD3    | SI      | RBM11     | SULT2B1  |
| S100A11   | SLC6A6  | SLC52A3  | SLC39A13 | RWDD2A   | STK40    | SSH3      | SLC17A5 | RNF223    | SYT4     |
| SCAND1    | SOWAHB  | SLC5A6   | SP110    | SELM     | STX12    | SYT7      | SLC43A2 | SLCO5A1   | SYTL2    |
| SHC4      | SYTL1   | STARD10  | STOM     | SNX32    | SYNGR2   | TLL1      | SLITRK5 | SMPDL3B   | TLE1     |
| SLC48A1   | TEP1    | STBD1    | STXBP5   | SPRY1    | TBX2     | TM7SF2    | SSTR1   | TMEM2     | TMEM45B  |
| TMEM135   | TET3    | SYNJ2BP  | TJP2     | THBS1    | TJP3     | TNFRSF12A | STON1   | TNFRSF11B | TNFRSF19 |
| TNFRSF21  | TINAGL1 | TMEM184B | TMC6     | TMEM79   | TPM4     | TNFSF9    | TMC4    | TP53INP1  | TSPAN1   |
| TRPV3     | TMEM8A  | TPD52L1  | TMEM238  | TNK2     | TSTD1    | TRPM4     | TMEM125 | TSPAN12   | ZDHC16   |
| TYRO3     | WWP2    | UNC13B   | TRADD    | TP53INP2 | VASN     | TRPS1     | TNFSF15 | TSPAN15   | ZNF385A  |
| WNT9A     | XKR8    | WDR45    | TTBK2    | TULP4    | ZNF385B  | TTC39A    | TRIB1   | TUFT1     |          |
| ZBTB7A    | ZDHC18  | ZG16B    | UNC45A   | ZFP36    | ZNF524   | ULK1      | ZCCHC3  | ULK3      |          |
| ZFP36L2   | ZNF341  | ZSCAN16  | VGLL4    | ZFYVE21  | ZNF827   | ZFP36L1   | ZSWIM4  | ZCCHC24   |          |

|           |              |          |           |           |           |              |           |         |           |
|-----------|--------------|----------|-----------|-----------|-----------|--------------|-----------|---------|-----------|
| AAED1     | ACADL        | ACBD7    | ADORA1    | AAK1      | AADAT     | AFAP1L1      | ABHD15    | ALDH1L2 | ARL1      |
| ACOT7     | AP1AR        | ADSSL1   | AGFG2     | AKAP5     | ACSF2     | ANKIB1       | ALDH6A1   | ALG12   | BMP8B     |
| ACTR3     | APOD         | ALCAM    | AMBRA1    | ARHGAP11B | ADAMTS3   | BLZF1        | ALG10     | ALKBH8  | C5orf22   |
| ALDH3A2   | ARHGEF10     | AP1S3    | ANKRD32   | ATP12A    | AKAP12    | C10orf118    | ARHGEF19  | BOD1    | C9orf37   |
| ANGPTL4   | ARHGEF37     | APLN     | APOLD1    | BMP6      | AP3M2     | C14orf101    | ARSB      | BRIX1   | CBLL1     |
| AR        | ARHGEF6      | ATCAY    | ATOH8     | BRCA2     | ARHGAP11A | CCDC71L      | B3GALTL   | C2CD5   | CCDC66    |
| ARMCX4    | ARL4C        | ATP6V0E1 | BCL2L11   | C4orf21   | ARHGAP19  | CCNB2        | BDP1      | CAMK2D  | CCDC90B   |
| AUNIP     | ATAD5        | BARX1    | BLM       | CDC6      | ARPC5     | CENPH        | BUB1B     | CBWD1   | CCNE1     |
| BRMS1L    | BARD1        | BRIP1    | BRCA1     | CDH23     | ATAD2     | CENPK        | C7orf60   | CCSER2  | CDKN1C    |
| C11orf92  | C11orf93     | C16orf55 | C17orf104 | CECR6     | ATG14     | CENPN        | CCNA2     | CDC20   | CDKN3     |
| CAST      | C1orf112     | C9orf40  | C19orf57  | CEP152    | AVPR1A    | CENPQ        | CCNB1     | CDK17   | CENPE     |
| CDC25C    | CCNE2        | CCDC177  | C1QL4     | COPS7B    | BORA      | CEP120       | CCNJ      | DSEL    | CENPF     |
| CHRNA2    | CCSAP        | CEP63    | C5        | CTSO      | C11orf82  | CNN3         | CDC40     | E2F1    | CLIC4     |
| CNOT6     | CDC42        | CLGN     | CASC5     | CUL4B     | C14orf37  | COTL1        | CDK19     | E2F7    | COPS6     |
| CSDA      | CENPI        | CLOCK    | CDC7      | DGCR5     | C18orf54  | CTR9         | CEP78     | EDN2    | COX15     |
| CSDAP1    | CIT          | CORO1A   | CDCA8     | DHRS2     | C1orf21   | DCTN5        | CKAP2     | FADS2   | CPS1      |
| DLX1      | CLDND1       | CTSL2    | CENPJ     | EPB41     | C7orf63   | DONSON       | CKAP2L    | FAM101B | CSRN2     |
| DMD       | CLSPN        | DAPL1    | CETN3     | EPHB2     | CASP8AP2  | DPYSL5       | CLIP1     | FBLN1   | DHFR      |
| DOCK8     | CNKS2        | DNA2     | CNTFR     | ERCC6L    | CCDC18    | FLVCR1-AS1   | DEK       | FBXO5   | DIAPH3    |
| EFEMP1    | CROT         | DNMT3A   | DMC1      | EXO1      | CDCA7     | FOXN2        | DMXL2     | FEN1    | DNAJC27   |
| EFHD1     | CRYM         | ELL2     | DTD1      | FAM120C   | CDK1      | FZD2         | EIF5A2    | FUT10   | DTX4      |
| FKBP5     | DCAF12       | ESR1     | EBAG9     | FAM72B    | CLIP2     | GEN1         | ESCO2     | GABPB2  | DUS4L     |
| FOXD2-AS1 | DEPDC1       | EXOSC1   | EME1      | G2E3      | CMC2      | GIN54        | EVA1C     | GIN53   | DYM       |
| FSTL1     | DUSP3        | FAM57B   | ERI2      | GADD45B   | CRYBG3    | GM2A         | FAM136A   | GOT1    | EIF2AK3   |
| GHRHR     | FAM104B      | FAM72D   | FANCI     | GNA13     | DGUOK     | GNG4         | GFPT2     | GSTM3   | EIF4EBP1  |
| GMPR      | FAM72A       | FMO4     | FKBP9     | GNAI2     | DYNLT1    | HIPK1        | HAUS3     | HAUS6   | FANCD2    |
| GNPDA1    | FBXW11       | GIPC3    | FMNL3     | GPLD1     | E2F2      | HK2          | KATNBL1   | HJURP   | HECTD1    |
| GRIN3A    | GABARAP      | GLT25D2  | GPR137C   | GPM6A     | EGFR      | HSPB11       | KCTD9     | HLTF    | HIVEP2    |
| HPGD      | GHR          | HOOK1    | H19       | GSG2      | ETNK2     | ITPRIP       | KIAA0232  | HMGB2   | HNRNPH3   |
| HSPB6     | GPC4         | HOPX     | HAPLN3    | HAS3      | FADS1     | KDM4A        | KIAA1731  | HOMER2  | IER5L     |
| IL1R1     | HOXA4        | KIF15    | ICAM3     | HELLS     | FAM120AOS | LOC100144603 | KIF18B    | IDH1    | KIF18A    |
| INMT      | IGF1         | KNSTRN   | ISCA1     | HMMR      | FAM213B   | LOC148709    | KIF5C     | LIG1    | KIF23     |
| KCNC4     | KCNMB4       | KNTC1    | ITGB3BP   | KIF14     | FAM222B   | LRCH1        | LDLRAD3   | LMNB1   | KIF5A     |
| KDEL2     | KIF11        | KRT19    | KCTD12    | KIF20B    | FANCM     | LRP11        | LOC284889 | LPAR3   | LMAN1     |
| KLHL42    | KLK3         | LRRC40   | KIAA1524  | KIF24     | FAS       | MKI67        | MAP4K2    | MAFB    | LOC730101 |
| LAMA3     | LINC00467    | MCAM     | KIF21B    | KLF11     | FGFR1     | MPHOSPH6     | MCFD2     | MAGOHB  | MANF      |
| LIN9      | LOC100499405 | MCM8     | KIF26A    | KLF9      | FOXD4     | MPHOSPH9     | MCM6      | MBOAT2  | MCM4      |
| LPGAT1    | MAP2K6       | METTL7A  | LCLAT1    | KLK2      | GIN51     | MT1E         | MELK      | METTL4  | MDM1      |
| MAOA      | MGME1        | MMS22L   | LPCAT4    | LIN52     | GPAM      | NDRG4        | NAE1      | MMD     | MIPOL1    |
| MAP1B     | MIS18A       | MTBP     | LPL       | LOC389831 | HSPA5     | NEIL3        | NDC80     | MOB1A   | MKL2      |
| MRPL11    | MT1X         | NEDD4L   | LRRCC1    | LOC645249 | INPP5B    | NUP35        | NIF3L1    | MSH2    | MT2A      |
| MYO1D     | MYBL1        | NUP54    | MAD2L1    | LOC81691  | IRS2      | OSBPL3       | NUCKS1    | MTFR2   | NEDD1     |
| NCAPG2    | NAMPT        | NXPE3    | MARCKS    | MLF1IP    | LMNB2     | PKIB         | PBRM1     | MTHFD2  | NES       |

|          |         |          |          |            |              |           |          |         |          |
|----------|---------|----------|----------|------------|--------------|-----------|----------|---------|----------|
| NEMF     | NFATC3  | PBK      | MASTL    | MT1G       | LOC100288637 | POLA1     | PGBD5    | NUP155  | NUP107   |
| NKX3-1   | NID1    | PDCD4    | MCM10    | MYLK       | LRR1         | POLA2     | PIGK     | ORC1    | OAS3     |
| NME4     | OTOP3   | PFKFB3   | MCM3     | NEURL      | LRR3         | PREP      | PNPLA4   | ORC3    | OPA1     |
| NPTX1    | PALLD   | PHF16    | MFAP3L   | NR2C1      | MAML1        | PTPLB     | PPP6R3   | PARP2   | PAN2     |
| NPTX2    | PREX2   | PMP22    | MIS18BP1 | NUF2       | NCAPG        | RHNO1     | PRR14    | PCTP    | PDS5B    |
| NSA2     | PRIMA1  | PNMA2    | MND1     | ONECUT2    | NUP43        | RTTN      | PTENP1   | PDP1    | PLK4     |
| NTAN1    | PSMG2   | PNRC2    | MTFR1L   | OSGEPL1    | OGDHL        | SENP1     | RWDD2B   | PI4K2A  | PPP1R3E  |
| PCNA     | PTMA    | POLE2    | PDK4     | PABPC4     | OIP5-AS1     | SFSWAP    | SDA4A    | POGK    | PRIM2    |
| PCNP     | PTPRB   | PPWD1    | PIM1     | PARPBP     | OSTC         | SH3PXD2B  | SFN      | PPIP5K2 | PRPS2    |
| PIGP     | RAD54B  | PSIMCT-1 | PLA2G7   | PCDHA11    | PLEKHG2      | SLC25A24  | SLC22A31 | PSMD5   | RNASE4   |
| PLXND1   | RB1     | PTCH1    | PLEKHF1  | PLOD2      | RAD51        | SMC1A     | SLC25A40 | PTK2B   | SLC10A3  |
| PMEP1    | RIMS3   | PTPDC1   | POLQ     | PSIP1      | RFC4         | SMC2      | SLC35A3  | RHOBTB2 | SMCHD1   |
| PPAP2A   | RRM2    | PTPRA    | PPDPF    | PTDSS1     | RFC5         | SMNDC1    | SMCR7    | RMI2    | SNHG1    |
| PRICKLE2 | SGOL1   | RAD54L   | PRIM1    | RAB11B-AS1 | RHOQ         | SRD5A3    | SNRK     | RPP30   | SPC25    |
| RGS2     | SLC47A1 | RNF10    | PSTPIP2  | RAD1       | RIBC2        | STK39     | SPAG5    | RTN4RL1 | STAM     |
| RRM1     | SMC1B   | RTKN2    | REEP3    | RAD51AP1   | RMI1         | SUZ12     | SPDL1    | SASS6   | STT3B    |
| SCYL2    | SMPDL3A | SBF2-AS1 | REEP4    | RAD51B     | SGOL2        | TAPT1     | ST7L     | SCN8A   | TAF5     |
| SGMS2    | SRRM4   | SEC24A   | REEP6    | RFC3       | SKA3         | TEAD1     | TCF7     | SDC2    | THSD4    |
| SLC17A7  | SS18    | SFXN1    | SCG3     | RPP14      | SLC31A2      | TMEM116   | TFAP4    | SLC38A3 | TUBE1    |
| SLC38A4  | STEAP2  | SLC2A12  | SEC61B   | RPRD1B     | SMIM13       | TNFAIP8L1 | TMEM209  | SMC6    | UBTD2    |
| SLC47A2  | STRIP2  | SRPK1    | SIGMAR1  | RPS6KB1    | SPSB1        | TRIM35    | TMEM47   | STYK1   | YBX2     |
| SLCO2A1  | SUDS3   | STIL     | SLC25A33 | SIMC1      | SYNGR3       | TSKU      | TP53TG1  | TMEM48  | ZDHHC8P1 |
| SNAI2    | TARP    | TBC1D1   | SSR1     | SMC4       | TOP2A        | TUSC2     | TRANK1   | TTC28   | ZNF185   |
| ST8SIA6  | TEX15   | TDH      | TBPL1    | SYT12      | UCHL3        | TYMS      | TRIP13   | UBE2C   | ZNF215   |
| STEAP1   | TMCC3   | TICRR    | TMEM14B  | THAP6      | USP31        | WDR17     | UBE2T    | USO1    | ZNF778   |
| STEAP1B  | TMED7   | TRMT112  | TMEM194A | TMEM14C    | WDR67        | ZBTB1     | UHRF1    | USP28   | ZWILCH   |
| TMPRSS2  | TMEM123 | TTK      | TMEM64   | TMPO       | WDR76        | ZBTB10    | USP7     | WWTR1   |          |
| TMX4     | TMEM143 | TXNDC16  | TMOD2    | UBE2E3     | ZAK          | ZMYM1     | XPOT     | ZNF273  |          |
| TNC      | TMEM60  | VASH2    | TMTC1    | WDHD1      | ZMIZ1        | ZNF114    | XRCC2    | ZNF326  |          |
| UNC5B    | TSC22D3 | WDR92    | TP73     | YTHDF3     | ZNF217       | ZNF318    | ZDHHC5   | ZNF473  |          |
| ZBTB16   | ZIC5    | ZBTB41   | TTL      | ZBTB8OS    | ZNF704       | ZNF519    | ZNF280C  | ZNF829  |          |
| ZNF789   | ZNF718  | ZNF492   | ZNF271   | ZNF480     | ZRANB3       | ZNF596    | ZNF530   | ZWINT   |          |

**Appendix F.** List of CRIME derived proteins in complex with dCas9 at AR CE3. Ensembl identifiers are provided.

| Identifier      | Protein name/Description   |
|-----------------|--|
| ENSP0000009041  | STARD3 N-terminal like   |
| ENSP00000215754 | macrophage migration inhibitory factor                                       |
| ENSP00000222482 | carboxypeptidase A4 (CPA4)   |
| ENSP00000222482 | carboxypeptidase A4  |
| ENSP00000229390 | serine and arginine rich splicing factor 9 (SRSF9)                           |
| ENSP00000231512 | chromosome 5 open reading frame 15   |
| ENSP00000233813 | insulin like growth factor binding protein 5                                 |
| ENSP00000235347 | PRAME family member 10   |
| ENSP00000241891 | opsin 4  |
| ENSP00000243077 | LDL receptor related protein 1 (LRP1)  |
| ENSP00000246151 | PITH domain containing 1   |
| ENSP00000253332 | A-kinase anchoring protein 12  |
| ENSP00000254942 | telomeric repeat binding factor 2  |
| ENSP00000258829 | NK2 homeobox 8   |
| ENSP00000261247 | JNK1/MAPK8 associated membrane protein                                       |
| ENSP00000261917 | hyperpolarization activated cyclic nucleotide gated potassium channel 4      |
| ENSP00000262633 | RNA binding motif protein 42 (RBM42)   |
| ENSP00000262982 | chromosome segregation 1 like  |
| ENSP00000264233 | DNA polymerase theta   |
| ENSP00000264434 | chromosome 2 open reading frame 42   |
| ENSP00000270233 | basal cell adhesion molecule (Lutheran blood group)                          |
| ENSP00000272102 | ADP ribosylation factor 1  |
| ENSP00000276204 | dedicator of cytokinesis 11  |
| ENSP00000283006 | centromere protein H   |
| ENSP00000284110 | heparan sulfate-glucosamine 3-sulfotransferase 3A1                           |
| ENSP00000285393 | ATPase H <sup>+</sup> transporting V0 subunit d2                             |
| ENSP00000288774 | peroxisomal biogenesis factor 10   |
| ENSP00000298283 | ribosomal protein L10 like   |
| ENSP00000302222 | zinc finger protein 25   |
| ENSP00000302896 | ribosomal protein S9   |
| ENSP00000313050 | nipsnap homolog 2  |
| ENSP00000313420 | protein kinase, DNA-activated, catalytic polypeptide (DNA-PKcs)              |
| ENSP00000313829 | KH RNA binding domain containing, signal transduction associated 1 (KHDRBS1) |
| ENSP00000318115 | translocase of inner mitochondrial membrane 50                               |
| ENSP00000318136 | intraflagellar transport protein 20 homolog                                  |
| ENSP00000321997 | La ribonucleoprotein domain family member 1B                                 |
| ENSP00000322250 | interleukin 17 receptor D  |
| ENSP00000323508 | adhesion G protein-coupled receptor A2                                       |
| ENSP00000324287 | ArfGAP with coiled-coil, ankyrin repeat and PH domains 2                     |
| ENSP00000325421 | presenilin associated rhomboid like  |
| ENSP00000334319 | piccolo presynaptic cytomatrix protein                                       |
| ENSP00000335158 | G protein-coupled receptor 142   |

|                 |  |
|-----------------|--|
| ENSP00000336799 | tubulin alpha 1b   |
| ENSP00000339889 | KIAA0368   |
| ENSP00000342216 | ATP binding cassette subfamily A member 9                |
| ENSP00000344668 | KRIT1, ankyrin repeat containing                         |
| ENSP00000345060 | GPR158-like  |
| ENSP00000346015 | ribosomal protein L27a                                   |
| ENSP00000348536 | ectonucleoside triphosphate diphosphohydrolase 4         |
| ENSP00000349960 | actin beta   |
| ENSP00000350583 | dipeptidyl peptidase 9                                   |
| ENSP00000351937 | chromosome 17 open reading frame 80                      |
| ENSP00000352085 | WD repeat and HMG-box DNA-binding protein 1              |
| ENSP00000353770 | ribonucleotide reductase regulatory subunit M2           |
| ENSP00000354483 | dual specificity phosphatase 27, atypical                |
| ENSP00000354739 | ribosomal protein L12                                    |
| ENSP00000355180 | collagen type VI alpha 1 chain                           |
| ENSP00000355886 | mitogen-activated protein kinase 4                       |
| ENSP00000356475 | SNF2 histone linker PHD RING helicase                    |
| ENSP00000358158 | Histone H2A  |
| ENSP00000358413 | zinc finger DHHC-type containing 6 (ZDHHC6)              |
| ENSP00000360907 | spermatogenesis associated 6                             |
| ENSP00000362298 | sphingosine-1-phosphate lyase 1                          |
| ENSP00000362330 | chromodomain helicase DNA binding protein 6              |
| ENSP00000362820 | serine and arginine rich splicing factor 3               |
| ENSP00000363489 | growth differentiation factor 5                          |
| ENSP00000363676 | ribosomal protein L11                                    |
| ENSP00000363714 | transmembrane protein 245                                |
| ENSP00000363826 | frizzled class receptor 8                                |
| ENSP00000363926 | par-3 family cell polarity regulator                     |
| ENSP00000364184 | kinesin family member 17                                 |
| ENSP00000365757 | phosphatidylinositol-5-phosphate 4-kinase type 2 alpha   |
| ENSP00000365908 | prune homolog 2  |
| ENSP00000367265 | cytoskeleton associated protein 4                        |
| ENSP00000367744 | phospholipase C, eta 2                                   |
| ENSP00000370912 | tec protein tyrosine kinase                              |
| ENSP00000371729 | sacsin molecular chaperone                               |
| ENSP00000372320 | neuroligin 4, Y-linked                                   |
| ENSP00000373191 | leukemia NUP98 fusion partner 1                          |
| ENSP00000373370 | filaggrin family member 2                                |
| ENSP00000373964 | smoothelin like 2  |
| ENSP00000375899 | SP140 nuclear body protein                               |
| ENSP00000376669 | Histone H4   |
| ENSP00000377640 | ribosomal protein L24                                    |
| ENSP00000377795 | mediator complex subunit 20                              |
| ENSP00000380415 | interleukin-2 receptor subunit alpha isoform 3 precursor |
| ENSP00000381698 | SHANK associated RH domain interactor                    |
| ENSP00000382239 | RNA binding motif protein 12B                            |

|                 |   |
|-----------------|---|
| ENSP00000382957 | zinc finger protein 705D                                  |
| ENSP00000385319 | HIC ZBTB transcriptional repressor 2                      |
| ENSP00000385332 | BSCL2, seipin lipid droplet biogenesis associated         |
| ENSP00000386336 | spermatogenesis associated serine rich 2 like             |
| ENSP00000387244 | FAM183A   |
| ENSP00000391800 | apelin  |
| ENSP00000392128 | maestro heat like repeat family member 2A                 |
| ENSP00000392541 | albumin   |
| ENSP00000393173 | putative RNA-binding protein Luc7-like 2 isoform 3        |
| ENSP00000395004 | leukocyte immunoglobulin like receptor B1                 |
| ENSP00000397552 | actin like 6A   |
| ENSP00000401246 | SNF related kinase  |
| ENSP00000402140 | nitric oxide synthase trafficking                         |
| ENSP00000403067 | FERM domain containing 5                                  |
| ENSP00000405975 | peptidylprolyl isomerase A                                |
| ENSP00000406229 | FAT atypical cadherin 1                                   |
| ENSP00000406293 | transcription elongation factor A3                        |
| ENSP00000415464 | La ribonucleoprotein domain family member 4               |
| ENSP00000415901 | acetylcholinesterase (Cartwright blood group)             |
| ENSP00000416786 | growth factor receptor bound protein 14                   |
| ENSP00000417764 | ALG2, alpha-1,3/1,6-mannosyltransferase                   |
| ENSP00000419389 | tRNA methyltransferase 10C, mitochondrial RNase P subunit |
| ENSP00000420817 | structural maintenance of chromosomes 4                   |
| ENSP00000424067 | unc-5 netrin receptor A                                   |
| ENSP00000427224 | GTF2H2 family member C                                    |
| ENSP00000427700 | DnaJ heat shock protein family (Hsp40) member C18         |
| ENSP00000428417 | regulating synaptic membrane exocytosis 1                 |
| ENSP00000430329 | minichromosome maintenance complex component 4            |
| ENSP00000434643 | ribosomal protein S3                                      |
| ENSP00000435169 | uncharacterized protein NKAPD1 isoform a                  |
| ENSP00000435389 | pleckstrin homology domain containing A7                  |
| ENSP00000435797 | anoctamin 1   |
| ENSP00000436849 | outer dense fiber protein 2-like isoform d                |
| ENSP00000445641 | caldesmon, CALD1  |
| ENSP00000450318 | CCR4-NOT transcription complex subunit 2                  |
| ENSP00000450697 | HECT domain E3 ubiquitin protein ligase 1                 |
| ENSP00000450806 | gamma-aminobutyric acid type A receptor alpha5 subunit    |
| ENSP00000452798 | importin 4  |
| ENSP00000453321 | Sperm equatorial segment protein 1                        |
| ENSP00000453357 | regulator of microtubule dynamics 3                       |
| ENSP00000455823 | microtubule-actin crosslinking factor 1                   |
| ENSP00000457552 | solute carrier family 12 member 3                         |
| ENSP00000457924 | ribosomal protein S15a                                    |

**Appendix G.** Pathways involved in cell cycle regulation are controlled by AR-Vs in CWR22Rv1-AR-EK cells.

| Biological process   | No. of genes | % of genes | Fold enrichment | P-value (Hypergeometric test) | Bonferroni method (corrected p-value) |
|--|--------------|------------|-----------------|-------------------------------|---------------------------------------|
| cell division  | 63           | 9.81       | 4.96            | 1.77393E-26                   | 2.10849E-22                           |
| DNA-dependent DNA replication  | 7            | 1.09       | 8.46            | 1.12367E-05                   | 0.133559243                           |
| DNA-dependent DNA replication initiation                                 | 17           | 2.65       | 13.12           | 1.05751E-15                   | 1.25696E-11                           |
| DNA replication  | 34           | 5.30       | 6.95            | 1.43037E-19                   | 1.70013E-15                           |
| mitotic cell cycle   | 20           | 3.12       | 4.56            | 1.35163E-08                   | 0.000160655                           |
| G1/S transition of mitotic cell cycle                                    | 21           | 3.27       | 5.61            | 1.0804E-10                    | 1.28416E-06                           |
| regulation of transcription involved in G1/S phase of mitotic cell cycle | 11           | 1.71       | 13.29           | 1.10301E-10                   | 1.31104E-06                           |
| G2/M transition of mitotic cell cycle                                    | 19           | 2.96       | 4.06            | 2.05028E-07                   | 0.002436959                           |

### Cell Division

KLHL42; NCAPG2; CDC25C; KIF11; RB1; CCNE2; MIS18A; CCSAP; CEP63; KNSTRN; KNTC1; MAD2L1; CETN3; ITGB3BP; CDCA8; REEP4; CDC7; MASTL; LRRCC1; MIS18BP1; REEP3; CENPJ; ERCC6L; NUF2; CDC6; KIF20B; HELLS; GNAI2; SMC4; KIF14; EPB41; DYNLT1; CDK1; NCAPG; BORA; SKA3; NUP43; CCNB2; SMC2; SMC1A; CCNA2; NDC80; HAUS3; SPAG5; KIF18B; BUB1B; CCNB1; SPDL1; UBE2C; FBXO5; HAUS6; TTC28; CDC20; LIG1; ZWINT; BOD1; SPC25; CENPE; ZWILCH; NEDD1; CCNE1; PDS5B; CENPF;

### DNA-dependent DNA replication

POLE2; POLQ; RFC3; WDHD1; RFC5; RFC4; POLA1;

### DNA-dependent DNA replication initiation

CCNE2; POLE2; MCM8; MCM10; CDC7; PRIM1; MCM3; CDC6; POLA2; GINS4; POLA1; MCM6; ORC3; ORC1; MCM4; PRIM2; CCNE1;

### **DNA replication**

RRM1; CDC25C; CLSPN; BARD1; RRM2; POLE2; MCM8; TICRR; DNA2; BRIP1; MCM10; BRCA1; CDC7; DTD1; MCM3; BLM; EXO1; RAD1; CDC6; RFC3; RFC5; RMI1; CDK1; RFC4; RHNO1; DONSON; POLA2; POLA1; MCM6; ORC3; FEN1; RMI2; ORC1; MCM4;

### **Mitotic cell cycle**

RRM1; KIF11; CIT; MYBL1; PBK; DNMT3A; KIF15; MASTL; CDC6; WDHD1; SKA3; NDC80; XRCC2; CLIP1; KIF18B; BUB1B; PBRM1; CENPE; TUBE1; CENPF;

### **G1/S transition of mitotic cell cycle**

RB1; CCNE2; POLE2; MCM8; MCM10; CDC7; PRIM1; MCM3; CDC6; RPS6KB1; CUL4B; POLA2; POLA1; MCM6; ORC3; ORC1; MCM4; EIF4EBP1; PRIM2; CCNE1; CDKN3;

### **Regulation of transcription involved in G1/S phase of mitotic cell cycle**

PCNA; RRM2; CDC6; KLF11; TYMS; POLA1; FBXO5; ORC1; E2F1; DHFR; CCNE1;

### **G2/M transition of mitotic cell cycle**

CDC25C; CIT; FBXW11; CEP63; MASTL; CENPJ; CEP152; CDK1; BORA; CCNB2; CCNA2; HAUS3; MELK; CEP78; CCNB1; HAUS6; NES; NEDD1; PLK4;

**Appendix H.** Pathways involved in DDR are controlled by AR-Vs in CWR22Rv1-AR-EK cells.

| Biological process  | No. of genes | % of genes | Fold enrichment | P-value (Hypergeometric test) | Bonferroni method (corrected p-value) |
|---|--------------|------------|-----------------|-------------------------------|---------------------------------------|
| DNA synthesis involved in DNA repair  | 15           | 2.34       | 11.91           | 3.13E-13                      | 3.73E-09                              |
| DNA repair  | 32           | 4.98       | 4.23            | 4.63E-12                      | 5.5E-08                               |
| DNA ligation involved in DNA repair   | 3            | 0.47       | 8.35            | 0.004611                      | 1                                     |
| DNA damage checkpoint   | 9            | 1.40       | 8.34            | 6.97E-07                      | 0.008287                              |
| DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest | 9            | 1.40       | 4.04            | 0.000354                      | 1                                     |
| double-strand break repair via homologous recombination                                       | 17           | 2.65       | 5.91            | 2.85E-09                      | 3.38E-05                              |

**DNA synthesis involved in DNA repair**

BARD1; DNA2; BRIP1; BRCA1; BLM; EXO1; BRCA2; RFC3; RAD51B; RAD51AP1; RMI1; RAD51; POLA1; XRCC2; RMI2;

**DNA repair**

CLSPN; TEX15; POLE2; RAD54L; TICRR; POLQ; BLM; EXO1; RAD1; PARPBP; RFC3; RAD51B; RAD51AP1; WDHD1; RFC5; CDK1; RFC4; ZRANB3; RAD51; ZBTB1; SMC1A; POLA1; XRCC2; UBE2T; UHRF1; FEN1; PARP2; USP28; MSH2; LIG1; FANCD2; PDS5B;

**DNA ligation involved in DNA repair**

HMGB2; PARP2; LIG1;

**DNA damage checkpoint**

CLSPN; CEP63; CLOCK; BRIP1; RAD1; RHNO1; DONSON; USP28; E2F1;

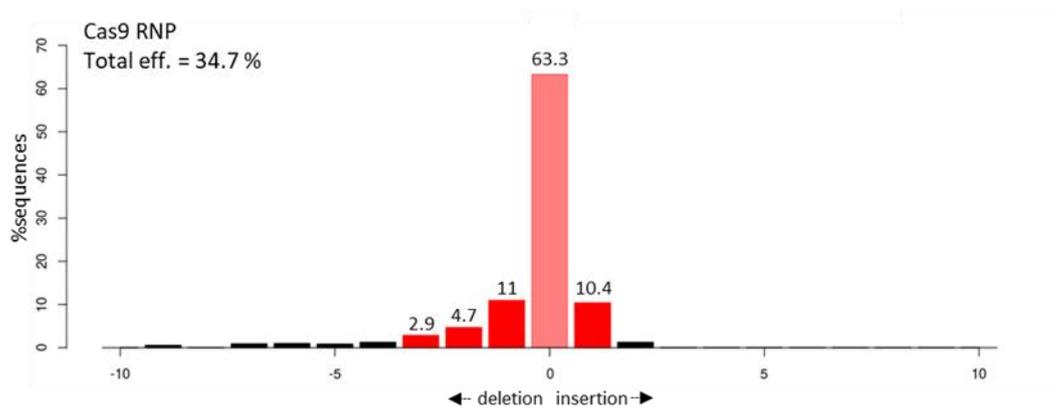
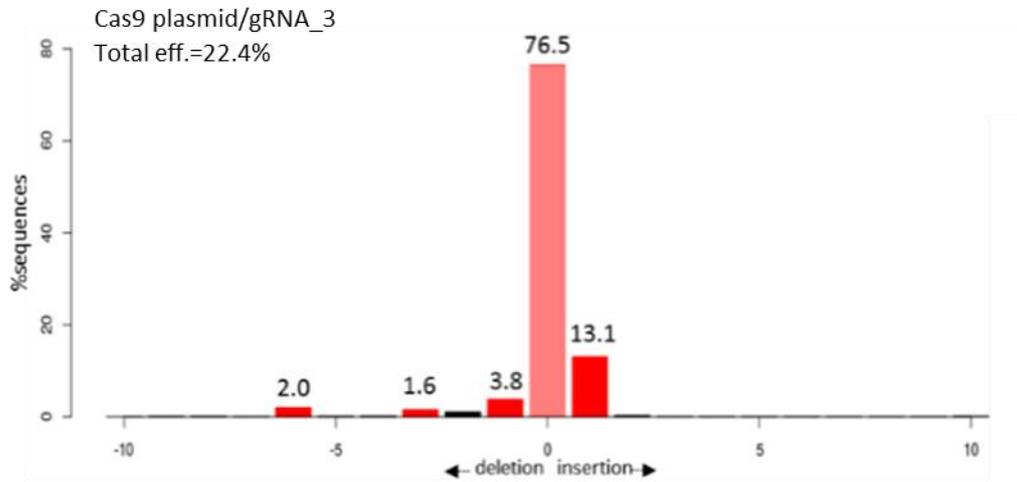
**DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest**

PCNA; CNOT6; CDC25C; CENPJ; CDK1; SFN; CCNB1; E2F7; E2F1;

### **Double-strand break repair via homologous recombination**

AUNIP; RAD54B; MCM8; RAD54L; MMS22L; POLQ; BRCA1; BLM; BRCA2; RAD51B; RAD51AP1;  
RAD51; GEN1; NUCKS1; XRCC2; FEN1; SMC6;

**Appendix I.** Performance of Cas9 RNP versus plasmid-derived Cas9 in CWR22Rv1 cells.



## References

- Abate-Shen, C., 2000. Molecular genetics of prostate cancer. *Genes Dev.* 14, 2410–2434. <https://doi.org/10.1101/gad.819500>
- Ahmed, M., He, H.H., 2017. SgTiler: a fast method to design tiling sgRNAs for CRISPR/Cas9 mediated screening. *bioRxiv* 217166. <https://doi.org/10.1101/217166>
- Aird, E.J., Lovendahl, K.N., St. Martin, A., Harris, R.S., Gordon, W.R., 2018. Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. *Commun. Biol.* 1. <https://doi.org/10.1038/s42003-018-0054-2>
- Alibhai, S.M.H., Gogov, S., Allibhai, Z., 2006. Long-term side effects of androgen deprivation therapy in men with non-metastatic prostate cancer: A systematic literature review. *Crit. Rev. Oncol. Hematol.* <https://doi.org/10.1016/j.critrevonc.2006.06.006>
- Allen, F., Crepaldi, L., Alsinet, C., Strong, A.J., Kleshchevnikov, V., De Angeli, P., Palenikova, P., Kosicki, M., Bassett, A.R., Harding, H., Galanty, Y., Muñoz-Martínez, F., Metzakupian, E., Jackson, S.P., Parts, L., 2018. Mutations generated by repair of Cas9-induced double strand breaks are predictable from surrounding sequence. *BioRxiv*. <https://doi.org/10.1101/400341>
- Anczukow, O., Krainer, A.R., 2016. Splicing-factor alterations in cancers. *RNA*. <https://doi.org/10.1261/rna.057919.116>
- Anczuków, O., Rosenberg, A.Z., Akerman, M., Das, S., Zhan, L., Karni, R., Muthuswamy, S.K., Krainer, A.R., 2012. The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nat. Struct. Mol. Biol.* 19, 220–228. <https://doi.org/10.1038/nsmb.2207>
- Anderson, E.M., Haupt, A., Schiel, J.A., Chou, E., Machado, H.B., Strezoska, Ž., Lenger, S., McClelland, S., Birmingham, A., Vermeulen, A., Smith, A. van B., 2015. Systematic analysis of CRISPR-Cas9 mismatch tolerance reveals low levels of off-target activity. *J. Biotechnol.* 211, 56–65. <https://doi.org/10.1016/j.jbiotec.2015.06.427>
- 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., Eisenberger, M.A., Luo, J., 2014. AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer. *N. Engl. J. Med.* 371, 1028–1038. <https://doi.org/10.1056/NEJMoa1315815>
- Arora, L., Narula, A., 2017. Gene editing and crop improvement using CRISPR-cas9 system. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2017.01932>
- Arora, V.K., Schenkein, E., Murali, R., Subudhi, S.K., Wongvipat, J., Balbas, M.D., Shah, N., Cai, L., Efstathiou, E., Logothetis, C., Zheng, D., Sawyers, C.L., 2013. Glucocorticoid receptor confers resistance to antiandrogens by bypassing androgen receptor blockade. *Cell* 155, 1309–22. <https://doi.org/10.1016/j.cell.2013.11.012>
- Asim, M., Tarish, F., Zecchini, H.I., Sanjiv, K., Gelali, E., Massie, C.E., Baridi, A., Warren, A.Y., Zhao, W., Ogris, C., McDuffus, L.A., Mascalchi, P., Shaw, G., Dev, H., Wadhwa, K.,

- Wijnhoven, P., Forment, J. V., Lyons, S.R., Lynch, A.G., O'Neill, C., Zecchini, V.R., Rennie, P.S., Baniahmad, A., Tavaré, S., Mills, I.G., Galanty, Y., Crosetto, N., Schultz, N., Neal, D., Helleday, T., 2017. Synthetic lethality between androgen receptor signalling and the PARP pathway in prostate cancer. *Nat. Commun.* 8. <https://doi.org/10.1038/s41467-017-00393-y>
- Attard, G., Reid, A.H.M., Auchus, R.J., Hughes, B.A., Cassidy, A.M., Thompson, E., Oommen, N.B., Folkerd, E., Dowsett, M., Arlt, W., De Bono, J.S., 2012. Clinical and biochemical consequences of CYP17A1 inhibition with abiraterone given with and without exogenous glucocorticoids in castrate men with advanced prostate cancer. *J. Clin. Endocrinol. Metab.* 97, 507–516. <https://doi.org/10.1210/jc.2011-2189>
- Auchus, R.J., Yu, M.K., Nguyen, S., Mundle, S.D., 2014. Use of Prednisone With Abiraterone Acetate in Metastatic Castration-Resistant Prostate Cancer. *Oncologist* 19, 1231–1240. <https://doi.org/10.1634/theoncologist.2014-0167>
- Baca, S.C., Garraway, L.A., 2012. The genomic landscape of prostate cancer. *Front. Endocrinol. (Lausanne)*. 3. <https://doi.org/10.3389/fendo.2012.00069>
- Bai, S., Zhang, B.Y., Dong, Y., n.d. Impact of taxanes on androgen receptor signaling. *Asian J. Androl.* 21, 249–252. [https://doi.org/10.4103/aja.aja\\_37\\_18](https://doi.org/10.4103/aja.aja_37_18)
- Basila, M., Kelley, M.L., Smith, A. van B., 2017. Minimal 2'-O-methyl phosphorothioate linkage modification pattern of synthetic guide RNAs for increased stability and efficient CRISPR-Cas9 gene editing avoiding cellular toxicity. *PLoS One* 12, e0188593. <https://doi.org/10.1371/journal.pone.0188593>
- Baurén, G., Wieslander, L., 1994. Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. *Cell* 76, 183–92. [https://doi.org/10.1016/0092-8674\(94\)90182-1](https://doi.org/10.1016/0092-8674(94)90182-1)
- 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. *Eur. Urol.* 63, 920–6. <https://doi.org/10.1016/j.eururo.2012.08.053>
- Ben-Shlomo, Y., Evans, S., Patel, B., Anson, K., Muir, G., Persad, R., Metcalfe, C., Chingwundoh, F., PROCESS Study Group, 2009. Differences in the epidemiology and presentation of prostate cancer in Black and White men in England: lessons learnt from the process study. *BJU Int.* 103, 723–4. <https://doi.org/10.1111/j.1464-410X.2008.08114.x>
- Bentley, D.L., 2014. Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* <https://doi.org/10.1038/nrg3662>
- Berget, S.M., Moore, C., Sharp, P.A., 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. U. S. A.* 74, 3171–3175. <https://doi.org/10.1073/pnas.74.8.3171>
- Beyer, A.L., Osheim, Y.N., 1988. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* 2, 754–765. <https://doi.org/10.1101/gad.2.6.754>

- Bielli, P., Busà, R., Paronetto, M.P., Sette, C., 2011. The RNA-binding protein Sam68 is a multifunctional player in human cancer. *Endocr. Relat. Cancer*.  
<https://doi.org/10.1530/ERC-11-0041>
- Bill-Axelson, A., Holmberg, L., Ruutu, M., Garmo, H., Stark, J.R., Busch, C., Nordling, S., Häggman, M., Andersson, S.-O., Bratell, S., Spångberg, A., Palmgren, J., Steineck, G., Adami, H.-O., Johansson, J.-E., SPCG-4 Investigators, 2011. Radical prostatectomy versus watchful waiting in early prostate cancer. *N. Engl. J. Med.* 364, 1708–17.  
<https://doi.org/10.1056/NEJMoa1011967>
- Bin Moon, S., Lee, J.M., Kang, J.G., Lee, N.E., Ha, D.I., Kim, D.Y., Kim, S.H., Yoo, K., Kim, D., Ko, J.H., Kim, Y.S., 2018. Highly efficient genome editing by CRISPR-Cpf1 using CRISPR RNA with a uridylate-rich 3'-overhang. *Nat. Commun.* 9.  
<https://doi.org/10.1038/s41467-018-06129-w>
- Black, D.L., 2003. Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annu. Rev. Biochem.* 72, 291–336. <https://doi.org/10.1146/annurev.biochem.72.121801.161720>
- Boel, A., De Saffel, H., Steyaert, W., Callewaert, B., De Paepe, A., Coucke, P.J., Willaert, A., 2018. CRISPR/Cas9-mediated homology-directed repair by ssODNs in zebrafish induces complex mutational patterns resulting from genomic integration of repair-template fragments. *DMM Dis. Model. Mech.* 11. <https://doi.org/10.1242/dmm.035352>
- Bohl, C.E., Miller, D.D., Chen, J., Bell, C.E., Dalton, J.T., 2005. Structural basis for accommodation of nonsteroidal ligands in the androgen receptor. *J. Biol. Chem.* 280, 37747–54. <https://doi.org/10.1074/jbc.M507464200>
- Bolla, M., Van Tienhoven, G., Warde, P., Dubois, J.B., Mirimanoff, R.O., Storme, G., Bernier, J., Kuten, A., Sternberg, C., Billiet, I., Torecilla, J.L., Pfeffer, R., Cutajar, C.L., Van der Kwast, T., Collette, L., 2010. External irradiation with or without long-term androgen suppression for prostate cancer with high metastatic risk: 10-year results of an EORTC randomised study. *Lancet Oncol.* 11, 1066–1073. [https://doi.org/10.1016/S1470-2045\(10\)70223-0](https://doi.org/10.1016/S1470-2045(10)70223-0)
- Brinkman, E.K., Chen, T., Amendola, M., Van Steensel, B., 2014. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* 42. <https://doi.org/10.1093/nar/gku936>
- Brinkmann, A.O., Blok, L.J., De Ruiter, P.E., Doesburg, P., Steketee, K., Berrevoets, C.A., Trapman, J., 1999. Mechanisms of androgen receptor activation and function, in: *Journal of Steroid Biochemistry and Molecular Biology*. pp. 307–313.  
[https://doi.org/10.1016/S0960-0760\(99\)00049-7](https://doi.org/10.1016/S0960-0760(99)00049-7)
- Brooke, G., Bevan, C., 2009. The Role of Androgen Receptor Mutations in Prostate Cancer Progression. *Curr. Genomics* 10, 18–25. <https://doi.org/10.2174/138920209787581307>
- Buoso, E., Galasso, M., Ronfani, M., Serafini, M.M., Lanni, C., Corsini, E., Racchi, M., 2017. Role of spliceosome proteins in the regulation of glucocorticoid receptor isoforms by cortisol and dehydroepiandrosterone. *Pharmacol. Res.* 120, 180–187.  
<https://doi.org/10.1016/j.phrs.2017.03.019>
- Cai, C., Wang, H., Xu, Y., Chen, S., Balk, S.P., 2009. Reactivation of androgen receptor-

- regulated TMPRSS2:ERG gene expression in castration-resistant prostate cancer. *Cancer Res.* 69, 6027–6032. <https://doi.org/10.1158/0008-5472.CAN-09-0395>
- Cai, Liuhong, Bai, H., Mahairaki, V., Gao, Y., He, C., Wen, Y., Jin, Y.C., Wang, Y., Pan, R.L., Qasba, A., Ye, Z., Cheng, L., 2018. A Universal Approach to Correct Various HBB Gene Mutations in Human Stem Cells for Gene Therapy of Beta-Thalassemia and Sickle Cell Disease. *Stem Cells Transl. Med.* 7, 87–97. <https://doi.org/10.1002/sctm.17-0066>
- Cai, Ling, 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 in Castrate-Resistant Prostate Cancer. *Mol. Cell* 72, 341-354.e6. <https://doi.org/10.1016/j.molcel.2018.08.029>
- 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, 1646–56. <https://doi.org/10.18632/oncotarget.1802>
- Cao, S., Zhan, Y., Dong, Y., 2016. Emerging data on androgen receptor splice variants in prostate cancer. *Endocr. Relat. Cancer.* <https://doi.org/10.1530/ERC-16-0298>
- Cappellari, M., Bielli, P., Paronetto, M.P., Ciccocanti, F., Fimia, G.M., Saarikettu, J., Silvennoinen, O., Sette, C., 2014. The transcriptional co-Activator SND1 is a novel regulator of alternative splicing in prostate cancer cells. *Oncogene* 33, 3794–3802. <https://doi.org/10.1038/onc.2013.360>
- Carlson-Stevermer, J., Abdeen, A.A., Kohlenberg, L., Goedland, M., Molugu, K., Lou, M., Saha, K., 2017. Assembly of CRISPR ribonucleoproteins with biotinylated oligonucleotides via an RNA aptamer for precise gene editing. *Nat. Commun.* 8. <https://doi.org/10.1038/s41467-017-01875-9>
- 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., Balk, S.P., Den, R.B., Trock, B.J., Karnes, R.J., Jenkins, R.B., Klein, E.A., Davicioni, E., Gruhl, F.J., Long, H.W., Liu, X.S., Cato, A.C.B., Lack, N.A., Nelson, P.S., Plymate, S.R., Groner, A.C., Brown, M., 2019. ARv7 Represses Tumor-Suppressor Genes in Castration-Resistant Prostate Cancer. *Cancer Cell* 35, 401-413.e6. <https://doi.org/10.1016/j.ccell.2019.01.008>
- Chan, S.C., Selth, L.A., Li, Y., Nyquist, M.D., Miao, L., Bradner, J.E., Raj, G. V, Tilley, W.D., Dehm, S.M., 2015. Targeting chromatin binding regulation of constitutively active AR variants to overcome prostate cancer resistance to endocrine-based therapies. *Nucleic Acids Res.* 43, 5880–97. <https://doi.org/10.1093/nar/gkv262>
- Chandrasekar, T., Yang, J.C., Gao, A.C., Evans, C.P., 2015. Mechanisms of resistance in castration-resistant prostate cancer (CRPC). *Transl. Androl. Urol.* 4, 365–80. <https://doi.org/10.3978/j.issn.2223-4683.2015.05.02>
- Chaytor, L., Simcock, M., Nakjang, S., Heath, R., Walker, L., Robson, C., Jones, D., Gaughan, L., 2019. The Pioneering Role of GATA2 in Androgen Receptor Variant Regulation Is Controlled by Bromodomain and Extraterminal Proteins in Castrate-Resistant Prostate

- Cancer. *Mol. Cancer Res.* 17, 1264–1278. <https://doi.org/10.1158/1541-7786.MCR-18-1231>
- 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. *Nat. Med.* 10, 33–9. <https://doi.org/10.1038/nm972>
- Chen, Z., Trotman, L.C., Shaffer, D., Lin, H.-K., Dotan, Z.A., Niki, M., Koutcher, J.A., Scher, H.I., Ludwig, T., Gerald, W., Cordon-Cardo, C., Pandolfi, P.P., 2005. Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature* 436, 725–30. <https://doi.org/10.1038/nature03918>
- 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., Wang, Q., 2018. Diverse AR-V7 cistromes in castration-resistant prostate cancer are governed by HoxB13. *Proc. Natl. Acad. Sci. U. S. A.* 115, 6810–6815. <https://doi.org/10.1073/pnas.1718811115>
- Claessens, F., Helsen, C., Prekovic, S., Van den Broeck, T., Spans, L., Van Poppel, H., Joniau, S., 2014. Emerging mechanisms of enzalutamide resistance in prostate cancer. *Nat. Rev. Urol.* 11, 712–6. <https://doi.org/10.1038/nrurol.2014.243>
- Clancy, S., 2008. RNA splicing: introns, exons and spliceosome. *Nat. Educ.*
- Clegg, N.J., Wongvipat, J., Joseph, J.D., Tran, C., Ouk, S., Dilhas, A., Chen, Y., Grillot, K., Bischoff, E.D., Cai, L., Aparicio, A., Dorow, S., Arora, V., Shao, G., Qian, J., Zhao, H., Yang, G., Cao, C., Sensintaffar, J., Wasielewska, T., Herbert, M.R., Bonnefous, C., Darimont, B., Scher, H.I., Smith-Jones, P., Klang, M., Smith, N.D., De Stanchina, E., Wu, N., Ouerfelli, O., Rix, P.J., Heyman, R.A., Jung, M.E., Sawyers, C.L., Hager, J.H., 2012. ARN-509: A novel antiandrogen for prostate cancer treatment. *Cancer Res.* 72, 1494–1503. <https://doi.org/10.1158/0008-5472.CAN-11-3948>
- Clinckemalie, L., Vanderschueren, D., Boonen, S., Claessens, F., 2012. The hinge region in androgen receptor control. *Mol. Cell. Endocrinol.* 358, 1–8. <https://doi.org/10.1016/j.mce.2012.02.019>
- 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., Climent, M.A., Duran, I., Gallardo, E., Rodriguez, A., Santander, C., Sáez, M.I., Puente, J., Gasi Tandefelt, D., Wingate, A., Dearnaley, D., PREMIERE Collaborators, Spanish Oncology Genitourinary Group, Demichelis, F., De Giorgi, U., Gonzalez-Billalabeitia, E., Attard, G., 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. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* 28, 1508–1516. <https://doi.org/10.1093/annonc/mdx155>
- 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 Dev.* 25, 445–59. <https://doi.org/10.1101/gad.2014311>

- Cottard, F., Asmane, I., Erdmann, E., Bergerat, J.P., Kurtz, J.E., Céraline, J., 2013. Constitutively Active Androgen Receptor Variants Upregulate Expression of Mesenchymal Markers in Prostate Cancer Cells. *PLoS One* 8. <https://doi.org/10.1371/journal.pone.0063466>
- 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. *J. Cell Sci.* 121, 957–68. <https://doi.org/10.1242/jcs.022103>
- De Laere, B., van Dam, P.-J., Whittington, T., Mayrhofer, M., Diaz, E.H., Van den Eynden, G., Vandebroek, J., Del-Favero, J., Van Laere, S., Dirix, L., Grönberg, H., Lindberg, J., 2017. Comprehensive Profiling of the Androgen Receptor in Liquid Biopsies from Castration-resistant Prostate Cancer Reveals Novel Intra-AR Structural Variation and Splice Variant Expression Patterns. *Eur. Urol.* 72, 192–200. <https://doi.org/10.1016/j.eururo.2017.01.011>
- Dehm, S.M., Schmidt, L.J., Heemers, H. V, Vessella, R.L., Tindall, D.J., 2008. Splicing of a novel androgen receptor exon generates a constitutively active androgen receptor that mediates prostate cancer therapy resistance. *Cancer Res.* 68, 5469–77. <https://doi.org/10.1158/0008-5472.CAN-08-0594>
- Dehm, S.M., Tindall, D.J., 2011. Alternatively spliced androgen receptor variants. *Endocr. Relat. Cancer* 18, R183-96. <https://doi.org/10.1530/ERC-11-0141>
- Denichenko, P., Mogilevsky, M., Cléry, A., Welte, T., Biran, J., Shimshon, O., Barnabas, G.D., Danan-Gotthold, 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. *Nat. Commun.* 10. <https://doi.org/10.1038/s41467-019-09523-0>
- Denmeade, S.R., Isaacs, J.T., 2002. Access : A history of prostate cancer treatment : Nature Reviews Cancer. *Nat. Rev. Cancer* 2, 389–96. <https://doi.org/10.1038/nrc801>
- Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H.W., Listgarten, J., Root, D.E., 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* 34, 184–191. <https://doi.org/10.1038/nbt.3437>
- Doudna, J.A., Charpentier, E., 2014. The new frontier of genome engineering with CRISPR-Cas9. *Science* (80-. ). <https://doi.org/10.1126/science.1258096>
- Duff, J., McEwan, I.J., 2005. Mutation of histidine 874 in the androgen receptor ligand-binding domain leads to promiscuous ligand activation and altered p160 coactivator interactions. *Mol. Endocrinol.* 19, 2943–54. <https://doi.org/10.1210/me.2005-0231>
- Dvinge, H., Bradley, R.K., 2015. Widespread intron retention diversifies most cancer transcriptomes. *Genome Med.* 7, 45. <https://doi.org/10.1186/s13073-015-0168-9>
- 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, 313–9. [https://doi.org/10.1016/0092-8674\(80\)90617-0](https://doi.org/10.1016/0092-8674(80)90617-0)
- El Marabti, E., Younis, I., 2018. The cancer spliceome: Reprogramming of alternative splicing in cancer. *Front. Mol. Biosci.* 5, 1–11. <https://doi.org/10.3389/fmolb.2018.00080>

- Everman, J.L., Rios, C., Seibold, M.A., 2018. Primary airway epithelial cell gene editing using CRISPR-Cas9, in: *Methods in Molecular Biology*. Humana Press Inc., pp. 267–292. [https://doi.org/10.1007/978-1-4939-7471-9\\_15](https://doi.org/10.1007/978-1-4939-7471-9_15)
- Fei, T., Chen, Y., Xiao, T., Li, W., Cato, L., Zhang, P., Cotter, M.B., Bowden, M., Lis, R.T., Zhao, S.G., Wu, Q., Feng, F.Y., Loda, M., He, H.H., Liu, X.S., Brown, M., 2017. Genome-wide CRISPR screen identifies HNRNPL as a prostate cancer dependency regulating RNA splicing. *Proc. Natl. Acad. Sci. U. S. A.* 114, E5207–E5215. <https://doi.org/10.1073/pnas.1617467114>
- Ferraldeschi, R., Sharifi, N., Auchus, R.J., Attard, G., 2013. Molecular Pathways: Inhibiting Steroid Biosynthesis in Prostate Cancer. *Clin Cancer Res* 19. <https://doi.org/10.1158/1078-0432.CCR-12-0931>
- Ferraldeschi, R., Welti, J., Powers, M. V., Yuan, W., Smyth, T., Seed, G., Riisnaes, R., Hedayat, S., Wang, H., Crespo, M., Rodrigues, D.N., Figueiredo, I., Miranda, S., Carreira, S., Lyons, J.F., Sharp, S., Plymate, S.R., Attard, G., Wallis, N., Workman, P., De Bono, J.S., 2016. Second-generation HSP90 inhibitor onalespib blocks mRNA splicing of androgen receptor variant 7 in prostate cancer cells. *Cancer Res.* 76, 2731–2742. <https://doi.org/10.1158/0008-5472.CAN-15-2186>
- 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. *Lancet Oncol.* 13, 983–992. [https://doi.org/10.1016/S1470-2045\(12\)70379-0](https://doi.org/10.1016/S1470-2045(12)70379-0)
- Fujita, T., Fujii, H., 2016. Isolation of specific genomic regions and identification of associated molecules by enChIP. *J. Vis. Exp.* 2016. <https://doi.org/10.3791/53478>
- Fujita, T., Yuno, M., Fujii, H., 2016. Efficient sequence-specific isolation of DNA fragments and chromatin by in vitro enChIP technology using recombinant CRISPR ribonucleoproteins. *Genes Cells* 21, 370–7. <https://doi.org/10.1111/gtc.12341>
- Ganesan, S., 2018. Tumor Suppressor Tolerance: Reversion Mutations in BRCA1 and BRCA2 and Resistance to PARP Inhibitors and Platinum. *JCO Precis. Oncol.* 1–4. <https://doi.org/10.1200/po.18.00001>
- Gao, W., Bohl, C.E., Dalton, J.T., 2005. Chemistry and structural biology of androgen receptor. *Chem. Rev.* <https://doi.org/10.1021/cr020456u>
- Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., Liu, D.R., 2017. Programmable base editing of T to G C in genomic DNA without DNA cleavage. *Nature* 551, 464–471. <https://doi.org/10.1038/nature24644>
- Gaughan, L., Logan, I.R., Cook, S., Neal, D.E., Robson, C.N., 2002. Tip60 and histone deacetylase 1 regulate androgen receptor activity through changes to the acetylation status of the receptor. *J. Biol. Chem.* 277, 25904–13. <https://doi.org/10.1074/jbc.M203423200>
- Gaughan, L., Stockley, J., Wang, N., McCracken, S.R.C., Treumann, A., Armstrong, K.,

- Shaheen, F., Watt, K., McEwan, I.J., Wang, C., Pestell, R.G., Robson, C.N., 2011. Regulation of the androgen receptor by SET9-mediated methylation. *Nucleic Acids Res.* 39, 1266–1279. <https://doi.org/10.1093/nar/gkq861>
- Ghosh, D., Venkataramani, P., Nandi, S., Bhattacharjee, S., 2019. CRISPR-Cas9 a boon or bane: The bumpy road ahead to cancer therapeutics 06 Biological Sciences 0604 Genetics. *Cancer Cell Int.* <https://doi.org/10.1186/s12935-019-0726-0>
- Graf, R., Li, X., Chu, V.T., Rajewsky, K., 2019. sgRNA Sequence Motifs Blocking Efficient CRISPR/Cas9-Mediated Gene Editing. *Cell Rep.* 26, 1098-1103.e3. <https://doi.org/10.1016/j.celrep.2019.01.024>
- Greco, G.E., Matsumoto, Y., Brooks, R.C., Lu, Z., Lieber, M.R., Tomkinson, A.E., 2016. SCR7 is neither a selective nor a potent inhibitor of human DNA ligase IV. *DNA Repair (Amst).* 43, 18–23. <https://doi.org/10.1016/j.dnarep.2016.04.004>
- 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 Res.* 69, 2305–2313. <https://doi.org/10.1158/0008-5472.CAN-08-3795>
- Gutierrez-Triana, J.A., Tavhelidse, T., Thumberger, T., Thomas, I., Wittbrodt, B., Kellner, T., Anlas, K., Tsingos, E., Wittbrodt, J., 2018. Efficient single-copy HDR by 5' modified long dsDNA donors. *Elife* 7. <https://doi.org/10.7554/eLife.39468>
- Haelens, A., Verrijdt, G., Callewaert, L., Peeters, B., Rombauts, W., Claessens, F., 2001. Androgen-receptor-specific DNA binding to an element in the first exon of the human secretory component gene. *Biochem. J.* 353, 611–20. <https://doi.org/10.1042/0264-6021:3530611>
- Hay, C., McEwan, I., 2012. The Impact of Point Mutations in the Human Androgen Receptor Receptor: Classification of Mutations on the Basis of Transcriptional Activity. *PLOS One.* 7(3). <https://doi.org/10.1371/journal.pone.0032514>
- 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 Res.* 63, 149–53.
- 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. *J. Biol. Chem.* 275, 22986–94. <https://doi.org/10.1074/jbc.M002807200>
- He, B., Minges, J.T., Lee, L.W., Wilson, E.M., 2002. The FXXLF motif mediates androgen receptor-specific interactions with coregulators. *J. Biol. Chem.* 277, 10226–10235. <https://doi.org/10.1074/jbc.M111975200>
- He, Y., Lu, J., Ye, Z., Hao, S., Wang, Liewei, Kohli, M., Tindall, D.J., Li, B., Zhu, R., Wang, Liguoguo, Huang, H., 2018. Androgen receptor splice variants bind to constitutively open chromatin and promote abiraterone-resistant growth of prostate cancer. *Nucleic Acids Res.* 46, 1895–1911. <https://doi.org/10.1093/nar/gkx1306>
- Heemers, H. V, Sebo, T.J., Debes, J.D., Regan, K.M., Raclaw, K.A., Murphy, L.M., Hobisch, A.,

- Culig, Z., Tindall, D.J., 2007. Androgen deprivation increases p300 expression in prostate cancer cells. *Cancer Res.* 67, 3422–30. <https://doi.org/10.1158/0008-5472.CAN-06-2836>
- Heemers, H. V, Tindall, D.J., 2007. Androgen receptor (AR) coregulators: a diversity of functions converging on and regulating the AR transcriptional complex. *Endocr. Rev.* 28, 778–808. <https://doi.org/10.1210/er.2007-0019>
- Hiort, O., 2013. The differential role of androgens in early human sex development. *BMC Med.* <https://doi.org/10.1186/1741-7015-11-152>
- Ho, C.K., Shuman, S., 1999. Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol. Cell* 3, 405–11. [https://doi.org/10.1016/s1097-2765\(00\)80468-2](https://doi.org/10.1016/s1097-2765(00)80468-2)
- Holly, J.M.P., Broadhurst, J., Mansor, R., Bahl, A., Perks, C.M., 2017. Hyperglycemia promotes TMPRSS2-ERG gene fusion in prostate cancer cells via upregulating insulin-like growth factor-binding protein-2. *Front. Endocrinol. (Lausanne)*. 8. <https://doi.org/10.3389/fendo.2017.00305>
- Horlbeck, M.A., Witkowsky, L.B., Guglielmi, B., Replogle, J.M., Gilbert, L.A., Villalta, J.E., Torigoe, S.E., Tjian, R., Weissman, J.S., 2016. Nucleosomes impede cas9 access to DNA in vivo and in vitro. *Elife* 5. <https://doi.org/10.7554/eLife.12677>
- Hsin, J.P., Manley, J.L., 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.* <https://doi.org/10.1101/gad.200303.112>
- Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., Cradick, T.J., Marraffini, L.A., Bao, G., Zhang, F., 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31, 827–832. <https://doi.org/10.1038/nbt.2647>
- Hu, J.H., Miller, S.M., Geurts, M.H., Tang, W., Chen, L., Sun, N., Zeina, C.M., Gao, X., Rees, H.A., Lin, Z., Liu, D.R., 2018. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*. <https://doi.org/10.1038/nature26155>
- 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 Res.* 69, 16–22. <https://doi.org/10.1158/0008-5472.CAN-08-2764>
- 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 Res.* 72, 3457–62. <https://doi.org/10.1158/0008-5472.CAN-11-3892>
- 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 Res.* 1.
- Ihry, R.J., Salick, M.R., Ho, D.J., Hoffman, G.R., Dolmetsch, R., Kaykas, A., 2019. Genome-

- Scale CRISPR Screens Identify Human Pluripotency-Specific Genes.  
<https://doi.org/10.1016/j.celrep.2019.03.043>
- Inui, M., Miyado, M., Igarashi, M., Tamano, M., Kubo, A., Yamashita, S., Asahara, H., Fukami, M., Takada, S., 2014. Rapid generation of mouse models with defined point mutations by the CRISPR/Cas9 system. *Sci. Rep.* 4. <https://doi.org/10.1038/srep05396>
- 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. *J. Biol. Chem.* 270, 7341–6. <https://doi.org/10.1074/jbc.270.13.7341>
- Jia, R., Ajiro, M., Yu, L., McCoy, P., Zheng, Z.-M., 2019. Oncogenic splicing factor SRSF3 regulates ILF3 alternative splicing to promote cancer cell proliferation and transformation. *RNA* 25, 630–644. <https://doi.org/10.1261/rna.068619.118>
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* (80-. ). 337, 816–821. <https://doi.org/10.1126/science.1225829>
- Johnson, S.A., Kim, H., Erickson, B., Bentley, D.L., 2011. The export factor Yra1 modulates mRNA 3' end processing. *Nat. Struct. Mol. Biol.* 18, 1164–71. <https://doi.org/10.1038/nsmb.2126>
- Jones, D., Wade, M., Nakjang, S., Chaytor, L., Grey, J., Robson, C.N., Gaughan, L., 2015. FOXA1 regulates androgen receptor variant activity in models of castrate-resistant prostate cancer. *Oncotarget* 6, 29782–29794. <https://doi.org/10.18632/oncotarget.4927>
- 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 Discov.* 3, 1020–9. <https://doi.org/10.1158/2159-8290.CD-13-0226>
- 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. *Br. J. Cancer* 119, 347–356. <https://doi.org/10.1038/s41416-018-0172-0>
- Kampmann, M., 2018. CRISPRi and CRISPRa Screens in Mammalian Cells for Precision Biology and Medicine. *ACS Chem. Biol.* 13, 406–416. <https://doi.org/10.1021/acscchembio.7b00657>
- Karantanos, T., Evans, C.P., Tombal, B., Thompson, T.C., Montironi, R., Isaacs, W.B., 2015. Understanding the mechanisms of androgen deprivation resistance in prostate cancer at the molecular level. *Eur. Urol.* 67, 470–479. <https://doi.org/10.1016/j.eururo.2014.09.049>
- 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 Res.* 3, canres.3965.2018. <https://doi.org/10.1158/0008-5472.can-18-3965>

- Kim, S., Kim, D., Cho, S., Kim, J., Kim, J.-S., 2014. Highly Efficient RNA-guide genome editing... *Genome Res.* 28, 1–32. <https://doi.org/10.1101/gr.171322.113>. Freely
- Kiskinis, E., Sandoe, J., Williams, L.A., Boulting, G.L., Moccia, R., Wainger, B.J., Han, S., Peng, T., Thams, S., Mikkilineni, S., Mellin, C., Merkle, F.T., Davis-Dusenbery, B.N., Ziller, M., Oakley, D., Ichida, J., Di Costanzo, S., Atwater, N., Maeder, M.L., Goodwin, M.J., Nemesh, J., Handsaker, R.E., Paull, D., Noggle, S., McCarroll, S.A., Joung, J.K., Woolf, C.J., Brown, R.H., Eggan, K., 2014. Pathways disrupted in human ALS motor neurons identified through genetic correction of mutant SOD1. *Cell Stem Cell* 14, 781–95. <https://doi.org/10.1016/j.stem.2014.03.004>
- Klotz, L., Zhang, L., Lam, A., Nam, R., Mamedov, A., Loblaw, A., 2010. Clinical results of long-term follow-up of a large, active surveillance cohort with localized prostate cancer. *J. Clin. Oncol.* 28, 126–131. <https://doi.org/10.1200/JCO.2009.24.2180>
- Klotz, L.H., Herr, H.W., Morse, M.J., Whitmore, W.F., 1986. Intermittent endocrine therapy for advanced prostate cancer. *Cancer* 58, 2546–50. [https://doi.org/10.1002/1097-0142\(19861201\)58:11<2546::aid-cnrc2820581131>3.0.co;2-n](https://doi.org/10.1002/1097-0142(19861201)58:11<2546::aid-cnrc2820581131>3.0.co;2-n)
- Kluth, L.A., Shariat, S.F., Kratzik, C., Tagawa, S., Sonpavde, G., Rieken, M., Scherr, D.S., Pummer, K., 2014. The hypothalamic-pituitary-gonadal axis and prostate cancer: Implications for androgen deprivation therapy. *World J. Urol.* <https://doi.org/10.1007/s00345-013-1157-5>
- Kobayashi, Y., Kitamoto, T., Masuhiro, Y., Watanabe, M., Kase, T., Metzger, D., Yanagisawa, J., Kato, S., 2000. p300 mediates functional synergism between AF-1 and AF-2 of estrogen receptor alpha and beta by interacting directly with the N-terminal A/B domains. *J. Biol. Chem.* 275, 15645–51. <https://doi.org/10.1074/jbc.M000042200>
- Kohli, M., Ho, Y., Hillman, D.W., Van Etten, J.L., Henzler, C., Yang, R., Sperger, J.M., Li, Y., Tseng, E., Hon, T., Clark, T., Tan, W., Carlson, R.E., Wang, Ligu, Sicotte, H., Thai, H., Jimenez, R., Huang, H., Vedell, P.T., Eckloff, B.W., Quevedo, J.F., Pitot, H.C., Costello, B.A., Jen, J., Wieben, E.D., Silverstein, K.A.T., Lang, J.M., Wang, Liewei, Dehm, S.M., 2017. Androgen receptor variant AR-V9 is coexpressed with AR-V7 in prostate cancer metastases and predicts abiraterone resistance. *Clin. Cancer Res.* 23, 4704–4715. <https://doi.org/10.1158/1078-0432.CCR-17-0017>
- Kornblihtt, A.R., Schor, I.E., Alló, M., Dujardin, G., Petrillo, E., Muñoz, M.J., 2013. Alternative splicing: A pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* <https://doi.org/10.1038/nrm3525>
- Korpala, 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 Discov.* 3, 1030–43. <https://doi.org/10.1158/2159-8290.CD-13-0142>
- Kote-Jarai, Z., Leongamornlert, D., Saunders, E., Tymrakiewicz, M., Castro, E., Mahmud, N., Guy, M., Edwards, S., O'Brien, L., Sawyer, E., Hall, A., Wilkinson, R., Dadaev, T., Goh, C., Easton, D., UKGPCS Collaborators, Goldgar, D., Eeles, R., 2011. BRCA2 is a moderate penetrance gene contributing to young-onset prostate cancer: implications for genetic

- testing in prostate cancer patients. *Br. J. Cancer* 105, 1230–4.  
<https://doi.org/10.1038/bjc.2011.383>
- Kuranaga, Y., Sugito, N., Shinohara, H., Tsujino, T., Taniguchi, K., Komura, K., Ito, Y., Soga, T., Akao, Y., 2018. SRSF3, a Splicer of the PKM Gene, Regulates Cell Growth and Maintenance of Cancer-Specific Energy Metabolism in Colon Cancer Cells. *Int. J. Mol. Sci.* 19. <https://doi.org/10.3390/ijms19103012>
- Lai, W.S., Carballo, E., Thorn, J.M., Kennington, E.A., Blackshear, P.J., 2000. Interactions of CCCH zinc finger proteins with mRNA. Binding of tristetraprolin-related zinc finger proteins to AU-rich elements and destabilization of mRNA. *J. Biol. Chem.* 275, 17827–17837. <https://doi.org/10.1074/jbc.M001696200>
- Langley, E., Kempainen, J.A., Wilson, E.M., 1998. Intermolecular NH<sub>2</sub>-/carboxyl-terminal interactions in androgen receptor dimerization revealed by mutations that cause androgen insensitivity. *J. Biol. Chem.* 273, 92–101.  
<https://doi.org/10.1074/jbc.273.1.92>
- Lawton, C.A., DeSilvio, M., Roach, M., Uhl, V., Kirsch, R., Seider, M., Rotman, M., Jones, C., Asbell, S., Valicenti, R., Hahn, S., Thomas, C.R., 2007. An Update of the Phase III Trial Comparing Whole Pelvic to Prostate Only Radiotherapy and Neoadjuvant to Adjuvant Total Androgen Suppression: Updated Analysis of RTOG 94-13, With Emphasis on Unexpected Hormone/Radiation Interactions. *Int. J. Radiat. Oncol. Biol. Phys.* 69, 646–655. <https://doi.org/10.1016/j.ijrobp.2007.04.003>
- Lee, S.C.W., Abdel-Wahab, O., 2016. Therapeutic targeting of splicing in cancer. *Nat. Med.*  
<https://doi.org/10.1038/nm.4165>
- Li, G., Liu, D., Zhang, X., Quan, R., Zhong, C., Mo, J., Huang, Y., Wang, H., Ruan, X., Xu, Z., Zheng, E., Gu, T., Hong, L., Li, Z., Wu, Z., Yang, H., 2018. Suppressing Ku70/Ku80 expression elevates homology-directed repair efficiency in primary fibroblasts. *Int. J. Biochem. Cell Biol.* 99, 154–160. <https://doi.org/10.1016/j.biocel.2018.04.011>
- Li, L., Karanika, S., Yang, G., Wang, J., Park, S., Broom, B.M., Manyam, G.C., Wu, W., Luo, Y., Basourakos, S., Song, J.H., Gallick, G.E., Karantanos, T., Korentzelos, D., Azad, A.K., Kim, J., Corn, P.G., Aparicio, A.M., Logothetis, C.J., Troncoso, P., Heffernan, T., Toniatti, C., Lee, H.-S., Lee, J.-S., Zuo, X., Chang, W., Yin, J., Thompson, T.C., 2017. Androgen receptor inhibitor-induced “BRCAness” and PARP inhibition are synthetically lethal for castration-resistant prostate cancer. *Sci. Signal.* 10.  
<https://doi.org/10.1126/scisignal.aam7479>
- Li, Y., Alsagabi, M., Fan, D., Bova, G.S., Tewfik, A.H., Dehm, S.M., 2011. Intragenic rearrangement and altered RNA splicing of the androgen receptor in a cell-based model of prostate cancer progression. *Cancer Res.* 71, 2108–17.  
<https://doi.org/10.1158/0008-5472.CAN-10-1998>
- 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 Res.* 73, 483–9. <https://doi.org/10.1158/0008-5472.CAN-12-3630>
- Li, Y., Hwang, T.H., Oseth, L.A., Hauge, A., Vessella, R.L., Schmechel, S.C., Hirsch, B.,

- Beckman, K.B., Silverstein, K.A., Dehm, S.M., 2012. AR intragenic deletions linked to androgen receptor splice variant expression and activity in models of prostate cancer progression. *Oncogene* 31, 4759–67. <https://doi.org/10.1038/onc.2011.637>
- Liang, X., Potter, J., Kumar, S., Zou, Y., Quintanilla, R., Sridharan, M., Carte, J., Chen, W., Roark, N., Ranganathan, S., Ravinder, N., Chesnut, J.D., 2015. Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J. Biotechnol.* 208, 44–53. <https://doi.org/10.1016/j.jbiotec.2015.04.024>
- 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 Res.* 67, 9001–5. <https://doi.org/10.1158/0008-5472.CAN-07-1072>
- Lin, S., Staahl, B.T., Alla, R.K., Doudna, J.A., 2014. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery 3, 4766. <https://doi.org/10.7554/eLife.04766>
- Linja, M.J., Savinainen, K.J., Saramäki, O.R., Tammela, T.L., Vessella, R.L., Visakorpi, T., 2001. Amplification and overexpression of androgen receptor gene in hormone-refractory prostate cancer. *Cancer Res.* 61, 3550–5.
- 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, 3140–50. <https://doi.org/10.1038/onc.2013.284>
- Liu, M., Rehman, S., Tang, X., Gu, K., Fan, Q., Chen, D., Ma, W., 2019. Methodologies for improving HDR efficiency. *Front. Genet.* <https://doi.org/10.3389/fgene.2018.00691>
- Liu, X., Zhang, Y., Chen, Y., Li, M., Zhou, F., Li, K., Cao, H., Ni, M., Liu, Y., Gu, Z., Dickerson, K.E., Xie, S., Hon, G.C., Xuan, Z., Zhang, M.Q., Shao, Z., Xu, J., 2017. In Situ Capture of Chromatin Interactions by Biotinylated dCas9. *Cell* 170, 1028-1043.e19. <https://doi.org/10.1016/j.cell.2017.08.003>
- Long, C., Li, H., Tiburcy, M., Rodriguez-Caycedo, C., Kyrychenko, V., Zhou, H., Zhang, Y., Min, Y.L., Shelton, J.M., Mammen, P.P.A., Liaw, N.Y., Zimmermann, W.H., Bassel-Duby, R., Schneider, J.W., Olson, E.N., 2018. Correction of diverse muscular dystrophy mutations in human engineered heart muscle by single-site genome editing. *Sci. Adv.* 4. <https://doi.org/10.1126/sciadv.aap9004>
- Lord, C.J., Ashworth, A., 2017. PARP Inhibitors: The First Synthetic Lethal Targeted Therapy. *Science* 355, 1152. <https://doi.org/10.1126/SCIENCE.AAM7344>
- Lu, J., Van der Steen, T., Tindall, D.J., 2015. Are androgen receptor variants a substitute for the full-length receptor? *Nat. Rev. Urol.* 12, 137–44. <https://doi.org/10.1038/nrurol.2015.13>
- Luna Velez, M. V., Verhaegh, G.W., Smit, F., Sedelaar, J.P.M., Schalken, J.A., 2019. Suppression of prostate tumor cell survival by antisense oligonucleotide-mediated inhibition of AR-V7 mRNA synthesis. *Oncogene* 38, 3696–3709. <https://doi.org/10.1038/s41388-019-0696-7>
- Luo, J., Attard, G., Balk, S.P., Bevan, C., Burnstein, K., Cato, L., Cherkasov, A., De Bono, J.S.,

- Dong, Y., Gao, A.C., Gleave, M., Heemers, H., Kanayama, M., Kittler, R., Lang, J.M., Lee, R.J., Logothetis, C.J., Matusik, R., Plymate, S., Sawyers, C.L., Selth, L.A., Soule, H., Tilley, W., Weigel, N.L., Zoubeidi, A., Dehm, S.M., Raj, G. V, 2018. Role of Androgen Receptor Variants in Prostate Cancer: Report from the 2017 Mission Androgen Receptor Variants Meeting. *Eur. Urol.* 73, 715–723. <https://doi.org/10.1016/j.eururo.2017.11.038>
- Ma, M., Zhuang, F., Hu, X., Wang, B., Wen, X.Z., Ji, J.F., Xi, J.J., 2017. Efficient generation of mice carrying homozygous double-floxp alleles using the Cas9-Avidin/Biotin-donor DNA system. *Cell Res.* <https://doi.org/10.1038/cr.2017.29>
- Maitland, N.J., 2015. *European Urology Today - Vol. 27 No.3 - June/July 2015* by European Association of Urology (EAU) - issuu [WWW Document]. URL [https://issuu.com/uroweb/docs/47461\\_eut\\_july\\_2015\\_lr/10](https://issuu.com/uroweb/docs/47461_eut_july_2015_lr/10) (accessed 9.24.19).
- Maitland, N.J., 2013. Stem Cells in the Normal and Malignant Prostate, in: Tindall, D.J. (Ed.), *Prostate Cancer: Biochemistry, Molecular Biology and Genetics*. Springer New York, New York, NY, pp. 3–41. [https://doi.org/10.1007/978-1-4614-6828-8\\_1](https://doi.org/10.1007/978-1-4614-6828-8_1)
- Mangelsdorf, D.J., Thummel, C., Beato, M., Herrlich, P., Schütz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P., Evans, R.M., 1995. The nuclear receptor superfamily: The second decade. *Cell* 83, 835–839. [https://doi.org/10.1016/0092-8674\(95\)90199-X](https://doi.org/10.1016/0092-8674(95)90199-X)
- Martinez-Montiel, N., Rosas-Murrieta, N.H., Ruiz, M.A., Monjaraz-Guzman, E., Martinez-Contreras, R., 2018. Alternative splicing as a target for cancer treatment. *Int. J. Mol. Sci.* 19, 1–28. <https://doi.org/10.3390/ijms19020545>
- Maruyama, T., Dougan, S.K., Truttmann, M.C., Bilate, A.M., Ingram, J.R., Ploegh, H.L., 2015. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* 33, 538–542. <https://doi.org/10.1038/nbt.3190>
- McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., Bentley, D.L., 1997. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* 385, 357–61. <https://doi.org/10.1038/385357a0>
- McNeal, J.E., 1969. Origin and development of carcinoma in the prostate. *Cancer* 23, 24–34. [https://doi.org/10.1002/1097-0142\(196901\)23:1<24::AID-CNCR2820230103>3.0.CO;2-1](https://doi.org/10.1002/1097-0142(196901)23:1<24::AID-CNCR2820230103>3.0.CO;2-1)
- Merkle, F.T., Neuhausser, W.M., Santos, D., Valen, E., Gagnon, J.A., Maas, K., Sandoe, J., Schier, A.F., Eggan, K., 2015. Efficient CRISPR-Cas9-mediated generation of knockin human pluripotent stem cells lacking undesired mutations at the targeted locus. *Cell Rep.* 11, 875–883. <https://doi.org/10.1016/j.celrep.2015.04.007>
- Mohammed, H., D’Santos, C., Serandour, A.A., Ali, H.R., Brown, G.D., Atkins, A., Rueda, O.M., Holmes, K.A., Theodorou, V., Robinson, J.L.L., Zwart, W., Saadi, A., Ross-Innes, C.S., Chin, S.F., Menon, S., Stingl, J., Palmieri, C., Caldas, C., Carroll, J.S., 2013. Endogenous Purification Reveals GREB1 as a Key Estrogen Receptor Regulatory Factor. *Cell Rep.* 3, 342–349. <https://doi.org/10.1016/j.celrep.2013.01.010>

- 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. *Nat. Protoc.* 11, 316–26. <https://doi.org/10.1038/nprot.2016.020>
- Moris, L., Van den Broeck, T., Tosco, L., Van Baelen, A., Gontero, P., Karnes, R.J., Everaerts, W., Albersen, M., Bastian, P.J., Chlosta, P., Claessens, F., Chun, F.K., Graefen, M., Gratzke, C., Kneitz, B., Marchioro, G., Salas, R.S., Tombal, B., Van Der Poel, H., Walz, J.C., De Meerleer, G., Bossi, A., Haustermans, K., Montorsi, F., Van Poppel, H., Spahn, M., Briganti, A., Joniau, S., 2016. Impact of Lymph Node Burden on Survival of High-risk Prostate Cancer Patients Following Radical Prostatectomy and Pelvic Lymph Node Dissection. *Front. Surg.* 3. <https://doi.org/10.3389/fsurg.2016.00065>
- Munkley, J., Livermore, K., Rajan, P., Elliott, D.J., 2017. RNA splicing and splicing regulator changes in prostate cancer pathology. *Hum. Genet.* 136, 1143–1154. <https://doi.org/10.1007/s00439-017-1792-9>
- Nyquist, M.D., Li, Y., Hwang, T.H., Manlove, L.S., Vessella, R.L., Silverstein, K.A.T., Voytas, D.F., Dehm, S.M., 2013. TALEN-engineered AR gene rearrangements reveal endocrine uncoupling of androgen receptor in prostate cancer. *Proc. Natl. Acad. Sci. U. S. A.* 110, 17492–17497. <https://doi.org/10.1073/pnas.1308587110>
- O'Neill, D., Jones, D., Wade, M., Grey, J., Nakjang, S., Guo, W., Cork, D., Davies, B.R., Wedge, S.R., Robson, C.N., Gaughan, L., 2015. Development and exploitation of a novel mutant androgen receptor modelling strategy to identify new targets for advanced prostate cancer therapy. *Oncotarget* 6, 26029–40. <https://doi.org/10.18632/oncotarget.4347>
- Okamoto, S., Amaishi, Y., Maki, I., Enoki, T., Mineno, J., 2019. Highly efficient genome editing for single-base substitutions using optimized ssODNs with Cas9-RNPs. *Sci. Rep.* 9. <https://doi.org/10.1038/s41598-019-41121-4>
- Paix, A., Folkmann, A., Goldman, D.H., Kulaga, H., Grzelak, M.J., Rasoloson, D., Paidemarry, S., Green, R., Reed, R.R., Seydoux, G., 2017. Precision genome editing using synthesis-dependent repair of Cas9-induced DNA breaks. *Proc. Natl. Acad. Sci. U. S. A.* 114, E10745–E10754. <https://doi.org/10.1073/pnas.1711979114>
- Paix, A., Schmidt, H., Seydoux, G., 2016. Cas9-assisted recombineering in *C. elegans*: Genome editing using in vivo assembly of linear DNAs. *Nucleic Acids Res.* 44, e128. <https://doi.org/10.1093/nar/gkw502>
- Pandya-Jones, A., 2011. Pre-mRNA splicing during transcription in the mammalian system. *Wiley Interdiscip. Rev. RNA* 2, 700–717. <https://doi.org/10.1002/wrna.86>
- Paquet, D., Kwart, D., Chen, A., Sproul, A., Jacob, S., Teo, S., Olsen, K.M., Gregg, A., Noggle, S., Tessier-Lavigne, M., 2016. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* 533, 125–129. <https://doi.org/10.1038/nature17664>
- Paschalis, A., Sharp, A., Welti, J.C., Neeb, A., Raj, G. V., Luo, J., Plymate, S.R., de Bono, J.S., 2018. Alternative splicing in prostate cancer. *Nat. Rev. Clin. Oncol.* <https://doi.org/10.1038/s41571-018-0085-0>

- Perrin-Vidoz, L., 2002. The nonsense-mediated mRNA decay pathway triggers degradation of most BRCA1 mRNAs bearing premature termination codons. *Hum. Mol. Genet.* 11, 2805–2814. <https://doi.org/10.1093/hmg/11.23.2805>
- Pinder, J., Salsman, J., Dellaire, G., 2015. Nuclear domain “knock-in” screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing. *Nucleic Acids Res.* 43, 9379–9392. <https://doi.org/10.1093/nar/gkv993>
- Polkinghorn, W.R., Parker, J.S., Lee, M.X., Kass, E.M., Spratt, D.E., Iaquina, P.J., Arora, V.K., Yen, W.-F., Cai, L., Zheng, D., Carver, B.S., Chen, Y., Watson, P.A., Shah, N.P., Fujisawa, S., Goglia, A.G., Gopalan, A., Hieronymus, H., Wongvipat, J., Scardino, P.T., Zelefsky, M.J., Jasin, M., Chaudhuri, J., Powell, S.N., Sawyers, C.L., 2013. Androgen receptor signaling regulates DNA repair in prostate cancers. *Cancer Discov.* 3, 1245–53. <https://doi.org/10.1158/2159-8290.CD-13-0172>
- Powell, I.J., Bollig-Fischer, A., 2013. Minireview: the molecular and genomic basis for prostate cancer health disparities. *Mol. Endocrinol.* 27, 879–91. <https://doi.org/10.1210/me.2013-1039>
- Puhr, M., Hoefler, J., Eigentler, A., Ploner, C., Handle, F., Schaefer, G., Kroon, J., Leo, A., Heidegger, I., Eder, I., Culig, Z., Der Pluijm, G. Van, Klocker, H., 2018. The glucocorticoid receptor is a key player for prostate cancer cell survival and a target for improved antiandrogen therapy. *Clin. Cancer Res.* 24, 927–938. <https://doi.org/10.1158/1078-0432.CCR-17-0989>
- Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., Lim, W.A., 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173–1183. <https://doi.org/10.1016/j.cell.2013.02.022>
- Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., Zhang, F., 2013. GeCKO - Genome-scale CRISPR Knock-Out. *Zhang Lab* 8, 2281–2308. <https://doi.org/10.1038/nprot.2013.143>
- Randall, V.A., 1994. Role of 5 alpha-reductase in health and disease. *Baillieres. Clin. Endocrinol. Metab.* 8, 405–31.
- Ratnadiwakara, M., Archer, S.K., Dent, C.I., De Los Mozos, I.R., Beilharz, T.H., Knaupp, A.S., Nefzger, C.M., Polo, J.M., Anko, M.L., 2018. SRSF3 promotes pluripotency through nanog mRNA export and coordination of the pluripotency gene expression program. *Elife* 7. <https://doi.org/10.7554/eLife.37419>
- Rees, H.A., Yeh, W.H., Liu, D.R., 2019. Development of hRad51–Cas9 nickase fusions that mediate HDR without double-stranded breaks. *Nat. Commun.* 10. <https://doi.org/10.1038/s41467-019-09983-4>
- Reid, J., Kelly, S.M., Watt, K., Price, N.C., McEwan, I.J., 2002a. Conformational analysis of the androgen receptor amino-terminal domain involved in transactivation. Influence of structure-stabilizing solutes and protein-protein interactions. *J. Biol. Chem.* 277, 20079–86. <https://doi.org/10.1074/jbc.M201003200>
- Reid, J., Murray, I., Watt, K., Betney, R., McEwan, I.J., 2002b. The androgen receptor interacts with multiple regions of the large subunit of general transcription factor TFIIF.

J. Biol. Chem. 277, 41247–53. <https://doi.org/10.1074/jbc.M205220200>

- Reinhardt, P., Schmid, B., Burbulla, L.F., Schöndorf, D.C., Wagner, L., Glatza, M., Höing, S., Hargus, G., Heck, S.A., Dhingra, A., Wu, G., Müller, S., Brockmann, K., Kluba, T., Maisel, M., Krüger, R., Berg, D., Tsytsyura, Y., Thiel, C.S., Psathaki, O.-E., Klingauf, J., Kuhlmann, T., Klewin, M., Müller, H., Gasser, T., Schöler, H.R., Sternecker, J., 2013. Genetic correction of a LRRK2 mutation in human iPSCs links parkinsonian neurodegeneration to ERK-dependent changes in gene expression. *Cell Stem Cell* 12, 354–67. <https://doi.org/10.1016/j.stem.2013.01.008>
- Renaud, J.B., Boix, C., Charpentier, M., De Cian, A., Cochenec, J., Duvernois-Berthet, E., Perrouault, L., Tesson, L., Edouard, J., Thinar, R., Cherifi, Y., Menoret, S., Fontanière, S., de Crozé, N., Fraichard, A., Sohm, F., Anegon, I., Concordet, J.P., Giovannangeli, C., 2016. Improved Genome Editing Efficiency and Flexibility Using Modified Oligonucleotides with TALEN and CRISPR-Cas9 Nucleases. *Cell Rep.* 14, 2263–2272. <https://doi.org/10.1016/j.celrep.2016.02.018>
- Robert, F., Barbeau, M., Éthier, S., Dostie, J., Pelletier, J., 2015. Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. *Genome Med.* 7. <https://doi.org/10.1186/s13073-015-0215-6>
- Robinson, D., Van Allen, E.M., Wu, Y.M., Schultz, N., Lonigro, R.J., Mosquera, J.M., Montgomery, B., Taplin, M.E., Pritchard, C.C., Attard, G., Beltran, H., Abida, W., Bradley, R.K., Vinson, J., Cao, X., Vats, P., Kunju, L.P., Hussain, M., Feng, F.Y., Tomlins, S.A., Cooney, K.A., Smith, D.C., Brennan, C., Siddiqui, J., Mehra, R., Chen, Y., Rathkopf, D.E., Morris, M.J., Solomon, S.B., Durack, J.C., Reuter, V.E., Gopalan, A., Gao, J., Loda, M., Lis, R.T., Bowden, M., Balk, S.P., Gaviola, G., Sougnez, C., Gupta, M., Yu, E.Y., Mostaghel, E.A., Cheng, H.H., Mulcahy, H., True, L.D., Plymate, S.R., Dvinge, H., Ferraldeschi, R., Flohr, P., Miranda, S., Zafeiriou, Z., Tunariu, N., Mateo, J., Perez-Lopez, R., Demichelis, F., Robinson, B.D., Schiffman, M., Nanus, D.M., Tagawa, S.T., Sigaras, A., Eng, K.W., Elemento, O., Sboner, A., Heath, E.I., Scher, H.I., Pienta, K.J., Kantoff, P., De Bono, J.S., Rubin, M.A., Nelson, P.S., Garraway, L.A., Sawyers, C.L., Chinnaiyan, A.M., 2015. Integrative clinical genomics of advanced prostate cancer. *Cell* 161, 1215–1228. <https://doi.org/10.1016/j.cell.2015.05.001>
- Robinson, J.L.L., Hickey, T.E., Warren, A.Y., Vowler, S.L., Carroll, T., Lamb, A.D., Papoutsoglou, N., Neal, D.E., Tilley, W.D., Carroll, J.S., 2014. Elevated levels of FOXA1 facilitate androgen receptor chromatin binding resulting in a CRPC-like phenotype. *Oncogene* 33, 5666–74. <https://doi.org/10.1038/onc.2013.508>
- Roche, P.J.R., Gytz, H., Hussain, F., Cameron, C.J.F., Paquette, D., Blanchette, M., Dostie, J., Nagar, B., Akavia, U.D., 2018. Double-Stranded Biotinylated Donor Enhances Homology-Directed Repair in Combination with Cas9 Monoavidin in Mammalian Cells. *Cris. J.* 1, 414–430. <https://doi.org/10.1089/crispr.2018.0045>
- Rossidis, A.C., Stratigis, J.D., Chadwick, A.C., Hartman, H.A., Ahn, N.J., Li, H., Singh, K., Coons, B.E., Li, L., Lv, W., Zoltick, P.W., Alapati, D., Zacharias, W., Jain, R., Morrissey, E.E., Musunuru, K., Peranteau, W.H., 2018. In utero CRISPR-mediated therapeutic editing of metabolic genes. *Nat. Med.* 24, 1513–1518. <https://doi.org/10.1038/s41591-018-0184-6>

- Ruan, J., Li, H., Xu, K., Wu, T., Wei, J., Zhou, R., Liu, Z., Mu, Y., Yang, S., Ouyang, H., Chen-Tsai, R.Y., Li, K., 2015. Highly efficient CRISPR/Cas9-mediated transgene knockin at the H11 locus in pigs. *Sci. Rep.* 5. <https://doi.org/10.1038/srep14253>
- 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., Flaig, T.W., Taplin, M.-E., Higano, C.S., de Souza, P., de Bono, J.S., Griffin, T.W., De Porre, P., Yu, M.K., Park, Y.C., Li, J., Kheoh, T., Naini, V., Molina, A., Rathkopf, D.E., COU-AA-302 Investigators, 2015. Abiraterone acetate plus prednisone versus placebo plus prednisone in chemotherapy-naive men with metastatic castration-resistant prostate cancer (COU-AA-302): final overall survival analysis of a randomised, double-blind, placebo-controlled phase 3 study. *Lancet Oncol.* 16, 152–60. [https://doi.org/10.1016/S1470-2045\(14\)71205-7](https://doi.org/10.1016/S1470-2045(14)71205-7)
- Sander, J.D., Joung, J.K., 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* <https://doi.org/10.1038/nbt.2842>
- Schenkwein, D., Ylä-Herttuala, S., 2018. Gene Editing of Human Embryos with CRISPR/Cas9: Great Promise Coupled with Important Caveats. *Mol. Ther.* <https://doi.org/10.1016/j.ymthe.2018.02.007>
- Scher, H.I., Beer, T.M., Higano, C.S., Anand, A., Taplin, M.-E., Efstathiou, E., Rathkopf, D., Shelkey, J., Yu, E.Y., Alumkal, J., Hung, D., Hirmand, M., Seely, L., Morris, M.J., Danila, D.C., Humm, J., Larson, S., Fleisher, M., Sawyers, C.L., Prostate Cancer Foundation/Department of Defense Prostate Cancer Clinical Trials Consortium, 2010. Antitumour activity of MDV3100 in castration-resistant prostate cancer: a phase 1-2 study. *Lancet (London, England)* 375, 1437–46. [https://doi.org/10.1016/S0140-6736\(10\)60172-9](https://doi.org/10.1016/S0140-6736(10)60172-9)
- Schwank, G., Koo, B.K., Sasselli, V., Dekkers, J.F., Heo, I., Demircan, T., Sasaki, N., Boymans, S., Cuppen, E., Van Der Ent, C.K., Nieuwenhuis, E.E.S., Beekman, J.M., Clevers, H., 2013. Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell* 13, 653–658. <https://doi.org/10.1016/j.stem.2013.11.002>
- 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. *Proc. Natl. Acad. Sci. U. S. A.* 101, 4758–63. <https://doi.org/10.1073/pnas.0401123101>
- Shen, H., Green, M.R., 2004. A pathway of sequential Arginine-Serine-Rich Domain-Splicing signal interactions during mammalian spliceosome assembly. *Mol. Cell* 16, 363–373. <https://doi.org/10.1016/j.molcel.2004.10.021>
- Shen, M.M., Abate-Shen, C., 2010. Molecular genetics of prostate cancer: New prospects for old challenges. *Genes Dev.* <https://doi.org/10.1101/gad.1965810>
- Shen, T., Li, H., Song, Y., Li, L., Lin, J., Wei, G., Ni, T., 2019. Alternative polyadenylation dependent function of splicing factor SRSF3 contributes to cellular senescence. *Aging (Albany, NY)*. 11, 1356–1388. <https://doi.org/10.18632/aging.101836>
- Siddiqui, Z.A., Krauss, D.J., 2018. Adjuvant androgen deprivation therapy for prostate cancer treated with radiation therapy. *Transl. Androl. Urol.* 7, 378–389. <https://doi.org/10.21037/tau.2018.01.06>

- Smits, A.M., 2019. Therapeutic gene editing, making a point. *Cardiovasc. Res.*  
<https://doi.org/10.1093/cvr/cvz038>
- Sontheimer, E.J., Barrangou, R., 2015. The Bacterial Origins of the CRISPR Genome-Editing Revolution. *Hum. Gene Ther.* <https://doi.org/10.1089/hum.2015.091>
- Sramkoski, R.M., Pretlow, T.G., Giaconia, J.M., Pretlow, T.P., Schwartz, S., Sy, M.S., Marengo, S.R., Rhim, J.S., Zhang, D., Jacobberger, J.W., 1999. A new human prostate carcinoma cell line, 22Rv1. *In Vitro Cell. Dev. Biol. Anim.* 35, 403–9.  
<https://doi.org/10.1007/s11626-999-0115-4>
- Steinestel, J., Luedeke, M., Arndt, A., Schnoeller, T.J., Lennerz, J.K., Wurm, C., Maier, C., Cronauer, M. V, Steinestel, K., Schrader, A.J., 2015. Detecting predictive androgen receptor modifications in circulating prostate cancer cells. *Oncotarget* 10, 4213–4223.  
<https://doi.org/10.18632/oncotarget.3925>
- 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. *Int. J. cancer* 100, 309–17. <https://doi.org/10.1002/ijc.10495>
- Sternberg, C.N., Petrylak, D.P., Madan, R.A., Parker, C., 2014. Progress in the treatment of advanced prostate cancer. *Am. Soc. Clin. Oncol. Educ. book. Am. Soc. Clin. Oncol. Annu. Meet.* 117–31. [https://doi.org/10.14694/EdBook\\_AM.2014.34.117](https://doi.org/10.14694/EdBook_AM.2014.34.117)
- Sternberg, S.H., Lafrance, B., Kaplan, M., Doudna, J.A., 2015. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* 527, 110–113.  
<https://doi.org/10.1038/nature15544>
- 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. *Sci. Rep.* 5, 13426.  
<https://doi.org/10.1038/srep13426>
- Su, Z., Wang, J., Yu, J., Huang, X., Gu, X., 2006. Evolution of alternative splicing after gene duplication. *Genome Res.* 16, 182–189. <https://doi.org/10.1101/gr.4197006>
- Sun, C., Shi, Y., Xu, L.L., Nageswararao, C., Davis, L.D., Segawa, T., Dobi, A., McLeod, D.G., Srivastava, S., 2006. Androgen receptor mutation (T877A) promotes prostate cancer cell growth and cell survival. *Oncogene* 25, 3905–3913.  
<https://doi.org/10.1038/sj.onc.1209424>
- 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. *J. Clin. Invest.* 120, 2715–2730.  
<https://doi.org/10.1172/JCI41824>
- Suvorova, E.S., Croken, M., Kratzer, S., Ting, L.M., de Felipe, M.C., Balu, B., Markillie, M.L., Weiss, L.M., Kim, K., White, M.W., 2013. Discovery of a Splicing Regulator Required for Cell Cycle Progression. *PLoS Genet.* 9. <https://doi.org/10.1371/journal.pgen.1003305>
- Suzman, D.L., Antonarakis, E.S., 2015. Does degree of androgen suppression matter in

- hormone-sensitive prostate cancer? *J. Clin. Oncol.* 33, 1098–1100.  
<https://doi.org/10.1200/JCO.2014.60.1419>
- Sveen, A., Kilpinen, S., Ruusulehto, A., Lothe, R.A., Skotheim, R.I., 2016. Aberrant RNA splicing in cancer; Expression changes and driver mutations of splicing factor genes. *Oncogene* 35, 2413–2427. <https://doi.org/10.1038/onc.2015.318>
- Tagawa, S.T., Antonarakis, E.S., Gjyrezi, A., Galletti, G., Kim, S., Worroll, D., Stewart, J., Zaher, A., Szatrowski, T.P., Ballman, K. V., Kita, K., Tasaki, S., Bai, Y., Portella, L., Kirby, B.J., Saad, F., Eisenberger, M.A., Nanus, D.M., Giannakakou, P., 2019. Expression of AR-V7 and ARV 567Es in circulating tumor cells correlates with outcomes to taxane therapy in men with metastatic prostate cancer treated in taxynergy. *Clin. Cancer Res.* 25, 1880–1888. <https://doi.org/10.1158/1078-0432.CCR-18-0320>
- Tan, M.H.E., Li, J., Xu, H.E., Melcher, K., Yong, E., 2015. Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacol. Sin.* 36, 3–23.  
<https://doi.org/10.1038/aps.2014.18>
- 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 antagonist. *Cancer Res.*
- 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 line. *Cancer Res.* 62, 6606–14.
- Thomas, C.E., Ehrhardt, A., Kay, M.A., 2003. Progress and problems with the use of viral vectors for gene therapy. *Nat. Rev. Genet.* <https://doi.org/10.1038/nrg1066>
- Tilgner, H., Knowles, D.G., Johnson, R., Davis, C.A., Chakraborty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T.R., Guigó, R., 2012. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* <https://doi.org/10.1101/gr.134445.111>
- Tran, C., Ouk, S., Clegg, N.J., Chen, Y., Watson, P.A., Arora, V., Wongvipat, J., Smith-Jones, P.M., Yoo, D., Kwon, A., Wasielewska, T., Welsbie, D., Chen, C.D., Higano, C.S., Beer, T.M., Hung, D.T., Scher, H.I., Jung, M.E., Sawyers, C.L., 2009. Development of a second-generation antiandrogen for treatment of advanced prostate cancer. *Science* 324, 787–90. <https://doi.org/10.1126/science.1168175>
- Tsai, M.J., O'Malley, B.W., 1994. Molecular mechanisms of action of steroid/thyroid receptor superfamily members. *Annu. Rev. Biochem.* 63, 451–86.  
<https://doi.org/10.1146/annurev.bi.63.070194.002315>
- Uo, T., Dvinge, H., Sprenger, C.C., Bradley, R.K., Nelson, P.S., Plymate, S.R., 2017. Systematic and functional characterization of novel androgen receptor variants arising from alternative splicing in the ligand-binding domain. *Oncogene* 36, 1440–1450.  
<https://doi.org/10.1038/onc.2016.313>
- 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 Res.* 77, 5228–5235. <https://doi.org/10.1158/0008-5472.CAN-17-0320>
- Veldscholte, J., Ris-Stalpers, C., Kuiper, G.G., Jenster, G., Berrevoets, C., Claassen, E., van Rooij, H.C., 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. *Biochem. Biophys. Res. Commun.* 173, 534–40. [https://doi.org/10.1016/s0006-291x\(05\)80067-1](https://doi.org/10.1016/s0006-291x(05)80067-1)
- Villiger, L., Grisch-Chan, H.M., Lindsay, H., Ringnalda, F., Pogliano, C.B., Allegri, G., Fingerhut, R., Häberle, J., Matos, J., Robinson, M.D., Thöny, B., Schwank, G., 2018. Treatment of a metabolic liver disease by in vivo genome base editing in adult mice. *Nat. Med.* 24, 1519–1525. <https://doi.org/10.1038/s41591-018-0209-1>
- 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. *Nat. Genet.* 9, 401–406. <https://doi.org/10.1038/ng0495-401>
- Wang, K., Tang, X., Liu, Y., Xie, Z., Zou, X., Li, M., Yuan, H., Ouyang, H., Jiao, H., Pang, D., 2016. Efficient Generation of Orthologous Point Mutations in Pigs via CRISPR-assisted ssODN-mediated Homology-directed Repair. *Mol. Ther. - Nucleic Acids* 5, e396. <https://doi.org/10.1038/mtna.2016.101>
- Wang, Z., Burge, C.B., 2008. Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA*. <https://doi.org/10.1261/rna.876308>
- Watson, P.A., Chen, Y.F., Balbas, M.D., Wongvipat, J., Socci, 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. *Proc. Natl. Acad. Sci. U. S. A.* 107, 16759–65. <https://doi.org/10.1073/pnas.1012443107>
- Will, C.L., Lührmann, R., 2011. Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* 3, 1–2. <https://doi.org/10.1101/cshperspect.a003707>
- Williams, T., 2005. 321. Splice-Switching of the Prostate Specific Membrane Antigen (PSMA) in LnCap Prostate Cancer Cells. *Mol. Ther.* 11, S125. <https://doi.org/10.1016/j.ymthe.2005.06.324>
- Wu, D., Sunkel, B., Chen, Z., Liu, X., Ye, Z., Li, Q., Grenade, C., Ke, J., Zhang, C., Chen, H., Nephew, K.P., Huang, T.H.-M., Liu, Z., Jin, V.X., Wang, Q., 2014. Three-tiered role of the pioneer factor GATA2 in promoting androgen-dependent gene expression in prostate cancer. *Nucleic Acids Res.* 42, 3607–22. <https://doi.org/10.1093/nar/gkt1382>
- Xiang, X., Li, C., Chen, X., Dou, H., Li, Y., Zhang, X., Luo, Y., 2019. CRISPR/Cas9-Mediated Gene Tagging: A Step-by-Step Protocol, in: *Methods in Molecular Biology*. Humana Press Inc., pp. 255–269. [https://doi.org/10.1007/978-1-4939-9170-9\\_16](https://doi.org/10.1007/978-1-4939-9170-9_16)
- 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 Res.* 65, 9611–6. <https://doi.org/10.1158/0008-5472.CAN-05-0817>

- Zhang, X.H., Tee, L.Y., Wang, X.G., Huang, Q.S., Yang, S.H., 2015. Off-target effects in CRISPR/Cas9-mediated genome engineering. *Mol. Ther. - Nucleic Acids*. <https://doi.org/10.1038/mtna.2015.37>
- Zhao, X.Y., Malloy, P.J., Krishnan, A. V, Swami, S., Navone, N.M., Peehl, D.M., Feldman, D., 2000. Glucocorticoids can promote androgen-independent growth of prostate cancer cells through a mutated androgen receptor. *Nat. Med.* 6, 703–6. <https://doi.org/10.1038/76287>
- Zhou, X., Yang, X., Sun, X., Xu, X., Li, Xi'an, Guo, Y., Wang, J., Li, Xia, Yao, L., Wang, H., Shen, L., 2019. Effect of PTEN loss on metabolic reprogramming in prostate cancer cells. *Oncol. Lett.* 17, 2856–2866. <https://doi.org/10.3892/ol.2019.9932>

## Online sources

Cancer Research UK, 2014-2016. Link: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer/incidence#heading-One>. Last updated: 20/06/19.

Cancer Research UK, 2016. Link: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer/incidence#heading-Zero>. Last updated: 20/06/19.

COSMIC, 2019. Link: <https://cancer.sanger.ac.uk/cosmic/search?q=androgen+receptor>. Last updated: 14/08/19.

FDA, 2012. Link: <https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-enzalutamide-castration-resistant-prostate-cancer>. Last updated: 16/07/18.