Androgen regulation controlled *NOVA1* gene expression drives alternative gene splicing in prostate cancer.

Rahul Sagar Advani

Biosciences Institute

A thesis submission for degree of Doctor of Philosophy

September 2024

Abstract

Prostate cancer is a genetic disease that is known to be regulated by androgen hormones that primarily arises in epithelial cells of the prostate gland. While epithelial splicing regulatory proteins (ESRP1 and ESRP2) have been established to be key players in prostate cancer progression, particularly through their androgen-dependent regulation leading to alternative splicing events, the role of mesenchymal – specific RNA binding proteins has not largely been explored. This study focuses on *NOVA1* gene, a mesenchymal RNA-binding protein, to determine its role in prostate cancer. The research objectives included assessing NOVA1 expression in various prostate cancer cell lines, analysing NOVA1 protein regulation under different androgen exposure conditions, and identifying cell lines with highest NOVA1 expression (PC3) and lowest NOVA1 expression (LNCaP) to understand its association with different prostate cancer phenotypes, particularly aggressive vs benign. Using a combination of lab techniques, I quantified NOVA1 levels and investigated its regulation mechanism. Additionally, RNA sequencing (RNAseq) was carried out following NOVA1 knockdown in PC3 cells to evaluate its impact on ESRP1 and ESRP2 target splicing events, with subsequent bioinformatic analysis approaches. My findings revealed that NOVA1 is inversely regulated by androgens in comparison to ESRP1 and ESRP2, with its expression significantly elevated in the highly aggressive PC3 prostate cancer cell line. Further a study by Zhang et al. in 2020, found high NOVA1 expression in neuroendocrine prostate cancer cell lines, known for their aggressive nature. NOVA1's influence on alternative splicing is diverse, mirroring ESRP1 and ESRP2's effects on certain events while exerting opposing effects on other events. NOVA1 has been predicted to have mesenchymal properties that contribute to cell mobility and tumour metastasis. These properties are supported by the differential expression analysis where several of the NOVA1 regulated genes associated with pathways such as EMT, G2M, and E2F signifying NOVA1 role in cell mobility, cell cycle regulation, and metastasis. In conclusion, NOVA1 emerges as a pivotal gene in prostate cancer, highlighting the critical role of mesenchymal RNA-binding proteins in the disease development. The regulatory connect between NOVA1 and ESRP1 and ESRP2 underscores the complexity of androgen-driven alternative splicing in prostate cancer. These findings point to NOVA1 as a potential target for therapeutic intervention, particularly in combating aggressive forms of the disease.

Impact of Covid 19

The onset of the COVID-19 pandemic in early 2020 had a profound impact on both my personal and academic life, shaping the trajectory of my PhD journey in ways that were unforeseen and challenging. I began my PhD program in December 2019, full of anticipation and excitement for the research that lay ahead. However, just a few months later, in March 2020, the world was hit by the COVID-19 pandemic, which led to the closure of universities for approximately four months, and I was unable to return for about 2 more months due to flight unavailability due to decreased operational capacity during Covid-19. This abrupt closure meant that I was no longer able to access the laboratory, a crucial component of my research work. The inability to conduct lab work significantly hindered the progress of my project, leaving me in a state of uncertainty and frustration.

During this period of disruption, I also discovered that I had been enrolled in the wrong course, a mistake that necessitated a course correction. This error required me to reapply for a visa, a process that was severely complicated by the pandemic. With embassies either closed or operating at minimal capacity, obtaining a visa became an arduous and time-consuming task, further delaying my return to the lab by approximately 5 months. The compounded effects of these delays meant that a considerable amount of lab work was either slowed down or could not be carried out at all.

When I was finally able to return to the lab, the working environment had changed drastically. Social distancing guidelines meant that lab space was limited, and the number of experiments that could be conducted simultaneously was significantly reduced. This limitation further slowed the pace of my research. A particularly challenging moment came when I successfully performed a knockdown experiment and needed to send the samples for RNA sequencing. The sequencing facility, having just reopened after a prolonged closure, was overwhelmed with a backlog of samples. As a result, there was a significant delay in processing my samples, which further pushed back the timelines of my research.

During the periods when I was unable to work on my main PhD project, I focused on writing a review paper. While this allowed me to remain productive, it was not a substitute for the experimental work that was essential to my thesis.

These delays culminated in the extension of my thesis submission deadline, as it became clear that the cumulative effects of the pandemic had significantly set back my research timeline. In total, I estimate that nearly 12 months of my PhD were lost to the inability to work in the lab, and another 12 months were spent working at only 25% to 50% of normal capacity. Despite these challenges, I am now able to submit my thesis. However, there is still some work that I would like to see carried forward, either by future candidates or through a potential post-doctoral position in the same research area.

This period has been incredibly stressful, and the extended duration of my PhD has tested my resilience in ways I never anticipated. Nevertheless, I have learned to adapt and persevere in the face of unprecedented challenges, and I am proud of the work I have been able to accomplish despite the numerous obstacles presented by the COVID-19 pandemic.

Declaration

I, Rahul Sagar Advani, declare that no portion of the work compiled in this thesis has been submitted in support of another degree or qualification at this or any other University or Institute of Learning. This thesis includes nothing which is the work of others, nor the outcomes of work done in collaboration, except where otherwise stated.

Rahul Sagar Advani

September 2024

Acknowledgements

I would like to extend my deepest gratitude to my supervisors, Professor David Elliott and Dr. Jennifer Munkley. David's kindness, patience, and unwavering support have been instrumental throughout my PhD journey. His resilience and extensive guidance, combined with his vast knowledge, have been invaluable. Despite my numerous questions and the delays in my work, often caused by visa issues and the COVID-19 pandemic, David remained a beacon of positivity and encouragement, always greeting me with a smile. His dedication and commitment to my success made even the toughest moments feel manageable, transforming this challenging journey into a remarkably smooth experience. I often wonder where I might have ended up without his steadfast support, particularly during the demanding process of writing the review paper. Alongside David, I am profoundly grateful to Dr. Jennifer Munkley for her immense kindness and the valuable guidance she provided. Jennifer's willingness to offer her expertise and, where needed, the materials essential for my research, ensured that I could carry on with my thesis without hindrance. Together, David and Jennifer have been the epitome of what one could hope for in supervisors, making this journey not only possible but also enriching and fulfilling.

I would also like to acknowledge the wonderful people in the lab who made this tough journey of a PhD manageable and even enjoyable. Dr. Gerald Hysenaj provided initial guidance and patiently taught me the lab protocols and techniques that became the foundation of my research. Dr. Saad Aldalaqan and Kayla Bastian, who joined alongside me, have been invaluable companions. Saad, in particular, was the friend I much needed at work and my go-to for PCR-related queries. Dr. Chile Siachisumo and Dr. Alice were the seniors everyone wishes for—Chile's infectious joy made the lab a brighter place. I'm also grateful to Dr. Caroline Dalgliesh for always being willing to answer my quick, and sometimes silly, questions with a smile. Dr. Ingrid Ehrmann, from the initial fire safety training to providing essential lab reagents and helping with those arduous calculations, your support has been greatly appreciated. Dr. Emma Scott's guidance through TCGA and bio-portal, and her assistance with datasets, was crucial for my research. Farimah and Irene, your addition to the lab brought joy and camaraderie that made work much lighter. Lastly, but most importantly, I want to thank

Sarah, without whom this PhD thesis would not have been possible. Sarah's expertise in bioinformatics analysis was a lifesaver, and she remains my go-to person for all bioinformatic questions and understanding complex results. Beyond the work, our discussions about Indian dishes, wedding dresses, recipes, and even Costco discounts on Indian rotis have been incredibly enriching and fun. Thank you, Sarah, for your unwavering support and friendship.

Additionally, the journey of my PhD has been marked by significant life changes. I got married in 2021 and became a father in 2024, both life events that brought immense joy and challenges. My deepest gratitude goes to my wife, Shilpa Advani, whose unwavering support and sacrifice made this possible. Living apart for two long years, she endured and persevered, embodying resilience and strength. Her patience and motivation were the driving forces that kept me going through the best and worst times. I also owe a debt of gratitude to my mother, Dr. Dayali Gul Sindhu Advani, who inspired me to pursue this PhD. It was her dream that I sought to fulfill, and she, along with my father, Dr. Raj Kumar Advani, bore the financial burdens with grace. My mother’s unwavering belief in me and my father’s excitement as the finish line nears have been sources of strength. To my siblings (Priyanka, Sahil, and Suraj), who have eagerly awaited the day I graduate, and to my father-in-law, Mr. Ramesh Lal, whose prayers have been a constant support, I extend my heartfelt thanks. My mother-in-law, Mrs. Asha Rameshlal, ensured I was well-nourished, sending love and food from afar, easing my burdens. My sisters-in-law (Sonia, Annie, Priya, Sindhu, and Jaya) and brothers-in-law (Vikash, Pawan, Vinaish, Kunal, and Yougesh), your excitement for my success has meant the world to me.

Lastly, I dedicate this thesis to my son, Armaan Advani. Though he is only a few months old, he has already filled my life with immeasurable joy and inspiration. His presence has fueled my determination to complete this work, and he remains a bright light guiding me forward.

Table of Contents

List of Figures and Tables

List of Abbreviations

3'SS	3' Splice Site
5'SS	5' Splice Site
ADT	Androgen Deprivation Therapy
AR	Androgen Receptor
ARE	Androgen Receptor Element
ATCC	American Type Culture Collection
BPH	Benign Prostate Hyperplasia
BPS	branch point sequence
BSA	Bovine Serum Albumin
cDNA	Complementary DNA
DBD	DNA Binding Domain
DHT	Dihydrotestosterone
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic acid
DRE	Digital Rectal Exam
LBD	Ligand Binding Domain
LHRH	Luteinizing Hormone Releasing Hormone
NCCN	National Comprehensive Cancer Network
NLS	Nuclear Localization Signal
NTD	N-terminal Domain
PCa	Prostate Cancer
PCA3	Progensa Prostate cancer gene 3
PHI	Prostate Health Index
PSA	Prostate Specific Antigen
SD	Steroid Deplete

Chapter 1: Introduction

1.1 Cancer:

The term "cancer" refers to a collection of diseases characterized by the uncontrolled growth of cells. This unregulated cell growth is typically a result of malfunctioning cell cycle regulation genes, which fail to identify and eliminate cells with damaged DNA through various mechanisms. Consequently, these cancerous cells have the potential to spread and develop secondary tumours through the blood or lymphatic system.[1,2] Cancer can be categorized into two types: benign tumours and malignant tumours.[3] Benign tumours are characterized by their ability to remain confined and localized as a lump without invading nearby tissues or spreading to other body parts.[4] However, it is essential to note that benign tumours can transform into malignant ones in certain instances.[5] On the other hand, malignant tumours are known for their invasive nature, as they infiltrate surrounding tissues and have the potential to metastasize, spreading throughout the body.[6]

Cancer, a disease that has plagued humanity for centuries, has been a significant cause of death throughout history.[7] Evidence of cancer can be traced back to approximately 1500 BC, and examination of specimens suggests that cancerous lesions may have existed even before written records.[7,8] The term "carcinoma," as it is known today, was introduced by Hippocrates (460-370 B.C.) as "karkinos." [8] However, it wasn't until the emergence of genetics and the implementation of whole genome sequencing (WGS) that a better understanding of cancer began to take shape. The introduction of next-generation sequencing (NGS) and the extensive sequencing of human genomes have further contributed to our comprehension of the underlying genetic mutations and mechanisms that drive the development of carcinomas.[7–9]

Despite numerous advancements in cancer diagnostics and treatment, the disease continues to prevail worldwide, with its incidence steadily increasing each year. It is predicted that cancer-related deaths will soon surpass those caused by cardiovascular diseases.[10,11] This conclusion is supported by studies indicating that countries with higher incomes, particularly upper-middle-income nations, experience higher rates of cancer mortality compared to deaths resulting from cardiovascular diseases.[10,11] According to the International Agency for

Research on Cancer, in 2018, Europe alone recorded approximately 4.23 million diagnosed cases of cancer, resulting in an estimated 2 million deaths.[12] Furthermore, there has been a notable rise in the incidence of major cancers such as lung, breast, prostate, colon and rectal cancers.[9,13,14]

1.2 Types of cancer

Cancer can be classified based on its site of origin, such as lung cancer, liver cancer, and prostate cancer, which originate in the lung, liver, and prostate, respectively. Another classification criterion is the tissue type, which is divided into six different categories as defined by the International Classification of Diseases for Oncology.[15,16] These categories include Carcinoma, Sarcoma, Myeloma, Leukaemia, Lymphoma, and Mixed types. Carcinomas are malignant cancers that develop from the epithelial layer lining the external parts of the body, such as the skin. Conversely, Sarcomas arise from connective and supportive tissues, such as muscles, bones, and fat. Myeloma is cancer originating in the bone marrow's plasma cells. Leukaemia is commonly known as blood cancer and affects the bone marrow. Lymphomas are a type of cancer that affects the lymphatic system. Finally, mixed types of cancers exhibit features from more than one category.[15,16]

In addition, it is worth noting that benign tumours are characterized by the suffix "-oma," while malignant tumours are identified by the suffix "carcinoma." For instance, an adenoma is a benign tumour that originates in the epithelial tissue of a gland. In contrast, adenocarcinoma is a malignant tumour with the same origin as adenoma.[15,16]

1.3 Tumour classification

The identification, diagnosis, and treatment of cancers rely on cancer staging, a crucial process that helps determine the extent of cancer progression and available treatment options. This staging is achieved through the implementation of the TNM classification system.[17]

The T in TNM represents the tumour size, ranging from 0 to 4. A higher number indicates a larger tumour that has invaded the surrounding tissue, while 0 signifies a small tumour that does not affect the surrounding area.[17] Next, the N represents lymph nodes and is assigned a number from 0 to 3. The higher the number, the more lymph nodes are invaded by cancer. Lastly, the M indicates metastasis and has values of 0 or 1. A value of 0 indicates no metastasis, while a value of 1 signifies that the tumour has spread to other body parts.[17] Based on this TNM classification, the cancer stage ranges from 1 to 4. Stage 4 cancer typically carries a poor prognosis, while stage 1 cancers are characterized by small size, localized growth, and no tissue invasion or metastasis.[17]

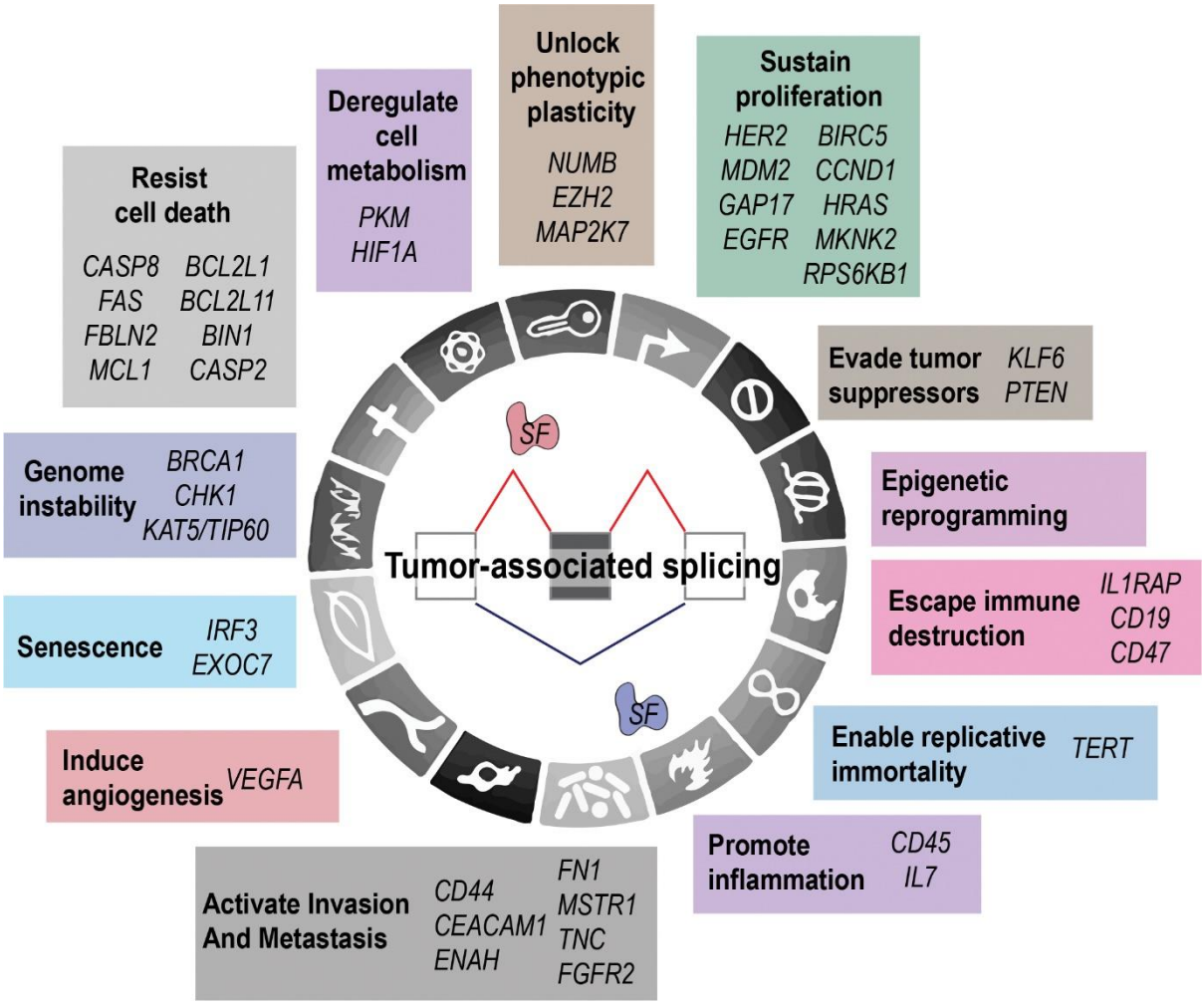


Figure 1.1: Hallmarks of cancer. *The figure shows hallmarks of cancer and related gene splice isoforms. The figure is taken from Bradley el al. 2023.*[20]

1.4 Hallmarks of cancer

Cancer is characterized by the accumulation of transformed cells that have lost their ability to regulate cell growth. Through extensive research, several hallmarks of cancer have been identified Figure 1.1.[18–20] These hallmarks can be described as acquired, evolutionary characteristics that facilitate the transformation of cells from a phenotypically normal state to a malignant one.[21] These hallmarks encompass various changes observed during the transformation of normal cells into cancer cells. They include alterations in cell morphology, the loss of contact inhibition, loss of anchorage dependence, tolerance of high saturation number, reduced reliance on growth factors, immortality through the loss of Hayflick's limit, evasion of telomere shortening, resistance to apoptosis, and increased glucose transport.[18,19,21,22] In summary, these hallmarks collectively outline the fundamental changes in cells as they progress from a normal phenotype to cancerous cells.

Tumour suppressor genes prevent cells with damaged or mutated DNA from progressing through the cell cycle and triggering cell death.[23] The proper functioning of these genes can be disrupted by mutations that affect both copies of the alleles.[23,24] This phenomenon was investigated by Knudson, who observed that individuals with an inherited mutation in the *Retinoblastoma* gene developed retinoblastoma at an earlier age than cases where individuals started with two unaffected copies of the gene. This finding confirmed the significance of "two hits" in the development of tumours, leading to what is known as Knudson's two-hit hypothesis.[23,24] Moreover, this underscores the importance of gene function in cancer occurrence. In addition, substances that contribute to cancer can be classified as initiators or promoters. Initiators are carcinogenic, and a single exposure to them can result in an irreversible mutation. On the other hand, promoters do not cause mutations themselves; instead, they play a role in stimulating cell proliferation. Applying promoters alone does not lead to cancer. However, when an initiator causes a mutation through initial exposure, subsequent application of a promoter triggers cell proliferation, leading to the formation of a tumour. This tumour can be either malignant or can become malignant upon further exposure to an initiator.[25]

1.5 The Prostate

The male reproductive system includes the prostate, an exocrine gland roughly the size and shape of a walnut.[26,27] Positioned anatomically between the bladder and penis shown in

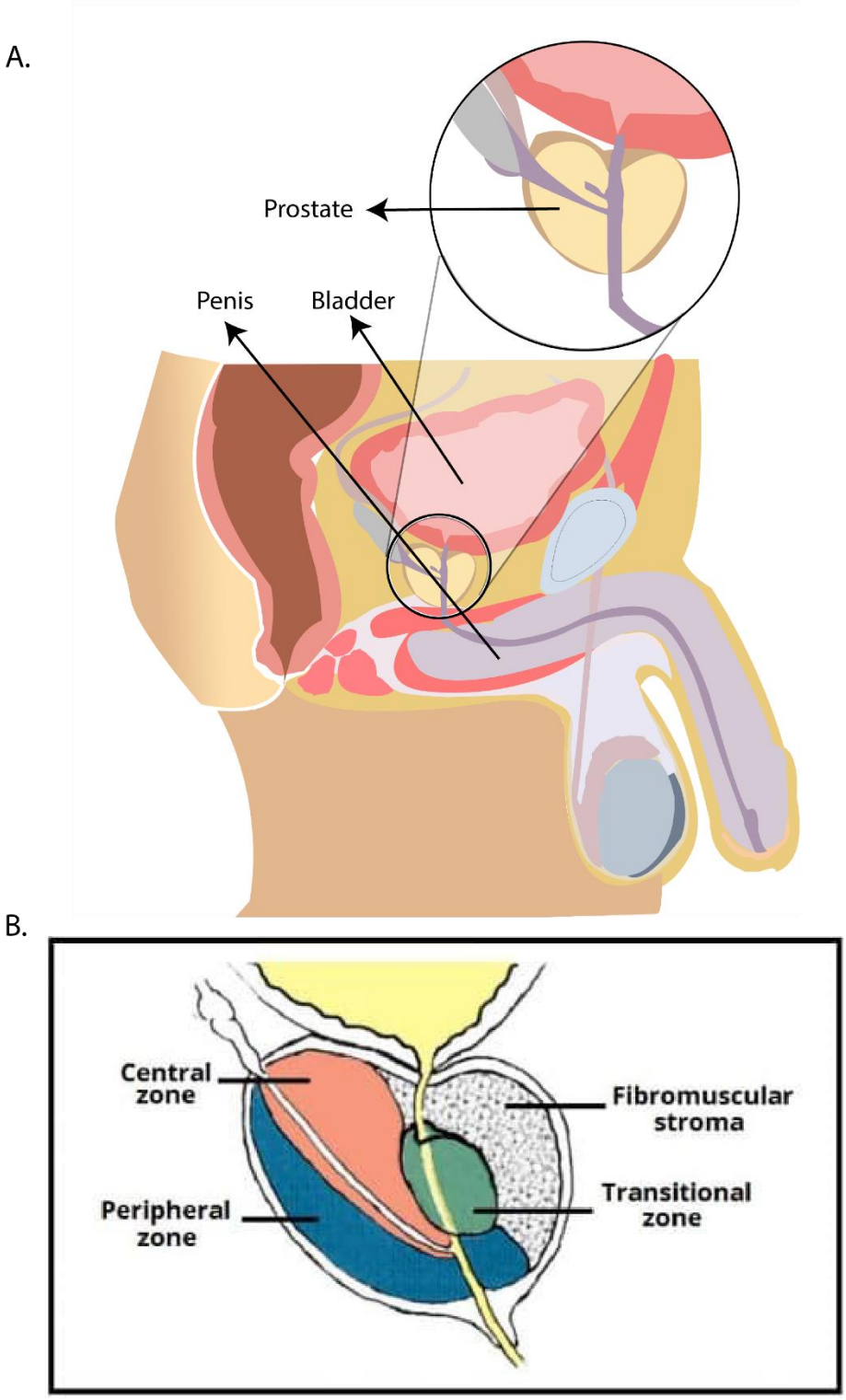


Figure 1.2: Anatomical location of Prostate gland. *A. The figure shows that walnut sized, prostate gland is located between bladder and penis. B. Shows the zones of prostate. Figure 1.2B is taken from https://teachmeanatomy.info.*[33]

Figure 1.2A. The prostate plays a crucial role in reproductive function[28,29]. It secretes fluid that nourishes sperm, enhancing their motility and prolonging their survival[28,30]. This fluid contains various biological components, including PSA (Prostate Specific Antigen), which serves as a biological marker for the detection of prostate cancer.

The current anatomical nomenclature for the prostate gland, as established by John McNeal, has brought clarity to its structure.[27,31] Initially, Lowsley's description of several prostate lobes posed challenges in their anatomical distinction among adult humans.[27,31,32] However, these lobes were later categorized into five specific lobes: middle, posterior, ventral, and two lateral lobes. John McNeal's nomenclature introduced a revised classification, dividing the human prostate into three distinct zones: the central zone, peripheral zone, and transition zone as shown in Figure 1.2B.[33] McNeal's central zone includes Lowsley's middle lobe and a portion of the posterior lobe. The peripheral zone comprises the remaining part of the posterior lobe, along with the two lateral lobes as described by Lowsley.[27,31,32] Additionally, despite the anatomical differences between the human and mouse prostate, it is noteworthy that approximately 80% of human prostate cancer cases manifest in the peripheral zone, which bears resemblance to the dorsolateral lobe in mice.[34,35] Conversely, the central zone demonstrates the lowest incidence of the disease.[27] Consequently, mouse models serve as valuable tools for investigating prostate cancer in men, given their similarity to the peripheral zone.

Figure 1.3 shows the key cellular components of the prostate gland involved in cell physiological and pathological functions. The normal prostate gland predominantly comprises epithelial cells that play an important role in glandular function maintenance. These epithelial cells are organized into two types: basal cells, which are situated on the periphery of the glandular acini, and luminal cells, which line the lumen of the acini. Luminal cells are responsible for the secretion of prostatic fluid, which is crucial in sperm motility. The basement membrane supports the epithelial cells, providing structural integrity and cellular interaction with other cell types, including neuroendocrine cells that modulate localized signalling within the gland. It is important to note that prostate cancer originates in these

epithelial cells. Initially, it begins as prostatic intraepithelial neoplasia within basal or glandular cells and then progresses to become a carcinoma of the prostate.[36]

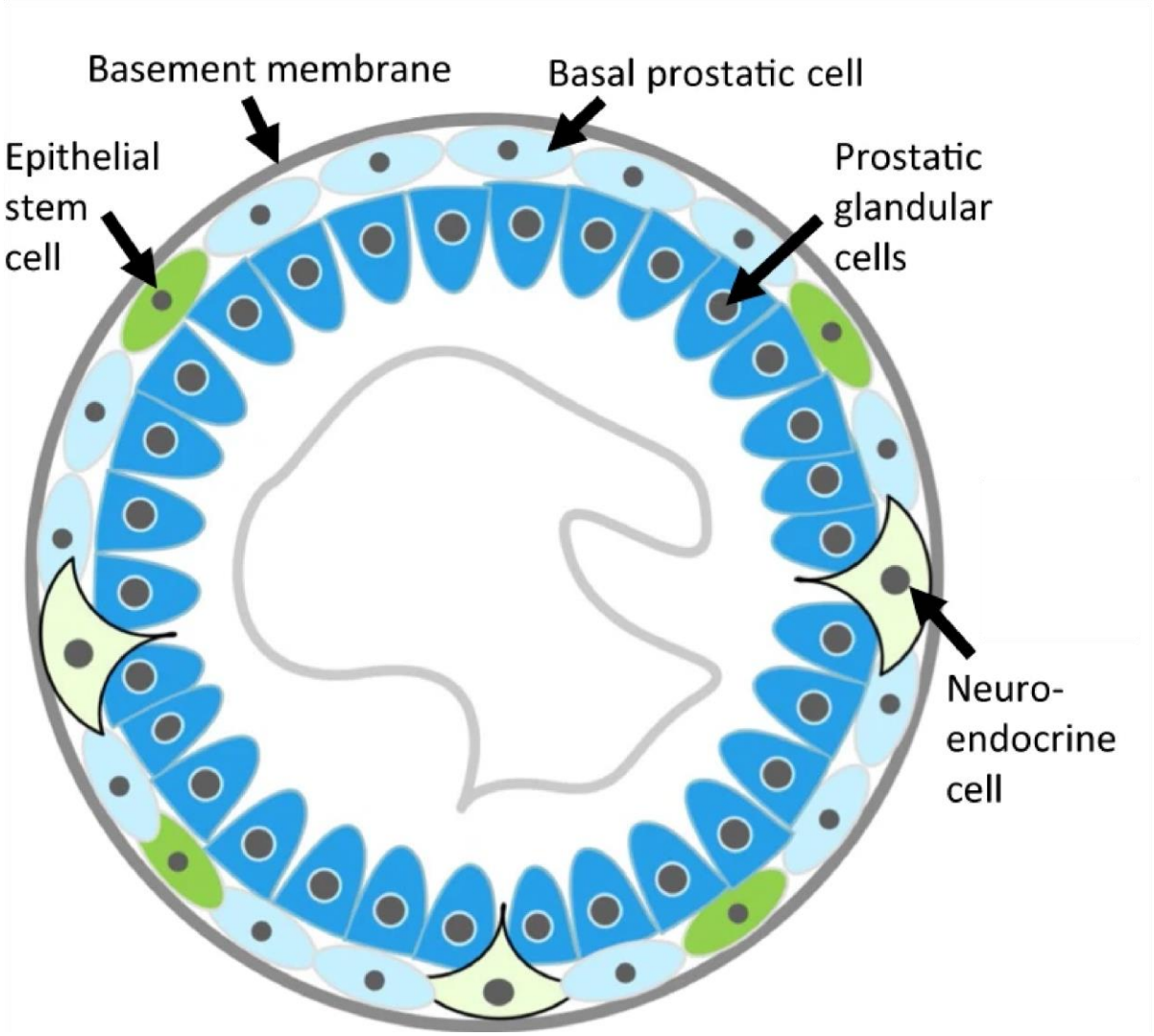


Figure 1.3: Prostate gland cellular structure. *The figure shows basement membrane on the outside. In the figure green cells show the epithelial stem cells. Additionally, Neuroendocrine, basal prostatic ad prostatic glandular cells can be seen. Picture taken from Advani et al.[36]*

1.6 Prostate Cancer

The prostate gland has become significantly important in medicine due to the notable increase in the prevalence of prostate cancer in Western countries. Among all cancers, prostate cancer is the second leading cause of death among men in United States.[37] Additionally, the prostate gland is associated with other conditions such as Benign Prostate Hyperplasia (BPH), characterized by prostate enlargement due to increase in number of cells,

and Prostatitis, which refers to inflammation of the prostate gland.[27] In a study published in 2015, prostate cancer was identified as the most prevalent cancer among men worldwide, with approximately 1.1 million reported cases and 307,500 annual deaths in 2012.[34,38] Despite advancements in prostate cancer diagnostics, data published by the International Agency for Cancer Research (IARC) in 2018 revealed a further increase, with approximately 1.3 million new cases and 358,989 reported deaths globally.[39] The statistics from 2020 have seen further increase in prostate cancer cases with incidence reported as 1.4 million and 375,304 deaths reported globally.[40,41] In UK approximately 55,100 new cases are diagnosed annually and about 12,000 prostate cancer deaths are reported. [42]

Given the rising incidence and burden of prostate cancer, it is crucial to delve into the molecular and genetic mechanisms that drive the disease. This understanding is essential for the development of new and effective diagnostic methods, enabling early detection when treatments are most impactful. Additionally, efforts should focus on developing treatments to combat the disease in advanced stages.

1.7 Molecular Mechanisms of prostate cancer

Androgens play a pivotal role in the progression of prostate cancer by activating the androgen receptor (AR), a transcription factor that drives the expression of genes critical for tumour growth. Locally advanced and metastatic prostate cancer is commonly treated with ADT, which initially proves effective in reducing tumour size. ADT involves the use of various drugs, including LHRH antagonists that inhibit the hypothalamic-pituitary-gonadal-axis to reduce testosterone production, abiraterone that disrupts enzymes essential for androgen biosynthesis, and AR antagonists like bicalutamide and enzalutamide that block the ligand-dependent activity of AR. Despite the initial success of ADT in shrinking the prostate tumour, most patients eventually experience a relapse, with the cancer becoming more aggressive and lethal. Additionally, prostate cancers often harbour a genetic fusion between *TMPRSS2* and *ERG* genes, which results in the *ERG* proto-oncogene being placed under the control of androgens downstream of the *TMPRSS2* promoter region.[36]

1.8 The Androgen Receptor

The androgen receptor (AR) is a protein that plays a central role in the regulation of male sexual development, as well as the growth and function of the prostate gland[43]. It belongs to the nuclear receptor superfamily and functions as a ligand-activated transcription factor, controlling the expression of target genes in response to androgen hormones, particularly testosterone and its more biologically active form dihydrotestosterone (DHT)[43].

Structure and Function of the Androgen Receptor

The AR is composed of four functional domains that arise from 8 canonical exons of the *AR* gene shown in Figure 1.4.[44] The N-terminal domain (NTD) is encoded by exon 1 of the *AR* gene and is connected to DNA-binding domain (DBD)[43,44]. DBD is encoded by *AR* exons 2 and 3.[44] It is located centrally and is responsible for binding to specific DNA sequences known as androgen response elements (AREs) within target gene promoters[43]. The DBD contains two zinc fingers that facilitate its interaction with DNA[43,45,46].

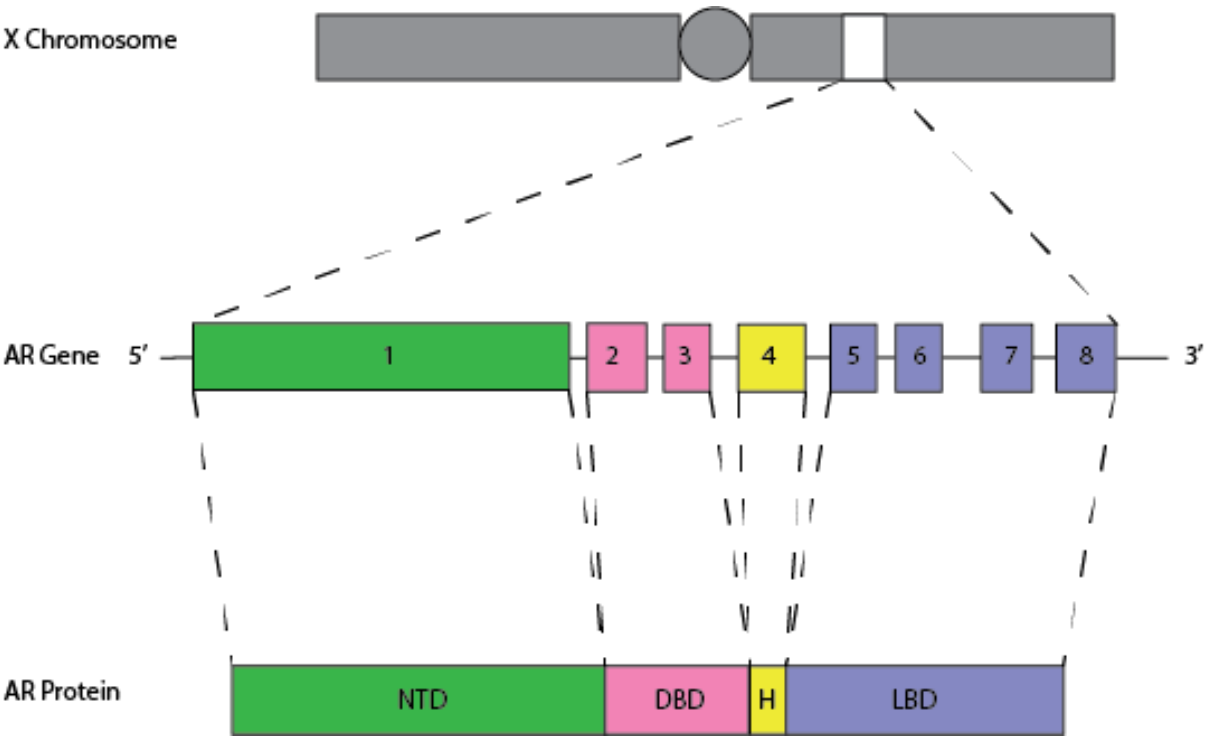


Figure 1.4: Formation of AR protein. *AR protein coding gene is present on X chromosome shown at the top of the figure. AR gene consists of 8 exons that are translated into AR protein.*

The hinge region connects the DBD to the ligand-binding domain (LBD)[43]. The hinge region of the AR is encoded by exon 4 of the *AR* gene[44]. This AR region contains a nuclear localization signal (NLS) that mediates the translocation of the AR into the nucleus upon ligand binding. The LBD is the biggest domain encoded by exons 5-8[44]. Furthermore, the LBD is responsible for ligand binding. Ligand binding induces a conformational change that allows the AR to

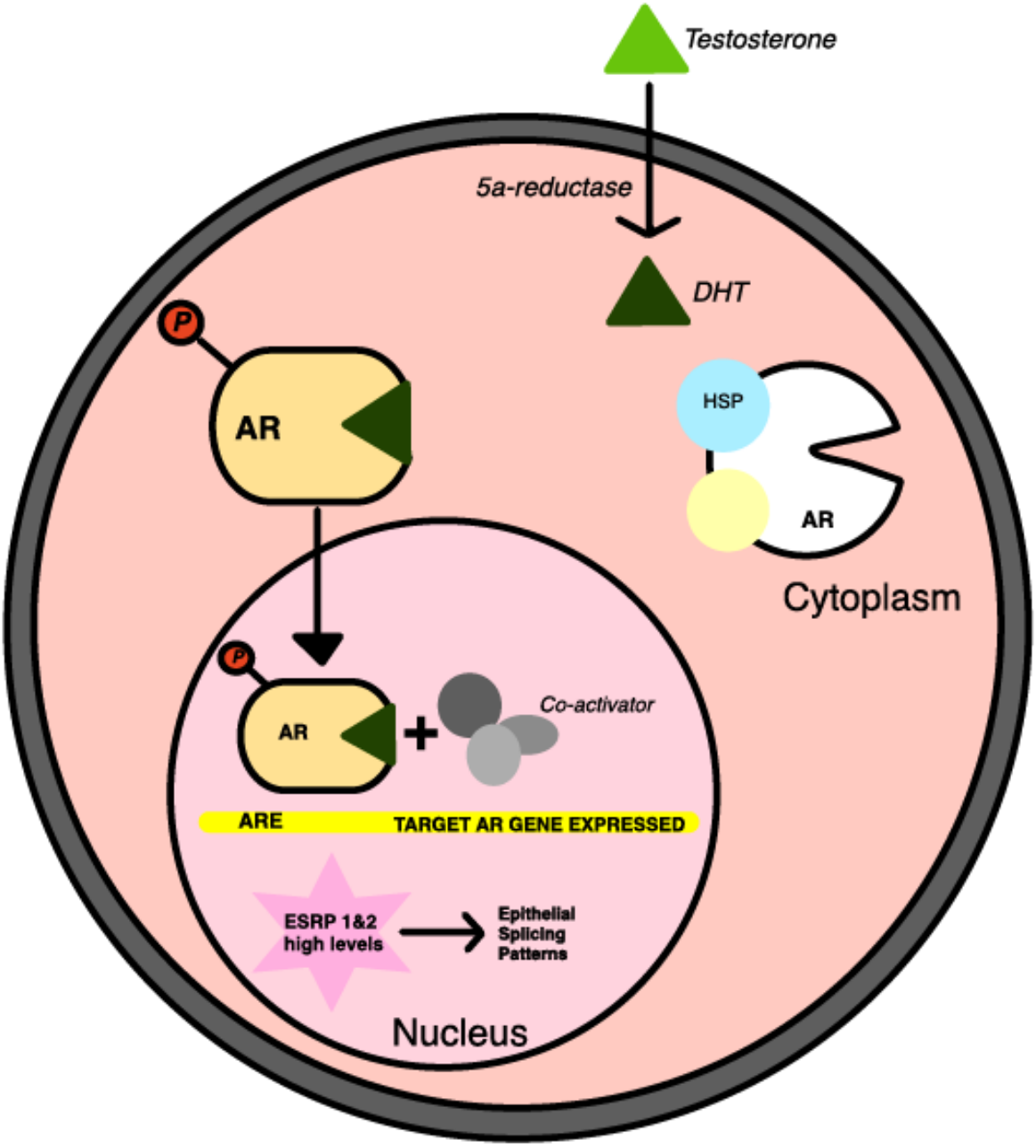


Figure 1.5: Mechanism of AR-ARE interaction. *Testosterone is converted to DHT that binds AR in cytoplasm and brings the conformational change. The AR move into nucleus and with co-activators binds to ARE and carry out its effects on target gene.*

dissociate from heat shock proteins, bind to DHT, and translocate to the nucleus, where it binds to AREs and regulates gene transcription Figure 1.5[43].

Upon ligand binding and translocation to the nucleus Figure 1.5, the AR interacts with co-regulators that can be co-activators or co-repressors of transcription, depending on whether target genes are activated or repressed [43]. The AR protein regulates the expression of genes involved in various biological processes, including prostate development. Munkley et al. 2016 studied the genes changing in response to androgen stimulation compared to androgen depletion in RNA sequencing dataset obtained from patient with prostate cancer.[47] The same study further found 700 genes that displayed expression changes in response to androgens and these were identified to be linked with the glycosylation.[47] Among the genes that showed reciprocal changes to androgen were two genes, *ESRP1* and *ESRP2* genes.[47] These genes have been followed up since then in Munkley et al. 2019 study.[48]

1.9 Biomarkers for Prostate Cancer Detection

Biomarkers play a crucial role in the detection and management of prostate cancer. Mainly, the prostate-specific antigen (PSA) test has been widely used as a blood-based biomarker for prostate cancer diagnosis. However, PSA has limitations in terms of specificity and sensitivity. Elevated PSA levels can be indicative of prostate cancer, but they can also be caused by other non-cancerous conditions such as prostatitis and BPH (Benign Prostate Hyperplasia).

Urine-based biomarkers, such as *PCA3* (Progensa Prostate cancer gene 3), 502083-IFU-PI_001.pdf (hologic.com), show increased expression from men with cancerous compared to benign prostate tissue following digital rectal examination.[49,50] ExoDx (Exosome Dx, Urine Test for Prostate Cancer | ExoDx Prostate Test (exosomedx.com)), also known as IntelliScore, is a non-invasive test that can detect prostate cancer at grade 2 or higher.[49,51] ExoDx calculates the risk of having clinically significant prostate cancer based on three RNA biomarkers, namely *ERG, PCA3, and SPDEF*. The score range is from 0 to 100, where a score of 15.6 or higher indicates a risk of clinically significant prostate cancer. Select MDx (What is the Select MDx test for prostate? (pcmarkers.com)) looks for the expression levels of two genes, *HOXC6 and DLX1*. These gene expression levels along with other factors, such as age, PSA levels, and DRE

(Digital Rectal Exam), are used to generate a risk score for prostate cancer. Select MDx generates an AUC (area under the curve) of 0.90 to identify a high-grade Gleason score of 7 or above PCa.[49,52–54] MiPS (Michigan Prostate Score), now known as MPS (MyProstateScore, MyProstateScore™ (MPS) 2.0 | MLabs (umich.edu)), combines measurements of TMPRSS2:ERG and PCA3 in liquid biopsy (urine sample collected post-digital rectal exam). MyProstateScore predicts the risk of prostate cancer with a Gleason score of 7 or above and validates the need for biopsy, as with the current PSA test, there is an unnecessary excess of biopsies performed.[49,55,56] These tests have shown promise in improving the specificity and sensitivity of prostate cancer detection. These biomarkers can differentiate between benign and malignant prostate tissue, aiding in the accurate identification of cancer and thus reducing the number of unnecessary prostate biopsies.[49]

Blood-based biomarkers, such as the 4Kscore and the Prostate Health Index (PHI), have also demonstrated the potential to enhance prostate cancer detection.[49] The 4Kscore (Detect Your Risk of Aggressive Prostate Cancer | 4Kscore® Test) combines detection of four kallikrein proteins (total PSA, free PSA, intact PSA, human kallikrein 2) to provide a personalized risk assessment for prostate cancer.[49,57] The 4Kscore improves the specificity and sensitivity of prostate cancer detection by considering multiple biomarkers. Similarly, the PHI test (Prostate Health Index Test (PHI) for Prostate Cancer - Prostate Cancer Specialists in NYC (newyorkurologyspecialists.com)) combines evaluation of total PSA, free PSA, and pro-PSA to enhance prostate cancer detection, offering a more comprehensive diagnostic approach.[49,58] Confirm MDx (Confirm mdx for Patients - mdxhealth) is another non-invasive PCR-based assay test that is used to analyze the expression levels of three genes: *APC, GSTP1*, and *RASSF*. The levels of these genes assist in the detection of prostate cancer. The Confirm MDx test is conducted on archived tissues from the biopsies conducted within the past 30 months in patients who have persistently shown elevated PSA levels and have been recommended for a repeat biopsy. The Confirm MDx assay identifies cancerous cells by detecting DNA methylation patterns associated with epigenetic changes. This test can also help identify the target areas for future biopsy procedures. The result for the Confirm MDx assay shows positive if DNA methylation is detected and negative if no DNA methylation is detected. This test has a 90% negative predictive value (accuracy in identifying patients that do not have disease).[59–61]

Oncotype Dx (GPSRE_Clinically-High-Risk-Overview.pdf (exactsciences.com)) is a test used for patients with confirmed cancer with biopsy results and undergoing active surveillance. Unlike traditional biopsies, this is a non-invasive type of biopsy procedure. The main target of this test is to evaluate the aggressiveness of the cancer and provide a personalized risk assessment. This risk assessment is classified by the National Comprehensive Cancer Network (NCCN) into three categories: very low risk, low risk, and intermediate risk. The Oncotype Dx test analyses 17 genes (12 specific to prostate cancer and five housekeeping genes) to predict the aggressiveness of prostate tumours. The Oncotype Dx test evaluates various cellular pathways, including stromal response, androgen signalling, proliferation stages, and cellular organization. The results are expressed as a Genomic Prostate Score (GPS) ranging from 0-100, with a higher score indicative of higher cancer aggressiveness.[59,62]

The Decipher test (Decipher Prostate | Guide Treatment for Prostate Cancer (decipherbio.com)) evaluates multiple biological pathways in cancer patients classified by the National Comprehensive Cancer Network (NCCN) as very low risk, low risk, or intermediate risk. This genome-wide test provides a score ranging from 0 to 1, independent of clinical and pathological features. The Decipher test assesses the likelihood of three key outcomes: 5-year metastasis, high-grade disease, and 10-year cancer-specific mortality. The Decipher test is typically performed on specimens from radical prostatectomy. It is beneficial for patients with adverse pathology, recurrence, or biochemical resistance, helping identify those who may benefit from salvage or adjuvant radiation therapy.[59,63–65] Salvage radiation therapy is given after the initial treatment when cancer recurs, while adjuvant radiation therapy is administered proactively after the initial treatment to prevent recurrence in high-risk patients.

The Prolaris tests Prolaris® Prostate Cancer Prognostic Test | Myriad Genetics analyses a combination of 46 genes, i.e. 31 cell cycle progression genes and 15 housekeeping genes as control with pathology and clinical features to predict ten years of metastasis risk after the disease-specific mortality and receiving proper treatment if managed conservatively. This tissue biopsy-based test can be helpful in post-prostatectomy specimens for predicting the

risk in biochemical recurrence over the next ten years to help subjects who might benefit from the adjuvant therapy.[59,66]

AR-V7 is a type of blood-based test that is very important in subjects having metastatic castrate-resistant prostate cancer, and those who have currently or previously received androgen receptor-targeted medications, and it is very helpful in determining appropriate future systemic therapy. In addition, androgen receptor AR-V7-positive patients have poor responses to blocking androgen receptors. Because of that, they might benefit more from chemotherapy or other non-androgen pathway therapies, while contrary to the androgen receptor-negative patients, they might respond to all therapeutic agents.[59,67,68]

1.10 Advancements in Imaging and Biopsy analysis of prostate tumours

Imaging techniques play a crucial role in the detection, staging, and monitoring of prostate cancer. Whereas prostate biopsy remains the gold standard for confirming the presence of prostate cancer[69],magnetic resonance imaging (MRI) has emerged as a valuable tool in prostate cancer diagnosis. Previously, patients with elevated PSA would have to undergo non-targeted transrectal ultrasound (TRUS) guided biopsy, which does lead to side effects such as haemorrhage and sepsis[70]. Additionally, conventional transrectal ultrasound (TRUS)-guided biopsy has limitations, including the fact that TRUS does not improve cancer detection without biopsy confirmation. This requires patients to have an increased number of biopsies that can lead to infections and ending up with a higher number of negative biopsies[71].

Recently multiparametric MRI (mpMRI), which combines anatomical and functional imaging, has improved the accuracy of prostate cancer detection by providing more detailed image about the tumour's location, size, and aggressiveness providing a way for an MRI-guided biopsy in men with positive prostate cancer[70]. In MRI-guided biopsy real time MRI images guide the biopsy needle to suspicious identified lesions [72]. Trans-perineal biopsy, which involves inserting a needle through the perineum, has shown higher sensitivity to prostate cancer detection. This approach allows for more systematic sampling of the prostate, reducing the likelihood of missing significant cancer foci. Furthermore, trans-perineal biopsy

has a lower risk of infection compared to the traditional transrectal approach, as it avoids passing through the rectum [73].

1.11 Therapeutic Approaches for Prostate Cancer

Prostate cancer treatment strategies are selected based on various factors, including disease staging, patient age, overall medical profile, and individual preferences. The treatment options for prostate cancer encompass active surveillance, surgery, chemotherapy, radiotherapy, and androgen deprivation therapy (ADT). The latest therapeutic options for prostate cancer are discussed below:

1.11.1 Active Surveillance

Active surveillance approach involves close monitoring of the cancer in men with low-risk disease through regular prostate-specific antigen (PSA) tests, digital rectal exams (DREs), and sometimes biopsies. This can delay or eliminate the need for immediate treatment and provides with an option to switch to active treatment options if disease progression is noted[74].

1.11.2 Surgery (Radical Prostatectomy)

Surgery is often recommended for localized prostate cancer and entails the removal of the prostate gland, seminal vesicles, nearby lymph nodes and surrounding tissue[75]. Radical prostatectomy can be performed through open surgery, laparoscopic surgery, or robot-assisted laparoscopic prostatectomy (RALP). Erectile dysfunction is the most common complication arising from the surgery[75].

1.11.3 Radiation Therapy

Radiation therapy is another widely used treatment for prostate cancer. It utilizes high-energy radiation to eliminate cancer cells and shrink tumours. External beam radiation therapy (EBRT) involves the delivery of radiation from an external machine, while brachytherapy

involves implanting radioactive seeds into the prostate gland. Radiation therapy is typically employed for localized or locally advanced prostate cancer, cancer not completely removed post radical prostatectomy, or it can be used for prostate cancers that have metastasized to the bone to relieve symptoms and control the spread for longest duration possible[76].

1.11.4 Hormone Therapy

Hormone therapy, also known as androgen deprivation therapy (ADT), aims to reduce the levels of male hormones (androgens), such as testosterone and dihydrotestosterone, in the body[77]. Lowering androgen levels can slow the growth of prostate cancer[75,77]. ADT can be achieved through surgical removal of the testicles or the use of medications that block the ligand dependent activity of AR such as Casodex (Bicalutamide). Casodex is also used to mimic the effects of ADT in vitro on human cells. Hormone therapy is given either with radiation therapy, or when the cancer resurfaces post-surgery and or radiation therapy. ADT is also employed in cases with widespread metastatic cancer and leads to shrinking of the prostate tumour. Patients that undergo ADT treatment develop CRPC and have a median survival of 18 months. Once patients develop CRPC the use of hormone therapy is limited. This raises the need to further understand and develop the treatment options post CRPC[75].

1.11.5 Chemotherapy

Chemotherapy involves the use of drugs (Docetaxel, Cabazitaxel, etc.) to target and kill cancer cells[78]. Chemotherapy is typically utilized for advanced prostate cancer that has spread to distant sites in the body, and often administered into a vein[78]. Chemotherapy can lead to multiple side effects of which most notable side effect commonly seen is a hair loss[78].

1.11.6 Immunotherapy

Immunotherapy stimulates the body's immune system to recognize and attack cancer cells. This is a relatively new treatment option for advanced prostate cancer that has not responded to other therapies. The Provenge vaccine (https://provenge.com) is an example of

immunotherapy that is personalized to each individual and develops the immune system to attack cancer cells[79,80].

1.11.7 Targeted Therapy

Targeted therapy involves the use of drugs that specifically target cancer cells by binding cell structures that differ from the normal cells, sparing healthy cells in the process[81].

1.11.8 Summary of current diagnostics and treatments

While significant advancements have been made in the field of prostate cancer diagnostics and treatments, there remains a critical challenge when the disease progresses to the CRPC stage. At this point, the available treatments often prove ineffective, leading to increased mortality rates. It is imperative to address this issue by identifying improved diagnostic and therapeutic targets. Additionally, the discovery of novel therapeutic targets is crucial to overcome the limitations of current treatment options for CRPC. These targets may include genes such as *ESRP1*, *ESRP2* and *NOVA1*, that may alter specific molecular pathways leading to cancer's aggressiveness.

1.12 Pre-mRNA, alternative splicing and the spliceosome

Gene expression is a highly regulated process that enables the synthesis of functional proteins essential for cellular processes. In humans, the pre-mRNA (an abbreviation of precursor mRNA) is the primary gene transcript that has to go through several steps of processing to make mRNA that can be translated into a protein. One of the important and crucial steps in pre-mRNA processing is splicing. Splicing is a complex and a tightly regulated process. During splicing introns are removed and exons are joined together to produce mature mRNA.[82] Alternative pre-mRNA splicing plays a very important role in expanding the information content of the genome, where ~20 thousand genes can produce a greater multitude of proteins, leading to protein diversity in human cells.[83]

1.12.1 Mechanism of Pre-mRNA Splicing

Pre-mRNA splicing is carried out in the spliceosome[84]. The spliceosome is a large nucleoprotein complex that is composed of small nuclear ribonucleoprotein particles (snRNPs) and numerous protein factors. The splicing process involves two sequential transesterification reactions catalysed by the spliceosome, resulting in the removal of introns and the joining of exons[84]. The first transesterification reaction, known as the branching reaction, initiates at the branchpoint adenosine. The first reaction is step 1 as seen in figure 1.6 and figure 1.7. The adenosine residue within the branchpoint forms a nucleophilic attack on the 5' splice site, resulting in the formation of a lariat structure and releasing the 5' exon[84]. The free 3' end of the 5' exon then undergoes a nucleophilic attack on the 3' splice site, leading to the second transesterification reaction of exon ligation seen as step 2 in figure 1.6 and figure 1.7. This reaction results in the covalent joining of the two exons, forming a mature mRNA molecule. The intron, in the form of a lariat, is subsequently released, debranched and degraded.[84]

Initially, the assembling spliceosome recognizes specific sequences at the intron-exon boundaries, known as splice sites Figure 1.6. The splice sites consist of a 5' splice site (5'SS), a branch point sequence (BPS) containing the key adenosine residue, and a 3' splice site (3'SS)[84]. The spliceosome assembles at the intron by recognizing the conserved sequences at the splice sites. The 5'SS is recognized by the U1 snRNP, and the BPS is recognized by the U2 snRNP[84]. The U4/U6.U5 tri-snRNP is subsequently recruited, leading to the formation of the pre-catalytic spliceosome[84]. Pre-catalytic spliceosome (B complex) is then activated by shedding U1 and U4, to become an activated spliceosome. This then matures into a catalytically active spliceosome (B* complex) where branching transesterification takes place. The catalytic step 1 spliceosome then becomes the step 2 catalytically active spliceosome where exon ligation takes place and, finally becomes a post-splicing (P complex) releasing a ligated exon mRNA and becoming an intron lariat spliceosome (ILS) complex that releases remaining intron lariat figure 1.6.

1.12.2 Regulation of Pre-mRNA Splicing

There are four splice site components: 5' splice site, 3' splice site, polypyrimidine tract and branch point site. These splice signals determine the introns and exons. Nevertheless, pseudo-exons also exist that have core splice site signals that match consensus elements. Therefore, it can be deduced that additional cis-acting elements must be involved in the process of pre-mRNA splicing that can differentiate between exons and pseudo-exons. There are additional cis-acting elements that have specific sequences within a pre-mRNA or in a very close vicinity of <500 nucleotide distance and can exert an effect on splicing. These cis-acting

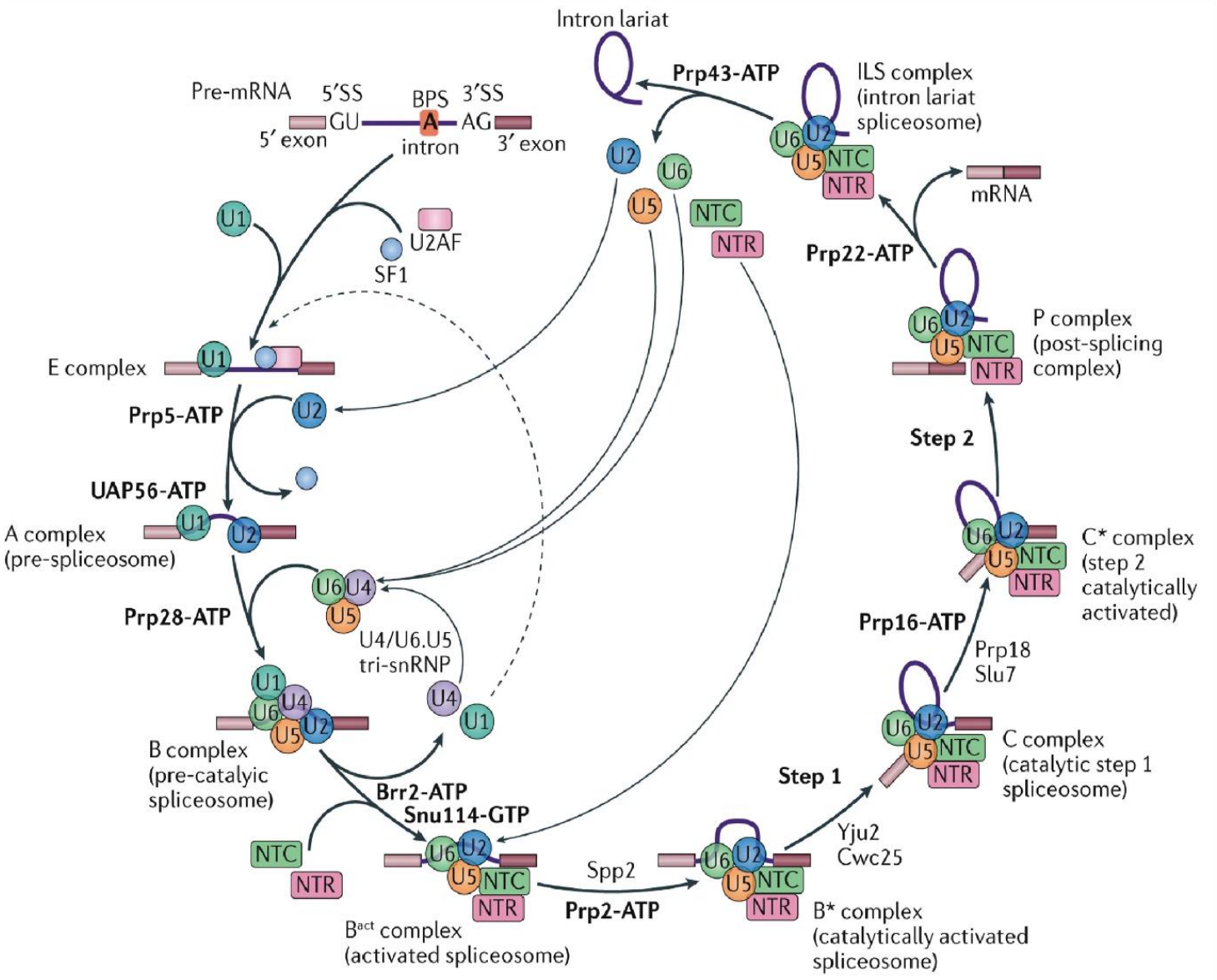


Figure 1.6: spliceosome assembly and splicing-reaction cycle. *U1 and U2 recognize the 5'SS and BPS and bind to it resulting in early spliceosome which through further processing becomes catalytically active. The end products of this cycle are an mRNA, intron lariat and disassembled snRNPs. Adapted from Shi et al.*[202]

elements that generate auxiliary splicing signals are called splicing regulatory elements (SREs). The SREs are grouped into two categories: exonic and intronic splicing enhancers ESE and ISE,

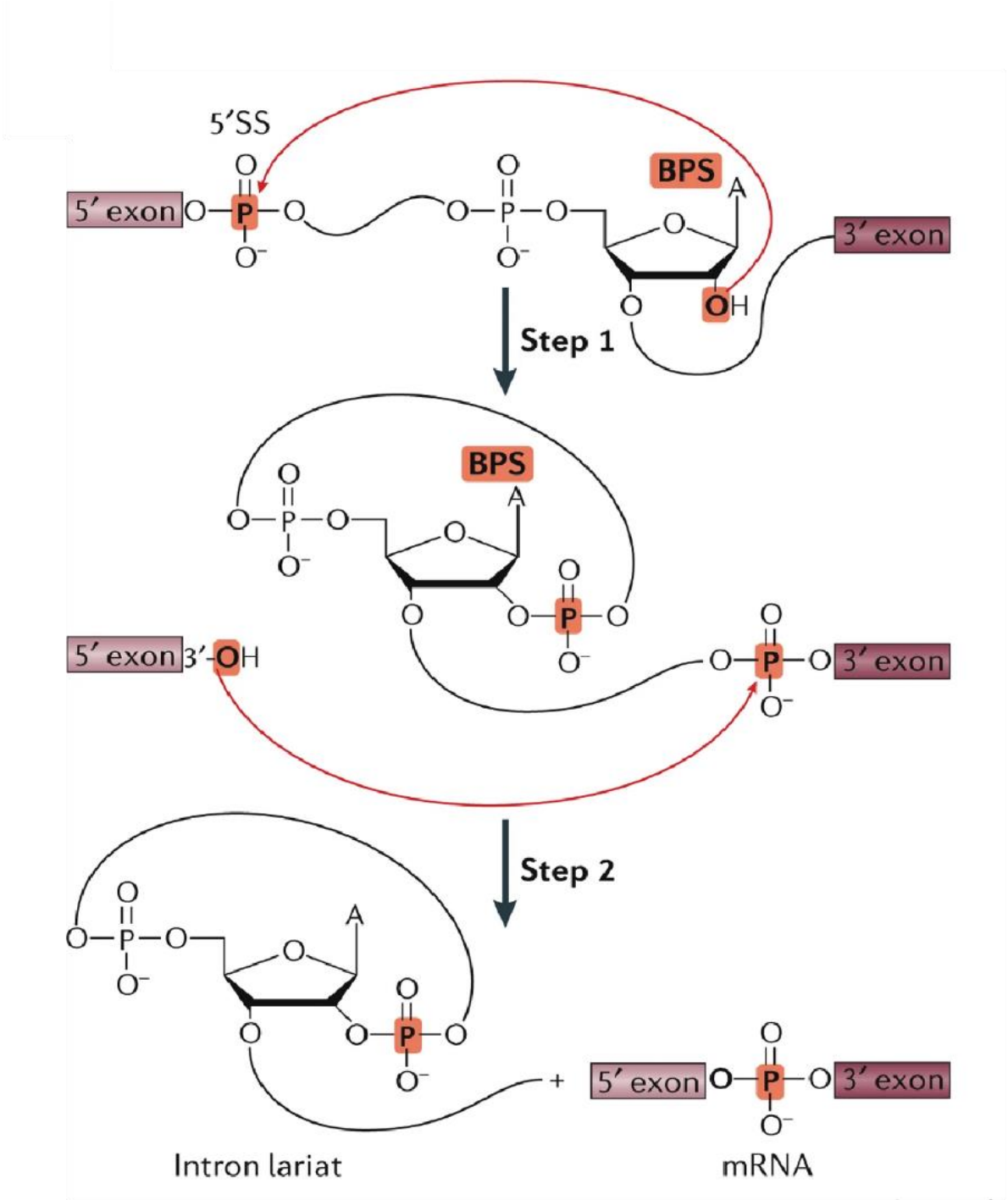


Figure 1.7: Transesterification steps in pre-mRNA splicing. *Step 1 (branching) releases a free 5′ exon and an intermediate of intron lariat - 3′ exon. In step 2 (ligation), the 5′ exon joins the 3′ exon and intron lariat is removed.*[202]

respectively, and exonic and intronic splicing silencers ESS or ISS, respectively. These SREs are recognized by either splicing silencer proteins or splicing enhancer proteins that, in turn, inhibit the splice site or enhance the splice site recognition, leading to the exclusion or inclusion of introns and exons, thus leading to alternative splicing.[85–88] These proteins are called trans-acting factors and are essential for the pre-mRNA splicing process. Trans-acting factors include serine/arginine (SR) proteins and heterogenous nuclear ribonucleoproteins (hnRNPs). These trans-acting factors bind to the pre-mRNA, regulate splicing, and exclude the pseudo-exons from the process of splicing. These factors can promote or inhibit splice site recognition, such as SR proteins functioning as splicing enhancers, whereas hnRNPs inhibit site recognition and work as strong splice site silencers. The complex interplay of these trans-factors, along with the strength of cis-acting nucleotide elements, affects the choices and strength of pre-mRNA splicing and, therefore, determines the outcome of the splicing. [85,89–91]

Epigenetic changes also affect pre-mRNA splicing by influencing the interaction between cis-elements and trans-acting factors. Epigenetic modifications can influence pre-mRNA splicing through histone modification and DNA methylation, alteration of chromatin structure, and splicing factor recruitment. Histone modifications such as histone acetylation can promote splicing, and histone methylation can repress splicing. Additionally, DNA methylation (addition of methyl group to DNA) can alter the binding of splicing factors to pre-mRNA.[92,93]

1.12.3 Significance of Pre-mRNA Splicing

Alternative splicing, a process driven by pre-mRNA splicing, plays a crucial role in expanding the coding capacity of the genome and generating protein diversity. Through alternative splicing, a single gene can produce multiple mRNA isoforms, each with distinct combinations of exons. Alternative splicing allows for the synthesis of different protein isoforms that can have diverse functions and cellular localization. Alternative splicing is particularly prevalent in multicellular organisms, where it contributes to tissue-specific gene expression and the complexity of biological processes.[82,94]

1.12.4 Alternative splicing of the *Androgen Receptor* gene

The *AR* gene is located on the X chromosome[95]. During prostate cancer disease progression, the *AR* gene can produce *AR* mRNA splice variants, some of which encode truncated proteins that can modulate transcription in the absence of androgens. These variants arise through alternative splicing of the *AR* gene and have been discovered to be an important factor in the CRPC or metastatic prostate cancer.

There are approximately 20 known AR variants as of 2016 that include AR-V1, AR-V3, AR-V7, AR-V9, AR-V567es[96]. Well-known AR variants include AR-V7 and AR-V9. AR-V7 and AR-V9 are generated through AS of the *AR* pre-mRNA[44]. the levels of AR-V7 are significantly higher in CRPC patients [43–45,96]. AR-V7 is encoded by exons 1-3 of the AR gene, with a cryptic exon 3 (cryptic exons are not normally recognised by the spliceosome). As AR-V7 arises from splicing of different cryptic exons located within intron 3 of the *AR* gene this introduces a translational stop codon after exon 3. This means that the coding information for the NTD and DBD are retained in the mRNA, but AR-V7 lacks the coding information for the ligand-binding domain and so is constitutively active, allowing it to regulate gene expression in the absence of androgen ligands[44,96]. AR-V7 is seen as highly clinically relevant as it is detected in circulating tumour cells in CRPC patients and is regarded as a prediction factor for treatment resistance to enzalutamide and abiraterone.

Another variant, AR-V1, arises from alternative splicing of exons 1 and 2. AR-V1 retains the NTD and DBD but lacks the LBD. AR-V1 along with AR-V9 is conditionally active[96]. AR-V9 is encoded by exons 1-3, and cryptic exon 5. Furthermore, other AR variants, such as AR-V4, AR-V5, AR-V12, and AR-V13 have been identified, each with distinct structural modifications and functional characteristics. These variants contribute to the complexity of androgen signalling and the regulation of gene expression[96].

1.12.5 Implications of AR Variants in Health and Disease

AR target genes regulate crucial cellular functions of growth, proliferation, and survival. This means that AR variants have significant implications in health and disease, particularly in the

context of prostate cancer[44]. CRPC, characterized by disease progression despite androgen deprivation therapy, is often associated with the emergence of AR variants, mainly where AR-V7 has been found to be elevated in CRPC[44,96]. AR-V7 enables cancer cells to bypass the need for androgen ligands, continue to drive tumour growth and survival, and contribute to treatment resistance to enzalutamide and abiraterone[44].

Beyond prostate cancer, AR variants have also been implicated in other diseases. For instance, AR is associated with Kennedy's disease, a neurodegenerative disorder characterized by muscle weakness and atrophy[97].

1.13 Epithelium-mesenchyme transition (EMT) and mesenchyme to epithelium transition (MET)

Epithelial cells are organized in sheets that adhere to each other tightly. They form the tissue lining and form barriers. Whereas mesenchymal cells are loosely held which allows them to have highly migratory properties and make them ideal cells to metastasize and invade human tissues. Epithelial cells are more replicative whereas mesenchymal cells do not replicate and are less susceptible to chemotherapeutic agents.[36,98,99]

An EMT (epithelial to mesenchymal transition) process is a mechanism where cells transform from epithelial cell state to mesenchymal cell state acquiring properties that aid in tumour metastasis. EMT process is proposed to be required for cancer cell dissemination and enabling cancer cell invasion. The epithelial-to-mesenchymal transition associated with cancers causes downregulation of E-cadherin, causing the loss of cell-cell adhesions that promote detachment of primary tumour mass, and activate cell migration.[100,101] Through the induction of epithelial to mesenchymal transition, tumour invasion into the rest of the body may be promoted by the loss of e-cadherin and its dissemination. The epithelial-to-mesenchymal transition of cancer progression has been widely studied worldwide using animal models and in vitro cell cultures.[100,102]

It has also been proposed that a reverse epithelial-to-mesenchymal transition or MET (mesenchymal-to-epithelial transition) occurs to colonize metastatic sites. MET enables the last step of metastasis, by downregulating metastatic genes and re-expressing its epithelial genes to survive in the migration/secondary organs. it is of particular importance to look at the histological examination of epithelial and mesenchymal markers and compare the results obtained in the metastatic samples to the primary samples to observe and understand if EMT and MET can be determined clinically. Understanding disease mechanisms and developing advanced diagnostic and therapeutic regimens is vital. A study by Chao et al. reported increased E-Cadherin expression levels in breast and prostate cancer metastatic tumours compared to the primary tumours.[100,103,104] Additionally, the study reported an induction of expression of E-Cadherin in otherwise E-cadherin-negative MDA-MB-231 breast cancer cell lines in a mouse model.[100,103] Additionally, to test whether total or partial MET had occurred, the study evaluated matched samples from primary and metastatic tumours in a more extensive set of breast cancer patients. It was seen that epithelial markers were higher in metastatic samples. In contrast, mesenchymal markers such as FSP1 and vimentin had variable changes in the same samples, which concluded that the MET process happened partially. This phenomenon of EMT and MET regarding cancer aggressiveness was further tested in unmatched prostate cancer samples from primary and metastatic samples, where it was seen that metastatic samples with increased metastatic size had lower E-cadherin expression compared to the primary tumour samples.[100] The conclusion from this observation could be drawn that MET was also reversible. Therefore, there is a possibility that not all metastatic tumours come from the primary tumour site. Still, metastasis can cause additional metastasis due to the reversible nature of EMT and MET processes.

Despite its importance in cancer biology, the role of epithelial to mesenchymal transition remains debatable.[100,101,105,106] Even though literature suggests that there is a strong correlation between EMT, tumor invasion, and poor patient prognosis with decreased expression of adhesion molecules. Samples taken from metastatic lesions often present with epithelial, well-differentiated phenotype.[100,106–109] This raises a question about whether EMT process at the site of origin is a reversible process? However, lack of evidence does not mean that EMT has not occurred because pathological specimens are mostly obtained in the late stage of cancer. Therefore, it can be deducted that process of EMT takes place at the site of

origin of cancer where epithelial cells acquire mesenchymal properties and mobilize to the location of metastasis where a reverse process of EMT know as mesenchymal-epithelial transition (MET) occurs.

1.14 RNA Binding Proteins

RNA binding proteins (RBPs) regulate gene expression and RNA interaction[110,111]. RBPs contain RNA binding domains such as RNA recognition motifs (RRMs), KH domains, and zinc finger motifs[112] that interact with target RNA molecules, thus exerting control on processes such as splicing, transport, localization, stability, and translation[110–112]. RNA binding proteins target molecules including messenger RNA (mRNA), non-coding RNA (ncRNA), etc. One of the most important processes regulated by RBPs is alternative splicing that leads to multiple isoforms of a single gene and contributing to the protein diversity of human proteome[110,111]. Some RNA binding proteins have tissue specific expression patterns that are involved in neuronal development, the immune response and metabolic regulation. Mutations can happen in RNA binding proteins that regulate splicing control and have been associated with neurodegenerative disorders. Furthermore, RNA binding proteins that are involved in functions such as RNA stability can contribute to cancer progression when damaged. Some RNA binding proteins of importance to the research topics in this thesis include but are not limited to ESRP1, ESRP2 and NOVA1.

1.14.1 ESRP 1 and 2: master regulators of epithelial splicing

ESRP1 (Epithelial Splicing Regulatory Protein 1) and ESRP2 (Epithelial Splicing Regulatory Protein 2) are epithelial cell-type-specific splicing regulators and RNA binding proteins. ESRP1 and ESRP2 play an important role in epithelial splicing therefore also called master regulators of splicing.[48,113] ESRP1 and ESRP2 are close paralogs with 60% overall similarity between the two genes.[36] ESRP 1 and ESRP2 both have RNA binding domains to bind to the target RNAs and exert the control. The RNA binding domains for ESRP 1 and ESRP2 are three RNA recognition motifs (RRMs), Figure 1.8.[36] ESRP proteins are present within epithelial cells,

thus organs that contain higher number of epithelial cells have high expression of ESRP1 and ESRP2 including prostate tissue as shown in Figure 1.9.

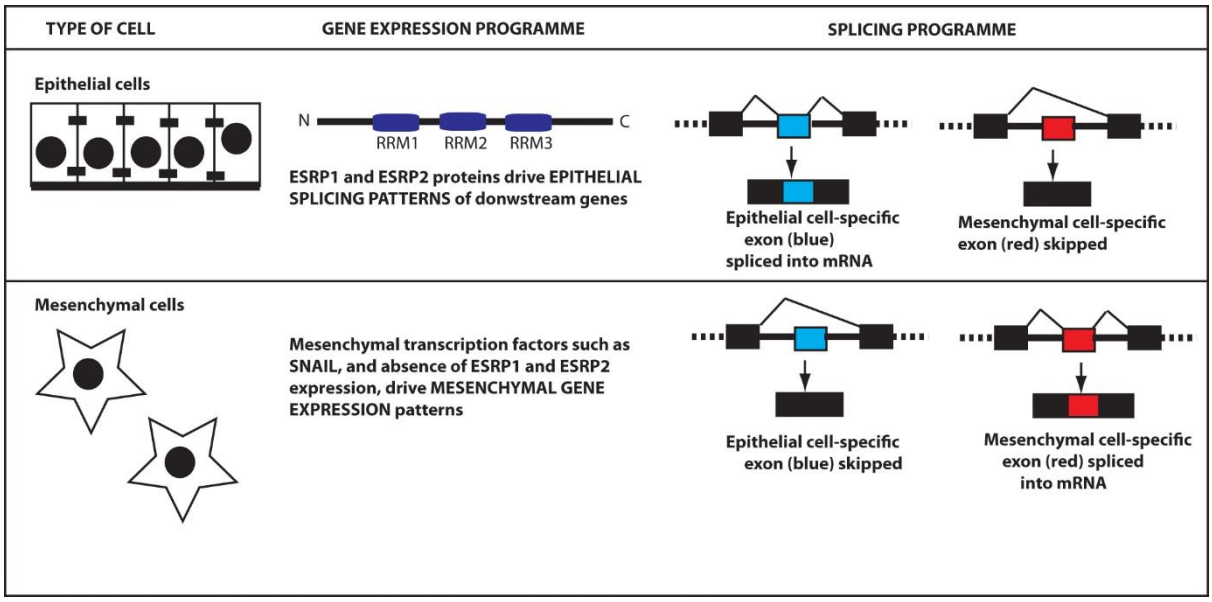


Figure 1.8: ESRP1 and ESRP2 regulate splicing in epithelial cells. *Epithelial cells are physically connected and grow in sheets expressing high levels of ESRP1 and ESRP2 proteins. ESRP1 and ESRP2 lead epithelial splicing patterns, and each consists of three RNA Recognition Motifs (RRMs). Mesenchymal cells differ in morphology to epithelial cells to aid in mobility and have differing gene expression patterns.*[36]

ESRP1 and ESRP2 proteins regulate production of epithelial-specific alternative mRNA isoforms. Additionally, ESRP1 and ESRP2 can compensate for each other in case where one of the two is absent. Moreover, in cases where both ESRPs are eliminated in mouse models this leads to improper organ formation because of a failure to form epithelial tissue.[114]

ESRPs bind to pre-mRNA to regulate splicing inclusion. ESRP proteins drive exon skipping if they bind upstream of target exons whereas, binding of ESRPs to the intronic sequences downstream regions of exons leads to splicing inclusion.[36,115,116] It has also been seen that in patients with higher expression of ESRP1 and ESRP2 the prostate tumours are more likely to be aggressive leading to poor patient prognosis.[36,48,113]

Although the AR was initially viewed as a transcription factor, it is known now that androgens change alternative splicing patterns. Munkley et al. showed that *ESRP2* gene and protein expression is directly proportional to the androgen expression.[48] It was also seen that androgens carry out epithelial splicing function via exerting a direct control over ESRP1 and ESRP2 expression levels.[36,48,113] ADT, a therapy used for the advanced prostate cancer treatment, reduces the expression of ESRP1 and ESRP2, and also impacts alternative splicing patterns.

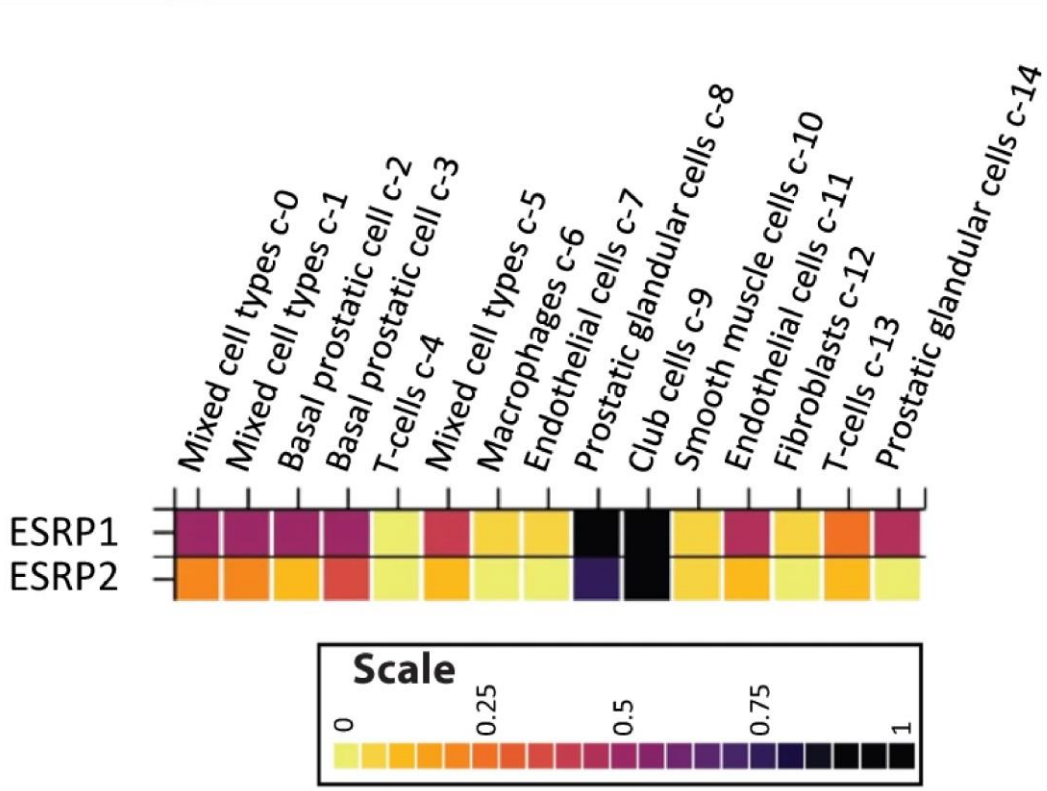


Figure 1.9: Cell types expressing ESRP1 and ESRP2. *ESRP 1 and ESRP2 are highly expressed in cells (Luminal Secretory cells, Basal cells, Neuroendocrine cells, Club cells, and Hillock cells) that give rise to prostate cancer. ESRP1 and ESRP2 can be survival oncogenes.[36]*

Figure 1.10A and 1.10B show distinct gene expression patterns in tumors that have either increased or decreased *ESRP1* gene expression levels. Tumors with high *ESRP1* gene expression are associated with G2M checkpoints as seen in figure 1.10A. Genes related to high ESRP1 tumour include *MKI67* (encodes proliferation marker Mki67); *CDK1, CCNA2,*

CCNB2 (cell cycle regulators); and *BIRC5* (inhibits apoptosis). The higher number of cell cycle associated genes suggests that epithelial cells are more proliferative.

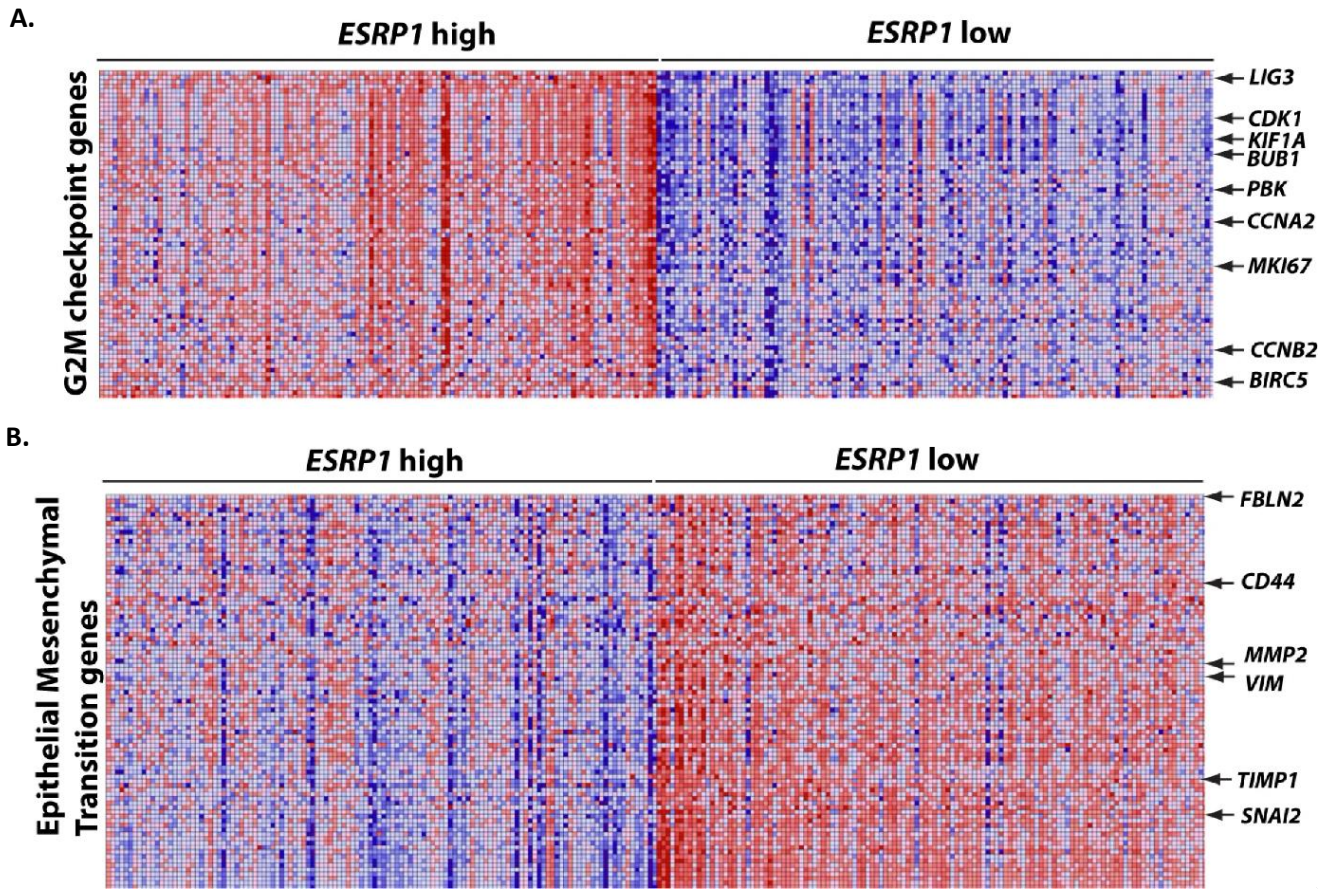


Figure 1.10: Heat map shows gene expression change is ESRP linked tumours. *A) Higher expression of G2M transition associated genes in high vs low ESRP1 tumours. B) Lower expression of genes linked to EMT in high vs low ESRP tumours. Red indicates high gene expression and blue indicates low gene expression.*[36]

1.14.2 NOVA1

Neuro-oncological ventral antigen (NOVA) is an RNA binding protein that was initially identified in patients with neurologic disorder known as paraneoplastic opsoclonus-myoclonus ataxia that was associated with known cases of breast and lung cancer.[117–119] *NOVA1* is located on human chromosome 14. NOVA proteins bind to a YCAY RNA sequence within target pre-mRNAs.[117,120–122] NOVA has been most studied in regard to neuronal disease pathology and about 200 NOVA RNA targets are known.[117,123]

The NOVA family consists of two members, called NOVA1 and NOVA2.[117,124] Both are known to bind to target pre-mRNAs and regulate approximately 700 exons. NOVA1 expression has been detected in ventral spinal cord and midbrain, whereas NOVA2 is expressed in neocortex and hippocampus. [117,125,126] NOVA1 is a splicing factor that regulates alternative splicing in neurons and is considered to be brain specific.[117,118,122,127,128] NOVA1 is critical for maintaining the viability of motor neurons.[117,121] NOVA1 has also been known to have impact on pancreatic beta cell function and have an effect on insulin release. NOVA1 expression in pancreatic beta cells was ten-fold more than other tissues.[117,129]

NOVA1 is also thought to be involved in multitude of diseases, not just neuronal disease.[117] Higher expression of NOVA1 during reperfusion after an ischaemic injury correlated with improved neuronal repair, suggesting that NOVA1 expression might be important in cases of cerebrovascular accidents.[117,130] Familial dysautonomia is a recessive genetic disorder that affects the development and survival of sensory and sympathetic neurons in sensory nervous system. In this disease it was seen that almost 26 miRNAs were altered out of which four targeted NOVA1.[117,131]

NOVA1 is also important in cancer. In tumours such as astrocytoma and oligodendroglioma, high levels of NOVA1 have been associated with poor survival, whereas downregulation of NOVA1 correlated with a decreased level of cell proliferation, cell invasion and migration suggesting NOVA1 as an oncoprotein.[117,132] In hepatocellular carcinoma (HCC), increased expression of NOVA1 correlated with poor prognosis and increased chance of HCC recurrence.[117,133,134] NOVA1 is also seen to have varying expression in different lymphomas having varying impacts.[117,135] On the contrary, NOVA1 downregulation has been correlated with poor prognosis in gastric cancer patients and thus an increased mortality.[117,136,137]

NOVA1 is a good potential candidate to study for its impact on prostate cancer as recent studies have shown higher NOVA1 expression in aggressive prostate cancers such as neuroendocrine prostate cancer.[138]

1.15 Background to this thesis

Previous work in the Elliott lab set out to identify genes that were activated in response to androgen exposure in cell culture and repressed within patients undergoing ADT, and vice versa.[47] Munkley et al. (2016) revealed that numerous genes exhibited differential expression in response to the presence of androgens, with some genes being upregulated and others downregulated. Although they initially focussed research on the genes involved in glycosylation in prostate cancer cells, another of these reciprocally regulated genes was *ESRP2*. Specifically, the *ESRP1* and *ESRP2* genes were identified as being upregulated by more than two-fold in the 2016 study, and this finding was later corroborated in a follow-up study in 2019. The 2019 study by Munkley et al. further established a strong connection between the *ESRP2* genes and androgen regulation. It demonstrated that the *ESRP2* expression levels changed proportionally to androgen stimulation, indicating that androgens played a critical role in controlling the expression of the *ESRP2* gene. Consequently, androgens indirectly controlled the splicing of genes through *ESRP2* gene regulation.

In the same vein, Munkley et al. (2016) also observed a significant downregulation of the *NOVA1* gene by more than two-fold in response to androgens. However, no subsequent investigations were carried out on this gene. Lu et al. (2015) conducted a study to investigate the changes in the expression of two genes, *ESRP2* and *NOVA1*, in prostate cancer and found compelling evidence suggesting that these genes were regulated inversely. Considering the implications of the three studies mentioned, possible androgen control of the *NOVA1* gene emerged as a potentially important gene that warranted further examination (Munkley et al., 2016; Munkley et al., 2019; Lu et al., 2015). In light of this significance, our project aimed to investigate the correlation between the *NOVA1* gene and androgens, with the primary objectives of confirming the androgen dependency of *NOVA1* gene expression and investigating its protein-level expression.

1.16 Aims:

The primary goal of this research is to understand the role of *NOVA1* in prostate cancer and its relationship with androgen receptor (AR) regulation. Therefore, the objectives for this research were as follows:

Objective 1. Confirm whether NOVA1 is a direct target of AR and what prostate cancer cell line expresses it most.

First objective was to investigate the correlation between *NOVA1* gene and androgens by looking at NOVA1 protein expression in presence and absence of androgen and trace the NOVA1 RNA expression change over time. Additionally, to find out the prostate cancer cell line that expresses the highest levels of NOVA1 protein. A recent study has reported higher NOVA1 gene expression in neuroendocrine prostate cancer cell lines.[138] Moreover, CWRRV122, DU-145, PC-3, LNCaP, and Neuroendocrine cell lines. Therefore, I aimed to assess NOVA1 expression in the PC3, CWR22RV1, LNCaP, and Neuroendocrine prostate cancer cell lines. Additionally, to analyse the binding patterns of AR to the NOVA1 gene.

Objective 2: KD cell line expressing most NOVA1 protein expression, carry-out RNA-Seq and verify if any known ESRP targets are regulated by NOVA1.

Perform *NOVA1* gene knockdown in the prostate cancer cell line that has the highest NOVA1 protein expression. Then, to extract a good quality RNA to perform RNA sequencing. Additionally, visualize the known *ESRP1* and *ESRP2* targets in the NOVA1 RNA sequencing data to identify the common targets between the two data sets and how they might regulate in each of the two datasets.

Objective 3: Use NOVA1 knockdown RNA-Seq data to analyse differential gene expression and differential splicing.

Carry out the differential expression analysis and gene ontology to find out genes that are down or upregulated in response to *NOVA1* knockdown. Furthermore, to identify the pathways that the NOVA1 regulated genes are associated with. Next, to perform functional analysis on the cell line to highlight any phenotypic changes that might occur. Lastly, differential splicing analysis would be performed to find out the novel splicing targets of NOVA1 and how the splicing targets of NOVA1 correlate with disease using Psichomics.

Chapter 2: Materials and Methods

2.1 Cell culture

Dr Gerald Hysenaj provided me with LNCaP cells. The Munkley Lab acquired these cells from ATTC (American Type Culture Collection) (LNCaP clone FGC - CRL-1740 | ATCC). The cells were maintained in RPMI medium with 10% fetal bovine serum and 1% penicillin and streptomycin antibiotics. They were stored in 75 cm tissue culture flasks and then incubated at 37°C in 5% CO2.

The cells were split every 4-5 days at a 1:5 ratio. Initially, the RPMI media from T75 tissue culture flasks was removed and rinsed with PBS (phosphate-buffered saline). Then, 2 ml of trypsin EDTA was added to the flask and placed in an incubator at 37°C for 5 minutes, after which the dislodged cells were seen under the microscope, and RPMI media was added to the flask. The media containing trypsin, RPMI, and cells were shifted to falcon tubes, centrifuged for one minute, and the supernatant was discarded. 10 ml of RPMI media was added to the pellet, and cells were shifted to a new T75 flask. The flask was examined under the microscope to confirm the cells, and the flask was then stored in an incubator at 37°C in 5% CO2.

For steroid depletion (SD), the cells were grown in 10% charcoal stripped FBS. A concentration of 10 nM synthetic androgen analogue (R1881) was added to the cells for androgen treatment (Androgen +). Additionally, LNCaP cells, referred to as DMSO or Vehicle, were treated with dimethyl-sulfoxide (DMSO).

The Munkley group provided me with additional cell lines (PC3, DU-145, CWR22RV1, and Neuroendocrine). PC3 and CWR22RV1 cells were originally sourced from ATCC (PC-3 - CRL-1435 | ATCC)(22Rv1 - CRL-2505 | ATCC). The Munkley group obtained neuroendocrine cell lines from Canada. The neuroendocrine cell lines were synthetically created and not isolated directly from a host. The cells were grown in RPMI medium with 10% fetal bovine serum and 1% penicillin and streptomycin antibodies. The cells were split every 3-5 days per the above protocol.

2.2 Cell preservation

After receiving the cells, they were split and cryopreserved for additional stocks when required. The cell pellet was resuspended in solution (RPMI media, 5% DMSO, and 10% FBS). The cells were then cryo-stored at -80°C.

2.3 SDS-PAGE

Protein samples were mixed with 2X protein loading dye and denatured by heating at 95°C for 10 minutes, followed by sonication to ensure sample homogeneity. Proteins were separated using SDS-PAGE on a 10% SDS (Sodium Dodecyl Sulfate) resolving gel was prepared using H_20, 1.5M Tris pH 6.8, Acrylamide, SDS, APS (Ammonium Persulfate), TEMED (Tetramethyl ethylenediamine) with a 10% stacking gel. Electrophoresis was performed at a constant voltage of 200 V for 1 hour using a standard vertical gel apparatus.

Following electrophoresis, proteins were transferred onto a nitrocellulose membrane using a wet transfer system at 350 mA for 1 hour. Transfer efficiency was confirmed by briefly staining the membrane with Ponceau solution. The membrane was then washed with TBST containing 5% non-fat dry milk to block non-specific binding sites.

Immunodetection was performed by incubating the membrane with the appropriate primary antibody overnight at 4°C. After washing, membranes were incubated with an HRP-conjugated secondary antibody for 1 hour at room temperature. Protein bands were visualized by chemiluminescence and exposed onto X-ray film for signal detection.

2.4 Western Blotting

Cells were dissolved in 2X protein loading dye (4% SDS, 20% glycerol, 200mM DTT, 0.01% bromophenol blue, and 0.1 M Tris HCL, pH 6.8) to prepare protein samples. Protein samples were then sonicated and then denatured for 5 minutes at 100°C. A 10% SDS-Gel was prepared, and individual protein samples along with a protein molecular size marker were

loaded on the gel wells and ran in an electrophoresis chamber, following which the proteins within the gel were transferred using a wet transfer onto a nitrocellulose membrane. The membrane was then blocked with 5% non-fat milk and 1% horse serum in TBST (Tris Buffer Saline with Tween 20) for 1 hour at room temperature. After the milk blocking step, the membrane was washed with TBST. Then it was incubated with primary antibody (Abcam Anti-Nova1 antibody, ab183024, host rabbit) diluted to a ratio of 1:1000 in blocking buffer overnight at 4°C. After primary incubation, the membrane was washed in TBST and incubated with a secondary antibody (anti-rabbit HRP, host: goat, from Jackson Labs) diluted at a ratio of 1:2000 in blocking buffer conjugated to horseradish peroxidase for 1 hour at room temperature. Specific antibody binding was detected using ECL (Enhanced Chemiluminescence) supplied by Cytiva Lifesciences, exposed to X-ray film. X-ray film was developed using the developer in the lab dark room. Western blot bands on the film were quantified using ImageJ (Fiji) software per developer instructions. ß-Actin (Sigma-Aldrich, A5441) was used as a loading control to probe Western blots.

Fiji image J software was then used to quantify the western blots. The western blot image was scanned and loaded using image J. Then, a rectangle tool was selected from the menu which was then used to mark the 1st lane on western blot. The 1st selected lane becomes the reference lane to trace the remaining rectangular areas highlighting bands in consecutive lanes. Same size rectangle tool was then used on the following lanes and numbered lane 2 onwards. Then the analyze option in the menu of the Fiji software was used to generate intensity peaks. Then straight-line tool was used to mark the baseline for each read following which the area under the curve was selected. Then to normalize read values, intensity of each band was divided against the reference read of the first lane band. The resulting value was used based on given formulas in Fiji software manual and the results were used in GraphPad prism to generate the bar graphs.

2.5 qPCR

In one flask, steroid deplete media was added whereas, in the second flask synthetic androgen R1881 media was added. Cells in both flasks were allowed to grow for 72 hours and then harvested. A cell pellet was obtained by microcentrifugation and used for RNA extraction

using a Qiagen RNA extraction kit. This RNA was then used to carry out RT-qPCR. Three replicates of each RNA sample were loaded to a 96-well plate. Three loading controls were used (actin, tubulin, and GAPDH) and their averages were used as a reference to calculate sample fold change. RT-qPCR returned the CT values for each sample. The average for all the loading samples was taken and used to calculate a delta CT (target gene CT – control gene CT) for each replicate. A delta-delta CT was then calculated using control group average and delta CT values which were add to the formulae (2^-delta-delta CT). A standard error was then calculated using an excel formula. Results were used in GraphPad PRISM to generate a bar graph.

2.6 siRNA knockdown

For siRNA transfection in chapter 3, LNCaP prostate cancer cells were seeded into six-well tissue culture plate and incubated for 24 hours at 37°C in a humidified incubator with 5% CO_2. Transfection was carried out using AR-specific siRNA (AR esiRNA, EHU025951, Sigma-Aldrich) and a negative control siRNA (IDT, 51-01-14-04) using RNAiMax lipofectamine transfection protocol. Cells were incubated for 72 hours at 37°C in an incubator with 5% CO_2 post-transfection to allow gene knockdown. Following incubation, cells were harvested, lysed, and mixed with 2× protein loading dye for downstream protein analysis.

NOVA1 knockdown in chapter 4 was achieved using siRNA (IDT NOVA1 siRNA, ID no. hs.Ri.NOVA1.13.1). Cells were transfected with siRNA using the Lipofectamine RNAiMAX transfection protocol (Thermofisher)(Lipofectamine™ RNAiMAX Transfection Reagent (thermofisher.com)). Negative control samples were transfected with negative control siRNA (Ambion)(Ambion™ In Vivo Negative Control #1 siRNA (thermofisher.com)).

To perform the siRNA NOVA1 knockdown in PC3 cell lines a master mix was prepared in 2 different universal tubes containing either a NOVA1 siRNA (Tube 1) or a negative control siRNA (Tube 2). A volume of 375 µl optimum media and 18 µl RNAiMax reagent were added to each tube. Then 15 µl (10 µM siRNA) of NOVA1 siRNA was added to the Tube 1 and 15 µl (10 µM siRNA) of negative control siRNA was added to the Tube 2. Tube 1 and Tube 2 were incubated at room temperature for 5 minutes before transfection. The knockdown incubation

was performed in a 6-well plate. Tube 1 was used to transfect PC3 cells to obtain the three NOVA1 knockdown samples in a 6 well-plate. Tube 2 was added to the three control wells of the 6-well plate. The 6-well plate was then stored in an incubator at 37°C and 5% CO_2 for 72 hours. The PC3 cells were seeded on to a 6-well plate 24 hours before the transfection was carried out at a confluency of approx. $1X10^6$. This cell density was established using a haemocytometer in the tissue culture lab. After 72 hours of incubation cells were harvested using trypsin. Cells were split into two halves for RNA purification (QIAGEN RNeasy Mini Kit) for RNA sequencing; and for protein isolation (added 2X protein loading dye) for western blotting.

Optimization of knockdown protocol.

PC3 cells received from the Munkley lab, cultured in RPMI media, were plated in a six-well plate and incubated for 24 hours at 37°C. Transfection was performed using RNAiMAX Lipofectamine with a 0.9µl siRNA of IDT NOVA siRNA design ID hs.Ri.NOVA1.13.1 (diluted to 100 µM) in 10 µl nuclease-free water. Then, it was distributed in 3 wells of the 6 well plate for knockdown samples compared against a negative control. Despite incubating at 37 degrees Celsius for 48 hours, the Western blot analysis from the first attempt did not reveal any discernible change in NOVA1 protein band intensity, indicating an unsuccessful knockdown. A subsequent attempt using a new PC3 cell line sample with a lower passage number using the same protocol was done that had a similar result to the first attempt with no change in bands, prompting a re-evaluation of the siRNA concentration.

In the third attempt, a higher siRNA solution amount of 1.5µL was used in 10µl of nuclease free water, slightly altering NOVA1 protein bands on the Western blot. Undeterred, a fourth attempt was made, using the same 1.5µl siRNA in 10µl nuclease free water and extending the incubation period to 72 hours. This fourth attempt led to visible NOVA1 protein band alteration. Thus, the knockdown on the western blot was confirmed. Then, the RNA was extracted using the trizol protocol, but the RNA obtained was of poor quality as it had less

than 100ng/µl concentration of RNA on nanodrop, precluding its usability for downstream RNAseq analysis.

A fifth attempt maintained the 1.5µl siRNA solution to a 10µl nuclease-free water ratio and extended the incubation to 72 hours.

2.7 cDNA and PCR quantification

The RNA from cells treated with siRNA were extracted by following instructions included in the QIAGEN RNeasy mini kit provided by the manufacturer (RNeasy Kits | RNA Extraction Kits | QIAGEN). In each case cDNA was synthesized using Superscript VILO cDNA synthesis kit (Invitrogen) following manufacturer's instructions (SuperScript™ VILO™ cDNA Synthesis Kit (thermofisher.com)). Primers were designed using Primer 3 Plus (Primer3Plus - Pick Primers) in order to analyse the splicing profiles of the alternative events identified by SUPPA2 for novel *NOVA1* splicing targets. The primers were tested in the UCSC *In-Silico* PCR tool. Primers from a previously published study by Munkley et al. 2019 were used to identify any changes of known ESRP1/2 targets in NOVA1 knockdown vs control cDNA samples. A full list of primers for all chapters is in the Appendix. All of the PCR reactions were performed using GoTaq G2 DNA polymerase kit from Promega in Table 2.1. All PCR products were examined using the QiAxcel capillary gel electrophoresis device made by Qiagen. The band concentrations obtained from QiAxcel were used to calculate amounts of RT-PCR products (three biological replicates were used for statistical significance). Below is the formula used to calculate PSI:

$$\frac{Concentration\ of\ exon\ included\ (ng/\mu l)}{Concentration\ of\ exon\ included + exon\ excluded\ (ng/\mu l)} X\ 100 = PSI\ (\%)$$

Table 2.1: PCR reaction mix

5X Reaction Buffer	4 µl
Forward Primer (10 µM)	1 µl
Reverse Primer (10 µM)	1 µl
GoTaq	0.1 µl
dNTPs (10 µM)	0.4 µl
dH_2O	12.5 µl
cDNA	1 µl
Total	20 µl

Above Table 2.1 shows the PCR reaction mix to prepare the PCR solution and run in a thermocycler based on the programs listed in Table 2.2.

Table 2.2: Thermocycle Steps.

Initial heating	95°C	5 minutes
Denaturation	95°C	20 seconds (30 Cycles)
Annealing	58°C	30 seconds (30 Cycles)
Extension	72°C	60 seconds (30 Cycles)
Final extension	72°C	10 minutes
Cooling	4°C	∞

RT- qPCR:

LNCaP cells were treated with steroid depletion, and androgen was harvested. RNA was extracted using TRI-reagent (Invitrogen, 15596–026), per the manufacturer's instructions (TRIzol™ Reagent (thermofisher.com)). RNA was treated with DNase 1 (Ambion). cDNA was generated by reverse transcription of 500 ng of total RNA using the Superscript VILO cDNA synthesis kit (Invitrogen, 11754–050). The cDNA obtained was then used to perform RT-qPCR on cDNA triplicates using SYBR green and *actin*, *tubulin*, and *GAPDH* as loading controls. Forward and reverse primers for *NOVA1* were designed using primer 3 plus online with a

sequence (Rev. 5’ TCAGTGCTTCAACCGTTCCC 3’; Fwd. 5’ GACCAATACGGGCGAAGACG 3’). The QuantStudio 7 Flex Real-Time PCR system was used.

LNCaP cells in different treatment conditions:

LNCaP cells were grown in tissue culture and split into two flasks. These flasks were further split into three different flasks. LNCaP cells were grown either with or without androgens, and in each case, they were either additionally treated with Casodex, DMSO, or untreated. These samples were then further divided into three technical replicates each and used for RNA extraction which was then analysed by RT-qPCR

Androgen treated LNCaP cells:

LNCaP cells were grown in cell culture split into different flasks and grown in triplicate in either steroid deplete media alone or media with the synthetic androgen R1881. After treatment cells were harvested at 72 hours and loading dye was added to the cell pellet and mixed to make a protein extract. This extract was then used to load an SDS-polyacrylamide gel and perform the western blot analysis.

.

2.8 GEPIA (Gene Expression Profiling Interactive Analysis) and ReMap2022.

GEPIA2 (GEPIA 2 (cancer-pku.cn)) is a web server that was initially introduced as GEPIA in 2017 and has since been updated to GEPIA2. This webserver utilizes data from TCGA and GTEx project database to perform gene expression analysis in tumor vs normal samples. GEPIA2 can be used to load a cancer type to be analysed. In this research it was PRAD (Prostate Adenocarcinoma) and I looked for gene expression in tumor vs normal samples. In this study, GEPIA2 was to look for gene expression of *NOVA1, ESRP1,* and *ESRP2* in prostate adenocarcinoma and generate the box plots to compare the levels of *NOVA1, ESRP1,* and *ESRP2* genes in normal tissues.[139]

Additionally, ReMap2022 (ReMap2022 (univ-amu.fr)) is an online resource that has quality-controlled ChIP-seq datasets to be used for analysis. In this study, ReMap2022 was used to download the androgen receptor binding tracks to load as custom tracks on to UCSC and identify the AR binding sites in relation to the *NOVA1* gene.[140]

2.9 Bioinformatic analysis of NOVA1 RNA-Seq data

The bioinformatics analysis was done by Dr. Sara Luzzi on my data. The RNAseq data obtained from the knockdown of *NOVA1* in PC3 cell lines was first checked for quality. Then, Salmon was used for the read alignment, and DESeq2 was used to carry out differential gene expression analysis.[141,142] Additionally, to identify differential splicing, SUPPA2 was used on the RNAseq data to identify the events with splicing.[143] Further, to identify biological pathways that are potentially regulated by the NOVA1 proteins Gene Ontology (GO) analysis was carried out on the list of targets that showed upregulation and downregulation.[144] This was done by Dr. Sara Luzzi from the NUBI (Newcastle University Biosciences Institute).
Additional analysis was then done using Psichomics, which is an R programming language-based tool .[145] R is a programming language designed for statistical analysis of large datasets. Psichomics (Bioconductor - psichomics) is a library package within R that initiates a graphical user interface in a browser to analyse alternative splicing and gene expression data and visualize it.[146]

2.10 Cell colony forming assay.

The PC3 cells were treated with NOVA1 siRNA and negative control siRNA. The knockdown was performed in three replicates and three replicates of negative control in a 6-well plate. Then, each replicate was loaded into the six wells of a 6-well plate each. The cells were seeded with a count of 200 cells per well using haemocytometer. This gave us six 6-well plates, three for knockdown and three for negative control. The well plates were incubated for 14 days in

incubator at 37°C and 5% CO_2. RPMI media was changed every 7 days for each of the six well plates. On the fourteenth day, RPMI media was removed, and wells were washed with PBS. Then PBS was discarded and 2ml of ice cold 100% methanol was loaded per well and incubated for 10 minutes in a tissue culture hood at room temperature. The wells of each well plate were then washed with PBS, PBS was then discarded, and 2ml per well PBS was then added again. Then, 2ml of crystal violet were added per well and incubated for 10 minutes at room temperature. The crystal violet was then removed and well plates were rinsed with the water to remove any excess stains left. The manual counting of the colonies was done once the plates had dried and the count was added to the pre-available excel sheet and the statistical analysis was then performed, and bar graphs were created using GraphPad PRISM.

Chapter 3: Identifying the relationship between NOVA1 expression and androgen signalling.

3.1 Introduction.

In patients with low-risk prostate cancer active surveillance of the disease is recommended [74,147]. As the disease risk increases radical prostatectomy and radiation therapy are used to treat the disease [147,148]. Circulating androgens control patterns of gene expression within prostate cancer cells that drive growth of the primary tumour. Androgen Deprivation Therapy (ADT) has long served as a standard treatment for recurrent or metastatic prostate cancer [147–150]. However, the disease ultimately progresses to a state known as Castrate Resistant Prostate Cancer (CRPC). This critical juncture of disease progression occurs when prostate cancer cells can grow in even low concentrations of androgens after ADT and leads to mortality. Therefore, it is of crucial importance to understand the effect of androgen regulated genes in prostate cancer. Androgen receptor inhibitors are used on cell lines grown in tissue culture laboratories to block or decrease androgen activity in prostate cancer cells and study downstream changes. Casodex works by inhibiting androgen receptors from binding to AR [48,151].

As researchers delved into the study trying to understand androgen regulation and its downstream effects, a plethora of genes controlled by androgen signalling began to emerge. In a study conducted by Munkley et al. in 2016, a novel exploration was initiated. This research employed RNA sequencing (RNAseq) to carefully examine the effects of androgen stimulation on cultured cells and the consequences of androgen deprivation in prostate cancer patients. Remarkably, this investigation unveiled a cohort of approximately 700 genes that underwent significant alterations in response to androgen stimulation and the reverse affect in patients after androgen [47]. Subsequently, in 2019, Munkley et al. conducted a follow-up study that unveiled a robust association between ESRP1, ESRP2 and androgen regulation[48]. Androgen exposure activated ESRP2 expression. However, Munkley et al. study of 2016 unveiled several genes that exhibited a contrasting expression pattern - being repressed by androgens [47]. Among these genes, one standout candidate with a marked inverse regulation to ESRPs and androgen signalling was the *NOVA1* gene [47].

In order to study prostate cancer under laboratory settings, various cell lines of prostate cancer are used with each cell line having its unique androgen receptor status. This distinction of AR status in cell lines makes them suitable to understand conditions the effects of androgen signalling as shown in Table 1 below. These cell lines can be categorized in two major categories. The first category is the cell lines that have AR positive status. The cell lines that have AR positive status are LNCaP, CWR22RV1, MDA pCa 2b, LAPC-4. The second category are AR negative. AR negative cell lines include DU-145 and PC3. In the study of prostate cancer described in this thesis, I used different cell lines to understand regulation, with each having its own qualities of androgen expression, but particularly LNCaP and PC-3 listed in Table 1 [152]. These cell lines are used in the laboratory as they are easier to grow in a tissue culture lab and represent different stages of prostate cancer ranging from LNCaP that models prostate adenocarcinoma that responds to androgen deprivation therapies, to PC3 cells that represent aggressive cancer which does not respond to androgen therapies. Additionally, I did western blotting to look at protein expression in samples.

Table 3.1: Table below shows the cell lines in the left column and androgen receptor expression status for each cell line in the right column.

Cell Line	*Androgen Receptor Expression*
LNCaP	Yes
MDA pCa 2b	Yes
LAPC-4	Yes
CWR22Rv1	Yes
DU145	AR Negative
PC-3	AR Negative

In this chapter, I describe experiments to embark on a comprehensive exploration of *NOVA1* gene regulation in response to both androgen stimulation and steroid depletion. My primary objective was to establish a conclusive link between *NOVA1* and androgen receptor signalling. I achieved this by looking at the timed effect of androgen stimulation and deprivation as well

as the expression levels of NOVA1 protein in different cell lines that all have different androgen profiles as mentioned in table 1. Additionally, I looked at public data sets to understand *NOVA1* gene expression in normal and tumour tissues, and at the relationship between the *NOVA1* gene and AR binding sites.

3.2 Aims of Chapter

Although identified as a candidate gene to be repressed by androgens, no further follow up analysis had been carried out on the *NOVA1* gene in prostate cancer. The aim of this Chapter was to validate androgen regulation of the *NOVA1* gene.

My specific aims were:

1. To monitor how NOVA1 gene expression changes after androgen stimulation. I analysed NOVA1 gene expression over a time-course in LNCaP cells after androgen stimulation.
2. To monitor if NOVA1 protein changes in response to androgens.
3. To monitor NOVA1 protein expression change in androgen receptor knockdown LNCaP cells and controls, to test the prediction that NOVA1 has higher expression in absence of AR.
4. To monitor NOVA1 protein expression in different prostate cancer cell lines based on their androgen receptor profiles.
5. To compare the GTEx project data against TCGA data for Prostate Adenocarcinoma to look for *NOVA1* gene expression in normal vs diseased tissue [153–155].
6. To analyse publicly available AR binding sites and their link to *NOVA1* gene location on the genome using UCSC browser [156].

3.3 Results

3.3.1 *NOVA1* has higher mRNA expression in steroid deplete LNCaP cells.

The Munkley et al. 2016 study generated a list of genes that changed in response to androgens, where *ESRP2* had a reciprocal expression between androgen stimulation and androgen deprivation.[47] In the same dataset the *NOVA1* gene also was identified to be regulated inversely in presence of androgens. As *NOVA1* was inversely regulated to ESRP2 protein it is likely repressed by androgens and the AR.

The 2016 Munkley et al paper identified *NOVA1* as a reciprocally changing gene in response to androgen stimulation, but this had not been independently verified as yet [47]. Therefore, in order to verify the observation by Munkley et al. 2016, I first set out to monitor *NOVA1* expression in steroid deplete and androgen positive conditions. LNCaP cells were grown in tissue culture and split into two separate flasks as shown in Figure 3.1B. To assess the effect of androgen stimulation on gene expression, RT-qPCR was performed on RNA extracted from cells treated with either steroid-depleted media or R1881-supplemented media, as described in the Materials and Methods. The resulting data were normalized using the average CT values of actin, tubulin, and GAPDH as internal controls, and relative gene expression levels were calculated using the 2^−ΔΔCT method. A bar graph summarizing these results is presented in Figure 3.1.

These data were analysed using a bar graph showing the relative expression change of NOVA1 in steroid deplete cells compared to androgen positive LNCaP cells (Figure 3.1). The primary objective of this graph is to provide a clear and quantitative representation of the changes in NOVA1 expression between these two experimental conditions. Figure 3.1A showed that there is a substantial difference in NOVA1 expression between the two experimental groups. Specifically, the bars corresponding to the "Steroid Deplete" condition exhibit a significantly higher value compared to the bars representing the "Androgen Positive" condition. This indicates that, on average, NOVA1 expression is considerably elevated in the cells subjected to steroid depletion when compared to the androgen-positive control cells.

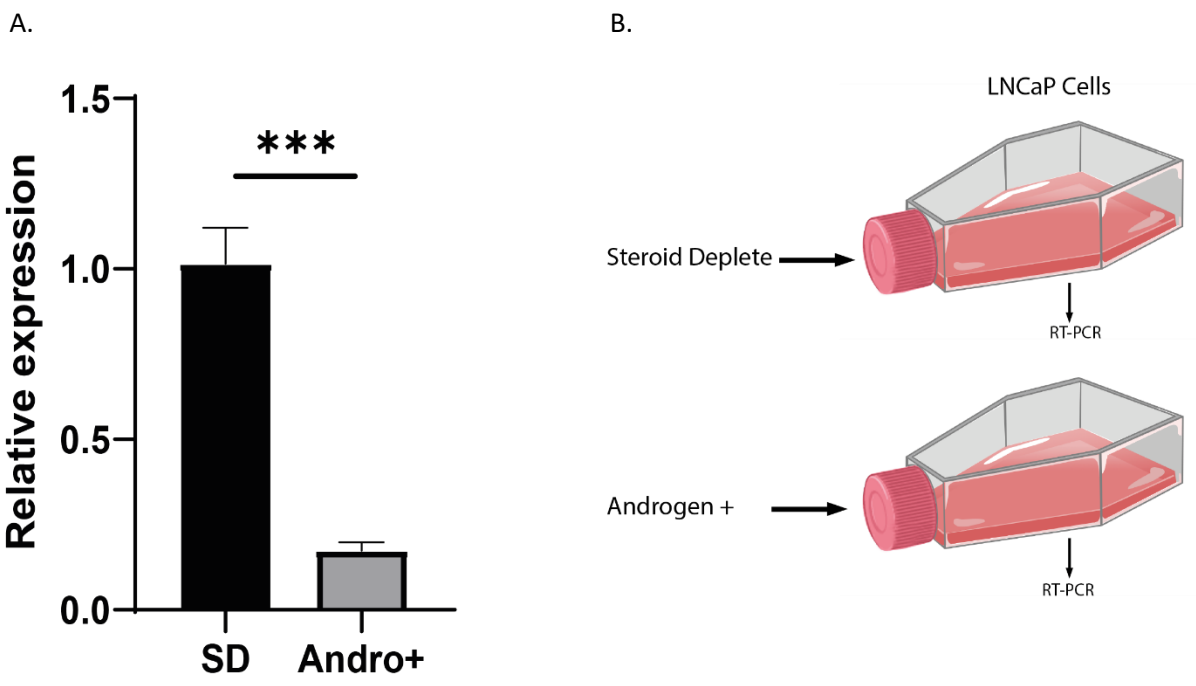


Figure 3.1: Relative expression of NOVA1 from qPCR LNCaP cell lines comparing steroid deplete vs androgen positive samples. *The horizontal axis of the graph separates the two experimental conditions: "Steroid Deplete" and "Androgen Positive." The vertical axis represents the relative expression change of NOVA1, which is typically expressed as a fold change. Each condition is represented by a set of bars, with one bar indicating the relative expression change in NOVA1 for that condition.*

The observed elevation in *NOVA1* mRNA expression in response to steroid depletion underscores the relationship between androgen signalling and NOVA1 regulation. These findings suggest that the addition of androgens compared to steroid deplete LNCaP cells leads to a significant down-regulation of *NOVA1* mRNA expression in LNCaP cells by 72 hours.

3.3.2 *NOVA1* has higher expression levels in steroid deplete LNCaP cells

I next carried out more detailed experiments to monitor the dynamics of how *NOVA1* gene expression responds to androgen exposure, by measuring the androgen response over time. There were two cell groups. One cell group containing LNCaP cells was maintained in a steroid deplete medium whereas second cells group was stimulated with androgen using R1881 media. Cells were harvested at different time point beginning at zero hours and ending at twenty-four hours.

Figure 3.2 presents this time course analysis of *NOVA1* mRNA expression in two distinct groups of LNCaP cells: one group maintained in steroid deplete medium (depicted by the blue line) and the second group treated with androgens (depicted by the green line). The primary objective of this analysis was to track how *NOVA1* gene expression levels change over time when cells are exposed to androgens, in contrast to being grown in steroid deplete medium. The results of this experiment shown in Figure 3.2 confirmed that *NOVA1* mRNA expression was repressed by androgen exposure in comparison to steroid deplete medium. Thus, I concluded that *NOVA1* gene expression is rapidly repressed by androgens.

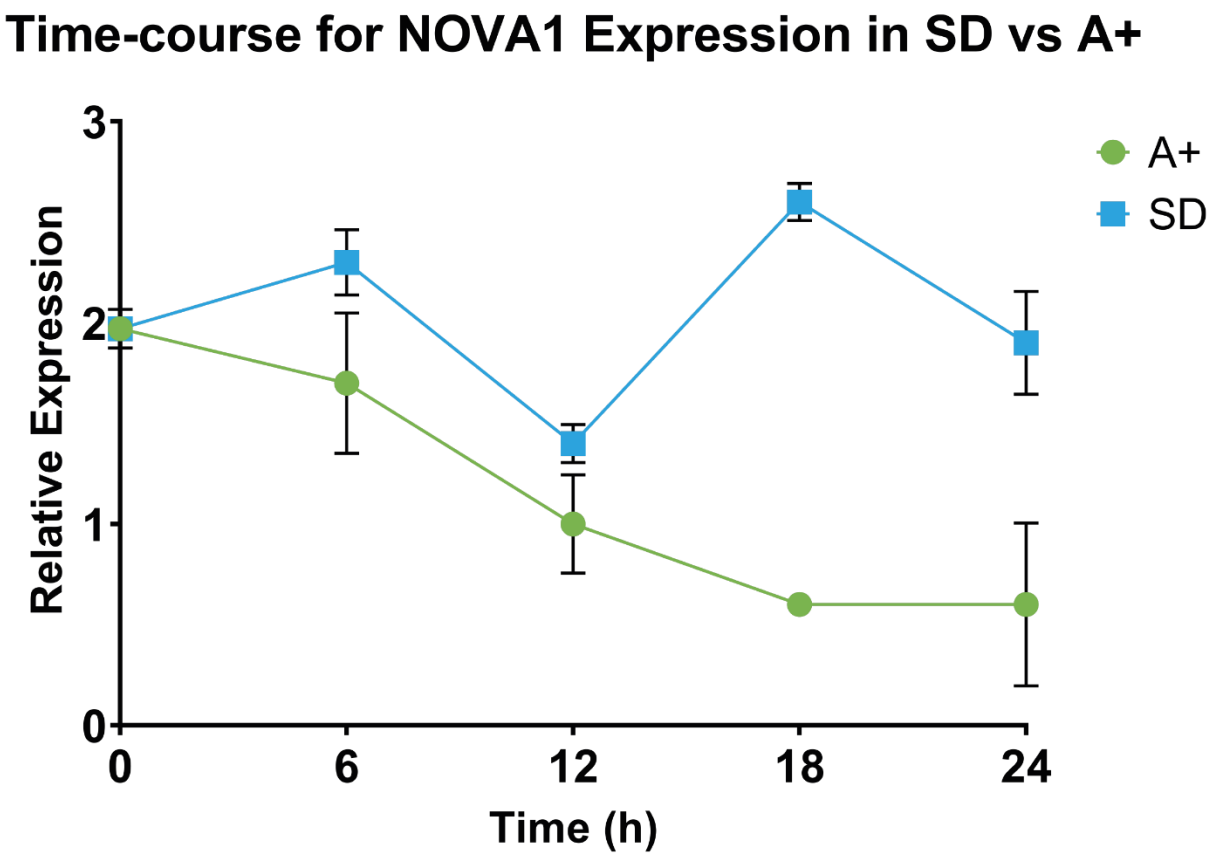


Figure 3.2: Quantitative analysis (real-time PCR). *qPCR of NOVA1 mRNA expression over a 24-hr time course following androgen exposure. RT-qPCR graph for LNCaP cells showing NOVA1 expression levels in steroid deplete vs androgen positive samples treated across 0 hours to 24 hours. The horizontal axis of the graph represents time in hours, ranging from zero hours (initial treatment) to 24 hours. The vertical axis represents relative NOVA1 protein expression levels.*

3.3.3 *NOVA1* expression is increased by Casodex treatment.

To further understand the connection between *NOVA1* gene expression and androgen regulation, I set out to see if there were any changes in steroid deplete vs androgen positive samples when they were additionally treated with Casodex in either of the two conditions. The hypothesis for this experiment was that if *NOVA1* is repressed by the AR, then Casodex treatment should enhance the *NOVA1* mRNA expression levels. To evaluate the impact of androgen presence and Casodex treatment on gene expression, LNCaP cells were cultured under different treatment conditions, as outlined in the Materials and Methods and illustrated in Figure 3.3B. RT-qPCR was performed on RNA extracted from these samples to assess differential expression levels.

The results from this RT-qPCR are shown in Figure 3.3A, a bar graph that illustrates the results of the RT-qPCR (quantitative polymerase chain reaction) analysis of RNA samples purified from LNCaP cells, comparing the levels of *NOVA1* mRNA expression in various treatment conditions for both steroid-depleted (SD) and androgen-treated (A+) LNCaP cells shown in Figure 3.3B. The horizontal axis of the graph distinguishes between the different treatment conditions for both steroid-depleted and androgen-positive cells. The vertical axis represents the levels of *NOVA1* mRNA expression, quantified as relative expression fold change. This fold change was related to the loading controls (*Actin*, *Tubulin*, and *GAPDH*). Each loading control had triplicate samples; initially the average for these triplicates was calculated. Then, the average of all loading control averages was taken as a standard. Each condition is represented by a set of bars, with each bar indicating the relative *NOVA1* expression levels in that condition. Error bars shown in the Figure 3.3A show the standard error, and a p-value was calculated using excel formulae.

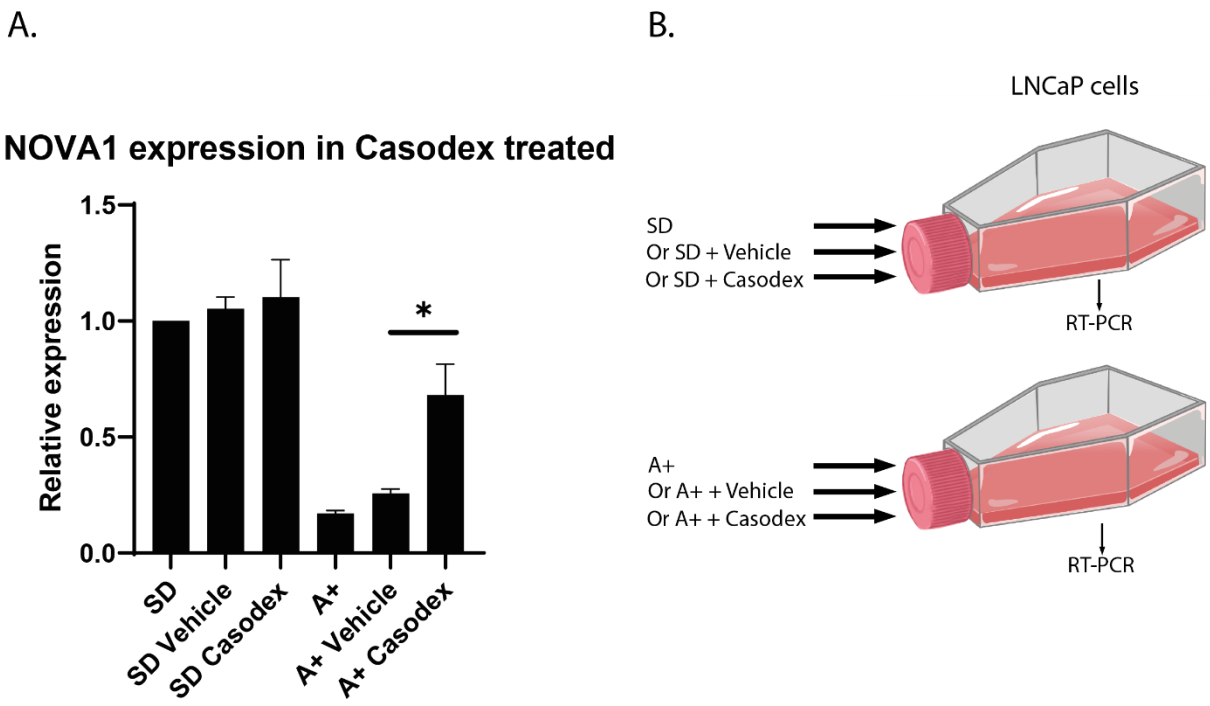


Figure 3.3: NOVA1 qPCR expression in different treatment condition in LNCaP cell line samples. ***A.*** *qPCR fold change has been quantified and is shown here. An * in the figure corresponds to p-value significance.* ***B.*** *Tissue culture flasks showing different treatment conditions. The LNCaP cells were treated with steroid deplete media represented by SD, or with androgen R1881 media shown as A+. SD Only: Steroid-depleted LNCaP cells without additional treatment; SD + Vehicle (DMSO): Steroid-depleted LNCaP cells treated with a vehicle control (DMSO); SD + Casodex: Steroid-depleted LNCaP cells treated with Casodex, an androgen receptor blocking drug; A+: Androgen treated LNCaP cells; A+ Vehicle (DMSO): Androgen treated LNCaP cells that were further treated with a vehicle control (DMSO); A+ Casodex: Androgen treated LNCaP cells were further treated with Casodex.*

Upon examining Figure 3.3A, several significant observations become evident: (1) Steroid Depletion Effect: The bars corresponding to the "SD Only" condition for steroid-depleted cells show notably higher NOVA1 expression compared to the androgen-positive conditions. This indicates that steroid depletion causes a significant upregulation of *NOVA1* expression. (2) Impact of Casodex: Among the androgen-positive conditions, the bars representing "Androgen Positive + Casodex" exhibit higher NOVA1 expression levels compared to the other androgen-positive treated samples. This shows that blocking the androgen receptor with Casodex results in an increase in NOVA1 expression but still at a lower level than observed in steroid-depleted cells. (3) Maximum Expression: The highest *NOVA1* expression is observed

in the "SD + Casodex" condition for steroid-depleted cells. This condition demonstrates the highest *NOVA1* expression, surpassing all other conditions in the graph.

In summary, Figure 3.3A provides a comprehensive view of how different treatments of LNCaP cells by Casodex and vehicle (DMSO) impact NOVA1 expression in both steroid-depleted and androgen-positive LNCaP cells. It reinforces the prediction that NOVA1 expression would be higher in response to steroid depletion, and that blocking the androgen receptor with Casodex can partially increase NOVA1 expression in androgen-treated cells. This figure adds to the evidence that androgen signalling does repress NOVA1 expression.

3.3.4 NOVA1 protein is repressed by androgens.

The above analysis only looked at *NOVA1* mRNA expression and did not analyse the NOVA1 protein. I thus next wanted to confirm these changes based on the protein samples of LNCaP cells that had been depleted of steroids and compare them to LNCaP cells that had been grown in synthetic androgen media.

My next aim was to compare NOVA1 protein expression after androgen stimulation. In order to examine protein levels, protein expression under androgen-depleted and androgen-stimulated conditions, LNCaP cells were cultured and treated as described in the Materials and Methods. Protein extracts prepared after 72 hours were analyzed by SDS-PAGE and western blot as shown in figure 3.4A. The prediction for the above analysis at the RNA level was that addition of synthetic androgens should block NOVA1 protein expression. To test this prediction, western blots were probed for NOVA1 protein. The western blot X-ray film was developed which was scanned into a JPEG format image using a scanner. The scanned image of the western blot of the protein sample and the loading control were analysed using a software called FIJI or ImageJ [157]. This software was used as detailed in Chapter 2 and captures the band intensities which are used to quantify the ratios of western blot bands visible in image 3.4A. This used the calculation system described in Chapter 2. The calculated values are in arbitrary units (AU) which were then analysed using GraphPad PRISM 9, that performed the statistical analysis of the values and generated the graph as shown in figure 3.4B.

In Figure 3.4A, I present a comparative analysis of NOVA1 protein expression levels in two distinct sets of LNCaP cell lines: those cultured in steroid-deplete media and those maintained in media with a synthetic androgen R1881 environment. The primary objective of this analysis was to ascertain whether the presence or absence of androgens had any discernible impact on NOVA1 protein expression within these cell lines.

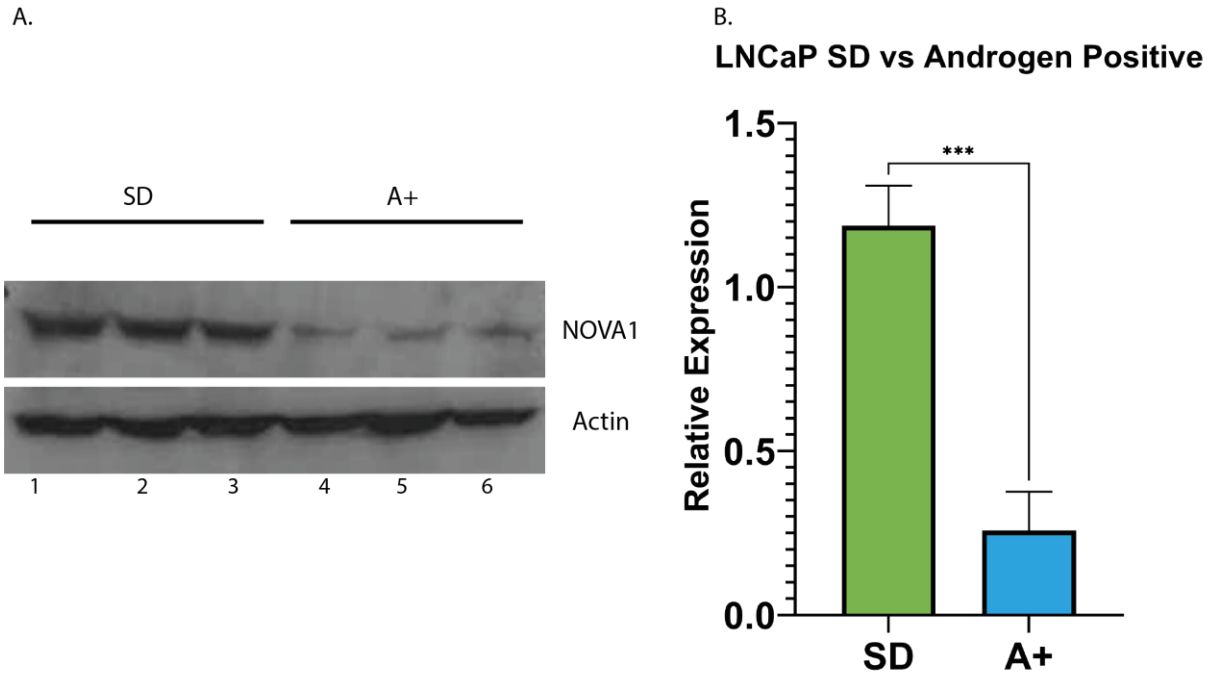


Figure 3.4: Western blot of NOVA1 protein in steroid deplete vs androgen positive treatment of LNCaP cells for 72 hours. *A) Original Western blot image after ECL detection on photographic film. B) Graph shows the quantified western blot data. Graph is created using PRISM GraphPad. The horizontal axis of the graph in figure 3.4B represents the two experimental conditions: "Steroid Deplete" and "Androgen Positive (A+)," while the vertical axis corresponds to NOVA1 protein expression levels, which are quantified using J Viewer, with the mean calculated in Excel and plotted using GraphPad PRISM9.*

Upon careful examination of Figure 3.4A and Figure 3.4B, it is readily apparent that NOVA1 protein exhibits contrasting expression patterns in response to the two different conditions. Specifically, NOVA1 protein expression is noticeably elevated in the "Steroid Deplete" condition compared to the "Androgen Positive (A+)" condition. Lanes 1-3 of Figure 3.4A show the steroid depleted LNCaP cells have a stronger band for NOVA1 protein expression, indicating higher presence of NOVA1 in these samples. In contrast, lanes 4-6 of Figure 3.4A contain protein from the androgen treated LNCaP cells where NOVA1 bands are thinner and

barely visible. This indicates that in presence of androgens NOVA1 protein is significantly downregulated, thus showing an inverse connection between androgen and NOVA1. This substantial difference in protein expression levels between the two conditions suggests a robust androgen-dependent regulatory mechanism governing NOVA1 expression in LNCaP cell lines. Figure 3.4B shows a graph where triplicate samples of LNCaP cells grown under steroid deplete conditions or after androgen stimulation have been averaged. It can be clearly seen that NOVA1 protein expression is much higher in steroid deplete and is statistically significant. Analysis of these data sets using a t-test confirmed that this result was statistically significant.

3.3.5 AR knockdown increases levels of NOVA1 protein expression.

The above section showed that NOVA1 protein expression was controlled by androgens. I next tested if NOVA1 protein was also controlled by the androgen receptor. To test this, I analysed NOVA1 protein expression in control LNCaP cells, and LNCaP cells after AR depletion using siRNA. The primary objective of this analysis was to investigate the influence of Androgen Receptor knockdown on NOVA1 protein expression, thus providing further evidence of androgen's impact on NOVA1 regulation.

Knockdown of the AR was achieved in LNCaP cells grown in tissue culture. The cells were then split and moved to a six well plate for 24 hours, after which an siRNA specific to the AR was used to knockdown AR in the LNCaP cells or treated with a control siRNA as detailed in chapter 2. The cells were harvested after 72 hours of siRNA treatment and used to prepare protein sample to carry out the western blotting analysis as shown in Figure 3.5A. This western blot was probed for NOVA1 protein, developed by ECL and then quantified and plotted with the help of ImageJ, Microsoft Excel, and GraphPad PRISM9 [157–159].

Figure 3.5A, presents the results of this Western blot analysis comparing NOVA1 protein expression levels in two distinct sets of LNCaP cell lines: those with an Androgen Receptor (AR) knockdown (shown in lanes 4-6 of Figure 3.5A) and those comprising the control siRNA treated LNCaP cell line (lanes 1-3 of Figure 3.5A). The Western blot shows bands corresponding to NOVA1 protein expression. A comparative analysis of these bands revealed

a contrast in NOVA1 protein expression patterns. Notably, there is a marked increase in NOVA1 protein expression in the Androgen Receptor knockdown samples as seen in lanes 4-6 Figure 3.5A, as evidenced by the more intense bands, when compared to the control LNCaP cell line. This outcome strongly indicates that the Androgen Receptor represses NOVA1 protein levels.

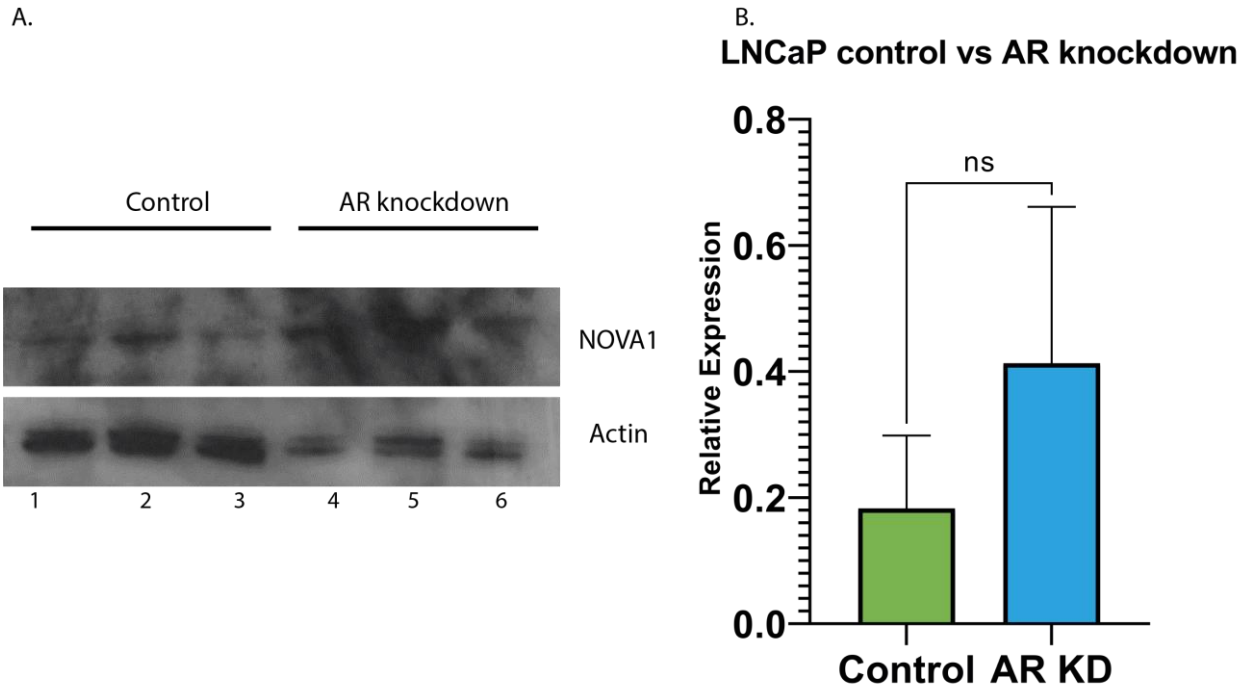


Figure 3.5: Western blot of NOVA1 protein in control vs Androgen Receptor knockdown in LNCaP cells for 72 hours. *A) Western blot image of NOVA1 protein expression detected by ECL. B) Graph shows the quantification of this western blot data. The graph was created using PRISM GraphPad. The horizontal axis of the graph delineates the two experimental conditions: "Control" and "Androgen Receptor Knockdown," each condition is represented by a distinct bar, and the height of each bar corresponds to the quantified NOVA1 protein expression levels.*

Quantification of this protein expression data are shown in Figure 3.5B. This shows a bar graph that quantifies the NOVA1 protein expression levels observed in the Androgen Receptor knockdown LNCaP cell lines in comparison to the control LNCaP cell line. This graph provides a numerical representation of the protein expression data depicted in Figure 3.5A. This graph was obtained similarly as explained in section 3.3.4. The greyscale scanned image of the western blot was loaded in Image J application [157] which gave value for the band intensities for each lane. These values were collected thrice for each band for samples as well as loading

control. Then using excel, the average of reads for each band was calculated and a ratio of sample bands and their respective control bands was taken which was then used to determine the p-value (default excel formulae) and to construct the bar graphs as shown in figure 3.5B. The excel values were used in GraphPad PRISM with p-value test value from excel manually added to the GraphPad PRISM.

As the graph clearly illustrates, NOVA1 protein expression is higher in the Androgen Receptor knockdown LNCaP cell lines when compared to the control group. This data reinforces the observation made in Figure 3.4A, providing evidence that the downregulation of the Androgen Receptor is indeed associated with a significant increase in NOVA1 protein expression. However, although there was an apparent decrease in NOVA1 protein expression on the Western blot, Figure 3.5B shows this difference to be non-significant. This could be due to the high background noise on the Western blot.

Collectively, Figures 3.1, 3.2, 3.3, 3.4 and 3.5 provide compelling evidence that the presence or absence of the Androgen Receptor plays an important role in the regulation of NOVA1 protein expression in LNCaP cell lines. These findings provide initial in-depth look into the role of androgens in modulating NOVA1 protein expression and lay the foundation for further investigation into the molecular mechanisms underlying this phenomenon.

3.3.6 NOVA1 is highly expressed in the PC3 cell line.

Now that I had examined the reciprocal effects of androgen stimulation and AR inhibition on NOVA1 gene and protein expression in LNCaP cells, I next analysed NOVA1 expression in a panel of prostate cancer cell lines. I wanted to identify a prostate cancer cell line that has the highest expression of NOVA1 to decide a suitable candidate to carry out NOVA1 knockdown, leading to a downstream RNA sequencing to study and understanding the downstream effects of NOVA1 expression. I predicted that AR-negative cell lines would have the highest expression of NOVA1. This was experimentally validated in different prostate cancer cell lines as they have a varying AR status as shown in table 1 and discussed in section 3.3.31. The cells were grown in tissue culture in RPMI media and were harvested without any additional cell treatments. Cells were then used to prepare protein samples that were used in western blotting.

Figure 3.6A displays the results of a Western blot analysis conducted on different prostate cancer cell lines, namely LNCaP, PC3, and DU-145. The primary objective of this analysis was to assess the relative levels of NOVA1 protein expression across these distinct cell lines.

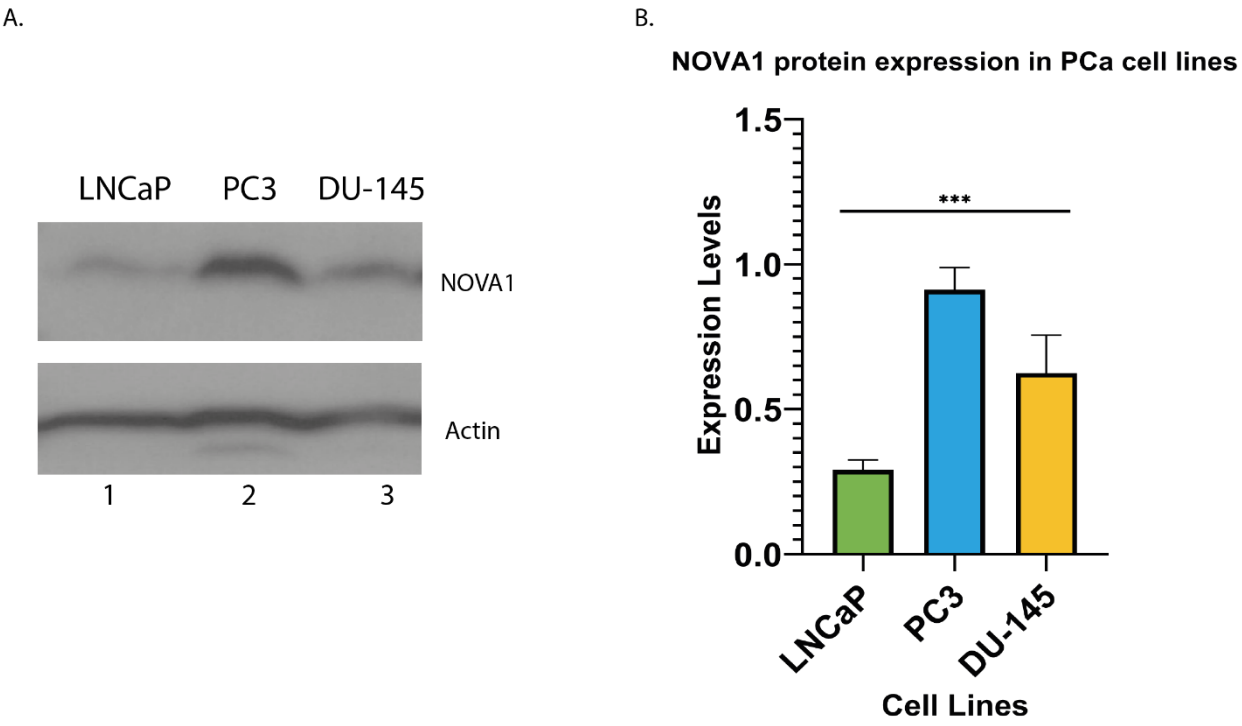


Figure 3.6: Western blot of NOVA1 protein in different cell lines (LNCaP, PC3, and DU-145). *A) It is a Western blot image in a greyscale. B) Graph shows the western blot data quantified. Graph is created using PRISM GraphPad. The horizontal axis of the graph distinguishes between the three cell lines, with each cell line represented by a distinct bar. The vertical axis represents NOVA1 protein expression levels. PC3 cell line shows highest NOVA1 protein expression among the cell lines examined.*

The Western blot image represented NOVA1 protein expression across these three specific cell lines. The presence and intensity of these bands are indicative of NOVA1 expression levels in the respective cell lines. Careful examination of Figure 3.6A, showed a striking disparity in NOVA1 expression across the three cell lines becomes evident. NOVA1 protein expression in LNCaP cell lines can be seen in lane 1 of Figure 3.6A. LNCaP cells are characterized as androgen-sensitive, exhibit the lowest levels of NOVA1 expression. This is depicted by the faint band corresponding to the LNCaP lane. This can also be seen in the bar graph shown in Figure 3.6B. PC3 cell lines, which are known for their aggressiveness and androgen independence, display the highest levels of NOVA1 expression [152]. This is evidenced by the prominent and intense band observed in the lane 2 of Figure 3.6A. DU-145 cell lines exhibit

moderate levels of NOVA1 expression, falling between the LNCaP and PC3 cell lines in terms of NOVA1 expression intensity. This is represented by the band in lane 3 of Figure 3.6A.
Figure 3.6B complements the Western blot results presented in Figure 3.6A by providing a quantitative representation of NOVA1 protein expression levels in the three prostate cancer cell lines: LNCaP, PC3, and DU-145. Upon examination of Figure 3.6B, it is evident that the bar graph corresponding to the NOVA1 protein expression in PC3 cell line is significantly higher than those representing the LNCaP and DU-145 cell lines. This indicates that PC3 cell lines indeed exhibit the highest levels of NOVA1 protein expression among the three cell lines. Conversely, the bar graph representing the NOVA1 protein expression in LNCaP cell lines is notably lower, reaffirming their status as having the lowest NOVA1 expression levels. DU-145 cell lines, while exhibiting higher NOVA1 expression than LNCaP, fall in between the other two cell lines in terms of NOVA1 protein expression.

These quantitative findings in Figure 3.6B underscore the suitability of PC3 cell lines as a candidate for subsequent NOVA1 gene knockdown and subsequent RNA sequencing analysis. The data presented in Figure 3.6A, and 3.6B clearly supports the rationale behind selecting PC3 as a cell line with high NOVA1 expression for further downstream research analysis. Furthermore, it also shows that AR positive cell lines have lower NOVA1 expression and cell lines that are AR negative have higher NOVA1 expression.

3.3.7 Does NOVA1 express highly in Neuroendocrine cell lines?

A study by Zhang et al. 2020 [160] published in nature communications brought in to light that NOVA1 was highly expressed in neuroendocrine cell lines in the NEPC (neuroendocrine prostate cancer) RNA-Seq data. Additionally, I was recently able to get CWR22 cell lines, I set out to test the NOVA1 protein expression in each of these cell lines to compare the expression as shown below in Figure 3.7.

The cell lines were grown in tissue culture in RPMI media and were harvested without additional treatment. Cell pellet was then used to prepare protein samples for each cell line. These protein samples were used for the western blot analysis.

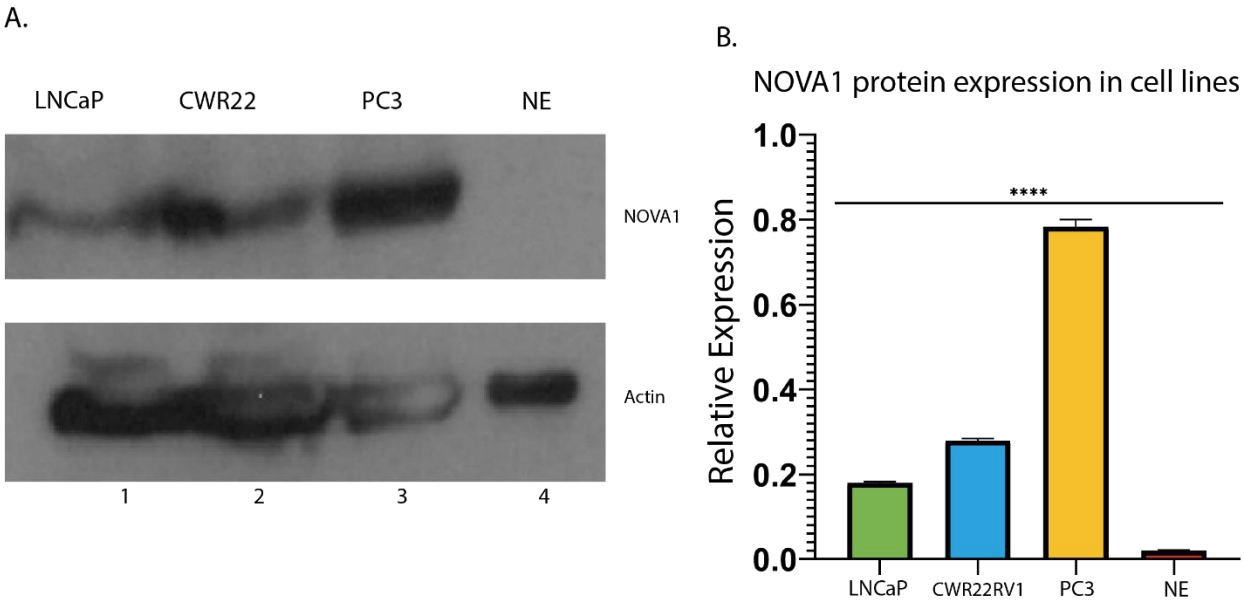


Figure 3.7: Western blot of NOVA1 protein in different cell lines*. A) It is a Western blot image in a greyscale. B) Graph shows the western blot data quantified. Graph is created using PRISM GraphPad. I examined more cell lines here including NE whereas, NE in lane 4 in 3.7A can be seen with no bands in NOVA1 sample. As previously seen, yet again this figure showed highest expression of NOVA1 in PC3 cells.*

In Figure 3.7A western blot figure is shown with cell line types marked at the top. It was seen that LNCaP cell had the lowest expression of NOVA1 protein whereas, PC3 cells had the highest expression which is consistent with the previous western blot analysis discussed in section 3.3.7. This can be visualized in bar graph in Figure 3.7B. Whereas, CWR22 cell lines in lane 2 of Figure 3.7A had expression higher than LNCaP cells but lower than PC3 cells. This further solidifies the evidence based for choosing PC3 cell lines as suitable candidate for *NOVA1* gene knockdown and downstream RNA-Seq analysis. Interestingly, I was expecting to see even higher expression of NOVA1 levels in neuroendocrine prostate cancer cell lines compared to PC3 cells but as it turned out that I did not see a band in lane 4. Whereas, loading control for the same revealed the presence of sample and that it was not an inappropriate loading in the lane. This could be due to several reasons.

3.3.8 The *NOVA1* gene has nearby direct AR binding sites.

Based on my work in previous sections I found *NOVA1* to be affected by androgens inversely which makes it possible that AR might have binding sites close to *NOVA1* in genome and therefore be exerting its effects on *NOVA1*. Hence, I started looking at UCSC genome browser to identify any binding sites for AR closer to *NOVA1*. In order to achieve this, I got iClip data for AR binding sites from Munkley labs previous work. I loaded the tracks on UCSC browser as shown in figure 3.8.

Figure 3.8 presents a visual representation of the comparison between Androgen Receptor (AR) binding sites and the NOVA1 gene, as observed in the UCSC (University of California, Santa Cruz) genome browser. These are iClip tags for AR data based on Munkley labs previous work. This analysis serves to elucidate the relationship between AR binding and the NOVA1 gene's genomic location. The figure underscores the significance of AR binding in the context of NOVA1 gene regulation.

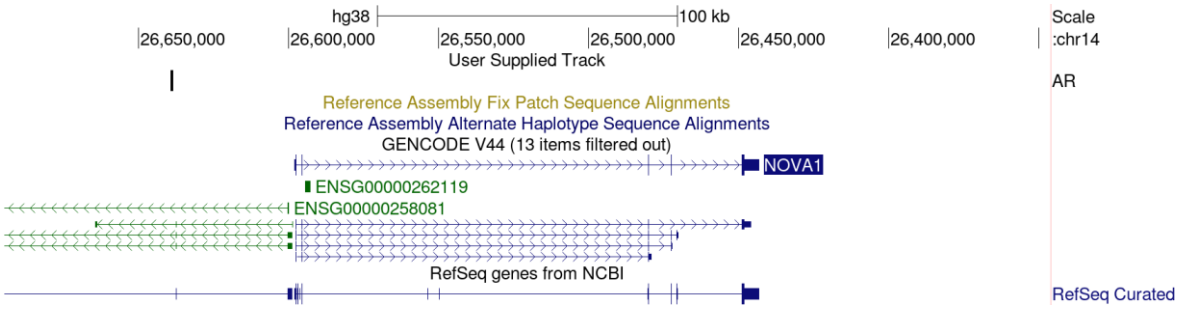


Figure 3.8: Hg.38 human genome assembly. *NOVA1 can be seen along with AR tracks loaded by custom tracks from iClip file from Munkley lab. In this track it could be seen that AR binding sites appear at a distance from NOVA1 gene binding sites.*

The UCSC browser image showcases genomic tracks that include AR binding sites, with particular emphasis on their proximity and overlap with the NOVA1 gene. AR iClip tags in Figure 3.8 initially suggested that AR binding sites were located at a considerable distance from the NOVA1 gene.

However recently, I came across a new Database 'ReMap' with help from Dr. Louis Renard.[1,2] This database has iClip tags for transcription factors including AR and has been extensively

done. In order to enhance the precision and accuracy of the analysis, the study utilized data from the ReMap database, which provides up-to-date information on transcription factor binding sites. The latest AR binding site tracks from ReMap were loaded to UCSC Custom Tracks[161].

The ReMap tracks loaded in UCSC genome browser assembly GRCH 38 are shown in Figure 3.9.[1,3] The graphical representation on the UCSC browser illustrates the relative positions of AR binding sites in relation to the *NOVA1* gene. Notably, the figure highlights that AR binding sites overlap with the *NOVA1* gene, indicating that the Androgen Receptor directly interacts with the genomic region associated with *NOVA1*.

The observed overlap of AR binding sites through the *NOVA1* gene serves as compelling evidence of a direct regulatory connection between the Androgen Receptor and *NOVA1*. This finding significantly strengthens the hypothesis that androgen signalling plays a pivotal role in the regulation of the *NOVA1* gene.

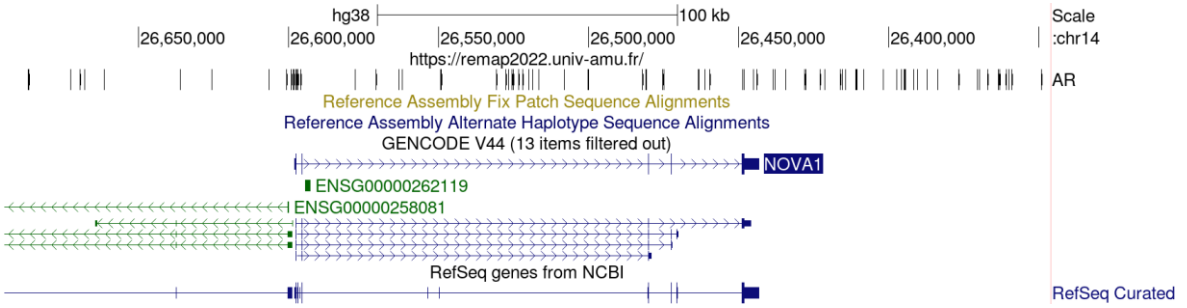


Figure 3.9: ReMap tracks for AR binding sites loaded as custom tracks to hg.38 assembly. *Recent data shows that AR binding sites overlap the NOVA1 gene establishing a connection between NOVA1 gene and AR.*

The figure also highlights that this new analysis using ReMap data represents a substantial improvement over the previous AR Clip tags data. The enhanced precision and granularity of the ReMap data contribute to a clearer understanding of the spatial relationship between AR binding and the *NOVA1* gene.

In conclusion, Figure 3.9 provides a visual representation of the physical proximity and overlap between AR binding sites and the *NOVA1* gene, reinforcing the connection between

androgen signalling and *NOVA1* regulation as initially hypothesized. This updated analysis, utilizing the ReMap database and UCSC browser, enhances the evidence supporting the regulatory relationship between the Androgen Receptor and the NOVA1 gene.

3.3.9 *NOVA1* has low expression in prostate adenocarcinoma tissue samples.

Based on the NOVA1 expression connection to AR I wanted to understand regulation in prostate cancer patients. Therefore, I utilised the TCGA available dataset for PRAD (Prostate Adenocarcinoma) and compared against the prostate tissue normal from the GTEx project. This was to look into interplay between NOVA1 and ESRP1 and ESRP2 which have been previously seen to be expressed in prostate cancer and suppressed after ADT treatment until the disease relapses[162]. I hypothesized that as NOVA1 is regulated opposite to that of ESRP1 and ESRP2, it would have lower expression in PRAD whereas, I suspected it to have a higher expression in aggressive or metastatic prostate cancer.

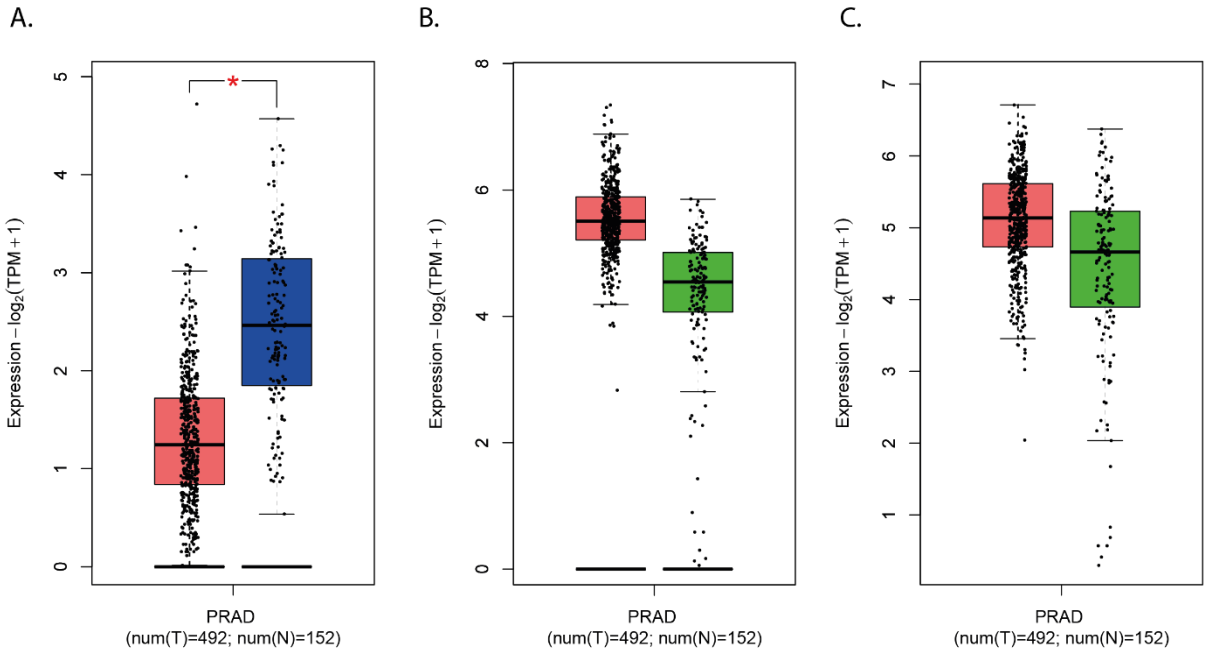


Figure 3.10: Comparative Box Plots of NOVA1 Gene Expression in Prostate Tissue. *Boxplots show the comparison of gene expression in normal vs prostate cancer tissue. This data is based on TCGA (The Cancer Genome Atlas) and GTEx (Genotype Tissue Expression) project. A) NOVA1 gene expression is shown to be lower in tumour samples (red) compared to normal tissue (blue). B) Shows ESRP1 gene expression which is higher in tumour samples (red) and lower in normal samples (green). C) ESRP2 gene expression levels can be seen in tumour samples (red) to be higher than normal samples (green).*

I utilized GEPIA2 (Gene Expression Profiling Interactive Analysis) webserver to load prostate adenocarcinoma samples from the cancer genome atlas database and compare them against GTEx project database [139].

I loaded the GEPIA2 webserver http://gepia2.cancer-pku.cn/ and opted for a box plot results for my analysis comparison. NOVA1, ESRP1, and ESRP2 were respectively used as genes of interest with a Log2FC cut-off of 1 and p-value cut-off at 0.01. PRAD cohort from TCGA was chosen as a database for comparison to GTEx data.

Box plot results for each gene are shown in Figure 3.10. Figure 3.10A shows a box plot for *NOVA1* where red box represents *NOVA1* gene expression in tumour samples from TCGA and blue box shows the *NOVA1* expression levels in normal prostate tissue from GTEx data. Furthermore, Figure 3.10B shows tumour samples in red and normal prostate tissue in green representing *ESRP1* gene expression in each of the two samples. Whereas. Figure 3.10C shows box plot for ESRP2 gene where red indicates the *ESRP2* expression in tumour prostate tissue and normal prostate tissue expression of *ESRP2* is shown in green.

From Figure 3.10 comparison of the genes, it can be seen that *NOVA1* gene has lower expression in tumour tissue for PRAD compared to normal tissue while, *ESRP1* and *ESRP2* have a higher expression in diseased prostate tissue compared to the normal prostate tissue. Thus, results are according to the initial hypothesis. The second part of hypothesis could not be verified where I considered NOVA1 expression to be higher once the CRPC cancer had relapsed after initial treatment with ADT leading to disease remission. This was due to lack of publicly available datasets or samples.

3.3.10 Survival analysis of NOVA1 correlates to GEPIA2 results.

I further wanted to investigate the clinical relevance of NOVA1 gene in patient survival relation. In order to achieve this, I utilized Psichomics an R based graphical application that can load TCGA data to perform different analysis on samples in a user-friendly interactive interface as discussed in chapter 2.

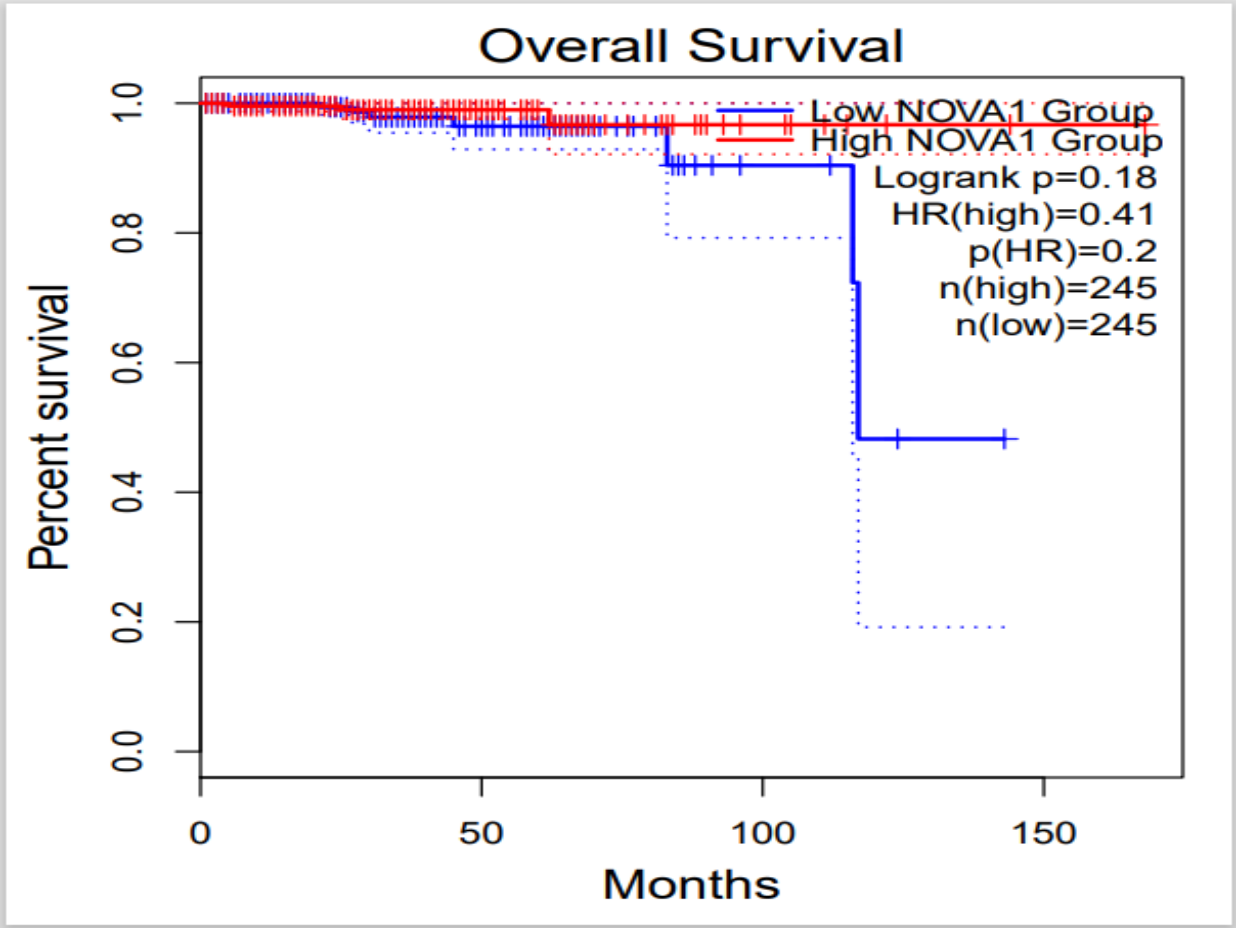


Figure 3.11: NOVA1 survival analysis. *Post-disease patient survival is shown in months on the x-axis and patient percent survival shown on y-axis. Red line shows higher NOVA1 expression and has higher patient survival as seen. Whereas blue line in the graph shows lower expression of NOVA1. In patients with lower expression of NOVA1, survival chances over time also decrease.*

Using Psichomics I loaded the TCGA PRAD dataset to perform a survival analysis for patients with different expression levels of the *NOVA1* gene as shown in Figure 3.11. Figure 3.11 shows the survival analysis performed for *NOVA1* gene, where survival with disease in months is shown on X-axis and percent of patient survival is shown on Y-axis. Furthermore, red line in the graph shows higher *NOVA1* expression which has higher chances of patient survival compared to low expression of *NOVA1* gene show in blue line which shows that lower NOVA1

gene expression can lead to lower patient survival. This shows an intricate relationship of NOVA1 gene with patient survival which develops more interest to understand the actual effects of *NOVA1* gene leading to patient survival chances. Also to note that the p-value in Figure 3.11 is not significant that can be sue to a smaller number of samples available. More importantly, the dataset had only four metastatic samples which can be a reason for a p-value that is not significant.

3.4 Discussion

In this chapter I observed that NOVA1 expression is inversely proportional to the androgen expression. In the steroid deplete environment which does not have androgen presence there was a higher expression of NOVA1 and vice versa when seen in LNCaP cells which are an androgen dependent cell line. Moreover, NOVA1 showed the highest expression in androgen independent PC3 cell lines which is a good candidate from available cell line choices to carry out the *NOVA1* gene knockdown and carry out RNAseq to look at downstream effects of *NOVA1*.

Initially, I looked at the prostate cancer cells that had been treated with synthetic androgen and steroid deplete media respectively. The treatment was performed for different time intervals up to 24 hours where steroid deplete cells had higher *NOVA1* expression. This was as I had initially predicted. Same observation was made when LNCaP cells that had been treated with androgen and steroid deplete for 72 hours were harvested and RT-PCR was performed. This was consistent with the cell lines that had additionally been treated with Casodex or DMSO where, steroid deplete samples treated with Casodex had highest *NOVA1* expression. This was consistent with the hypothesis that androgens do have an effect on *NOVA1* gene expression. The 72 hour time frame used for cells treated in tissue culture was used because through previous experiments in our lab cells were seen to have optimum growth and results at 72 hour time frame and to have enough confluence for downstream processing.

Additionally, I looked into protein samples to confirm what cell lines showed highest NOVA1 protein expression. Cell lines compared were LNCaP, PC3, DU-145, CWR22RV1, and NE (Neuroendocrine) cells where, androgen independent cell lines showed higher NOVA1 expression compared to androgen dependent cell lines. Among these cell lines PC3 showed the highest NOVA1 gene expression. This was in line with the initial prediction.

I could not see a band for neuroendocrine cell lines whereas, loading control shows presence of protein in the lane. This could be due to neuroendocrine samples being not from the original metastatic source. These cell lines came from Canada and were made synthetically rather than originally extracted from patient. Therefore, reliability of this cell lines could not be well-established.

Based on the above observations, I set out to see how NOVA1 was regulated in prostate tumour were normal prostate tissue and found that *NOVA1* had lower expression in prostate adenocarcinoma. But I was not able to obtain data to look for *NOVA1* expression profile in aggressive patient tumour data due to lack of samples available. Moreover, survival curve was also consistent with the comparison between diseased and normal tissue where people with lower *NOVA1* expression had higher survival. However, the clinical data used above for box plot and survival curves was based on the samples from PRAD cohort which are the primary cancer and did not include castrate resistant and metastatic samples due to the lack of ample data sets for these types of prostate cancer in TCGA.

Additionally, I was able to identify a connection between *NOVA1* and AR binding sites using ReMap database. I was able to achieve all targets successfully for this chapter where I was able to successfully identify a link between androgen regulation and *NOVA1* gene. NOVA1 is indeed inversely regulated to androgens and is possibly controlled by androgen and androgen receptor interaction.

Chapter 4: *NOVA1* gene expression Knockdown: Refining Techniques and Unveiling Gene Expression Alterations

4.1 Introduction

In this chapter, I wanted to knockdown *NOVA1* gene expression in prostate cancer PC3 cells grown in tissue culture, using siRNA transfection. This knockdown would enable me to perform RNAseq analysis on the *NOVA1* gene knockdown RNA samples, and to understand normal functions of the *NOVA1* gene. Before detailed RNAseq analysis, I planned to western blot the protein sample prepared from the transfected PC3 knockdown cells to confirm the knockdown of NOVA1 protein shown in Figure 4.6. Moreover, I wanted to use the RNAseq data to look at the common events between *NOVA1* gene knockdown and *ESRP1* and *ESRP2* knockdown from Munkley et al. 2019, to identify the regulation differences and overlaps between the common events[162]. The schematic shown in Figure 4.1 summarizes the steps planned for this chapter. Because ESRP1 and ESRP2 genes have been associated with epithelial splicing patterns and as NOVA1 is regulated opposite to the ESRP genes and is thought to promote mesenchymal splicing patterns, I predicted that NOVA1 would have opposite effects on splicing control and possibly on the same events controlled by ESRP genes.

EMT (epithelial-mesenchymal transition) is crucial to cancer progression and metastasis. During EMT, cells switch from an epithelial to mesenchymal phenotype, gaining invasiveness and migratory (metastatic) cell behavior. EMT can result from AS controlled by splicing factors, i.e. expression of ESRP1 and ESRP2. Low expression levels of ESRPs are seen to initiate or play a role in EMT, where cell-cell adhesion is diminished, and cell motility is promoted, leading to metastasis. The study by Lu et al. 2015.[163] prepared in-vitro and in-vivo metastatic models to look at PC3 cells with different metastatic properties.[163] In this study Lu et al. 2015 showed that the down-regulation of ESRP1 and ESRP2 caused increased metastatic potential, and further observed a regulatory change in the *NOVA1* gene, which regulates mesenchymal AS patterns in the process of EMT. [163] The same Lu et al study looked at a metastatic cell line from the liver with lower *ESRP* expression, whereas the opposite was true for *NOVA1*, which had a higher expression level in the liver metastatic cell line. It was also noted that ESRP1 and

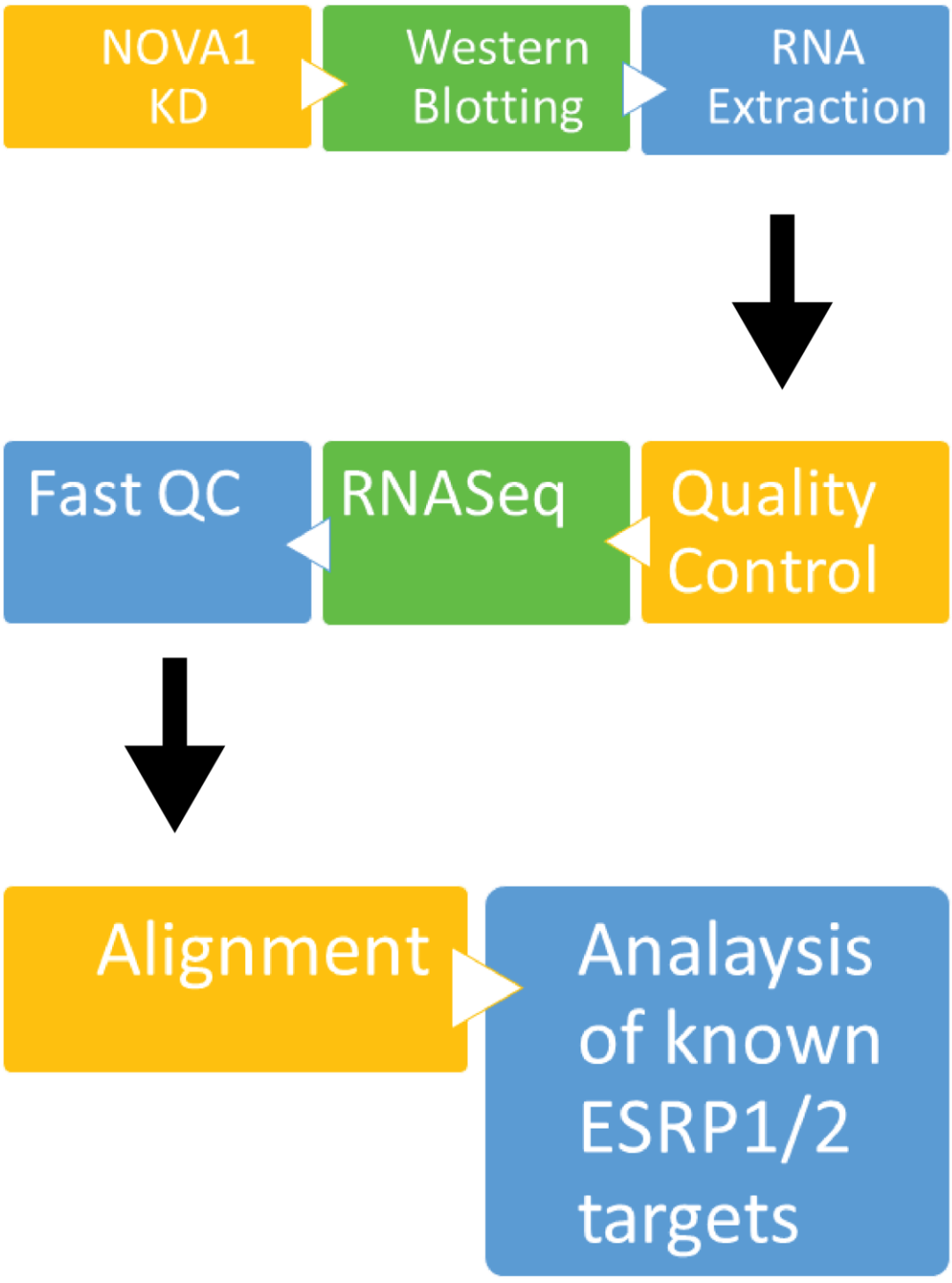


Figure 4.1: Flow chart shows a summary of processes undertaken for the analysis.

NOVA1 were among the three most differentiated splicing factors among the original PC3

cells and metastatic liver samples. The study also looked at splicing levels of GRHL1 (a transcription factor highly expressed in epithelial cells compared to metastatic cells), which were enhanced by ESRP1 and repressed by NOVA1. The study highlighted the opposing roles of ESRP1, ESRP2 and NOVA1 and the vital role of NOVA1 in several cancers. However, it could not provide a functional role and significance of NOVA1, suggesting further work to be done on NOVA1[163].

siRNA (small interfering ribonucleic acid) are complementary silencer nucleotide sequences that can bind to a target RNA and temporarily disrupt the gene by blocking translation and causing mRNA degradation [164]. Transfection is the process by which siRNA/foreign nucleic acid is introduced into the cells containing the target gene [165].

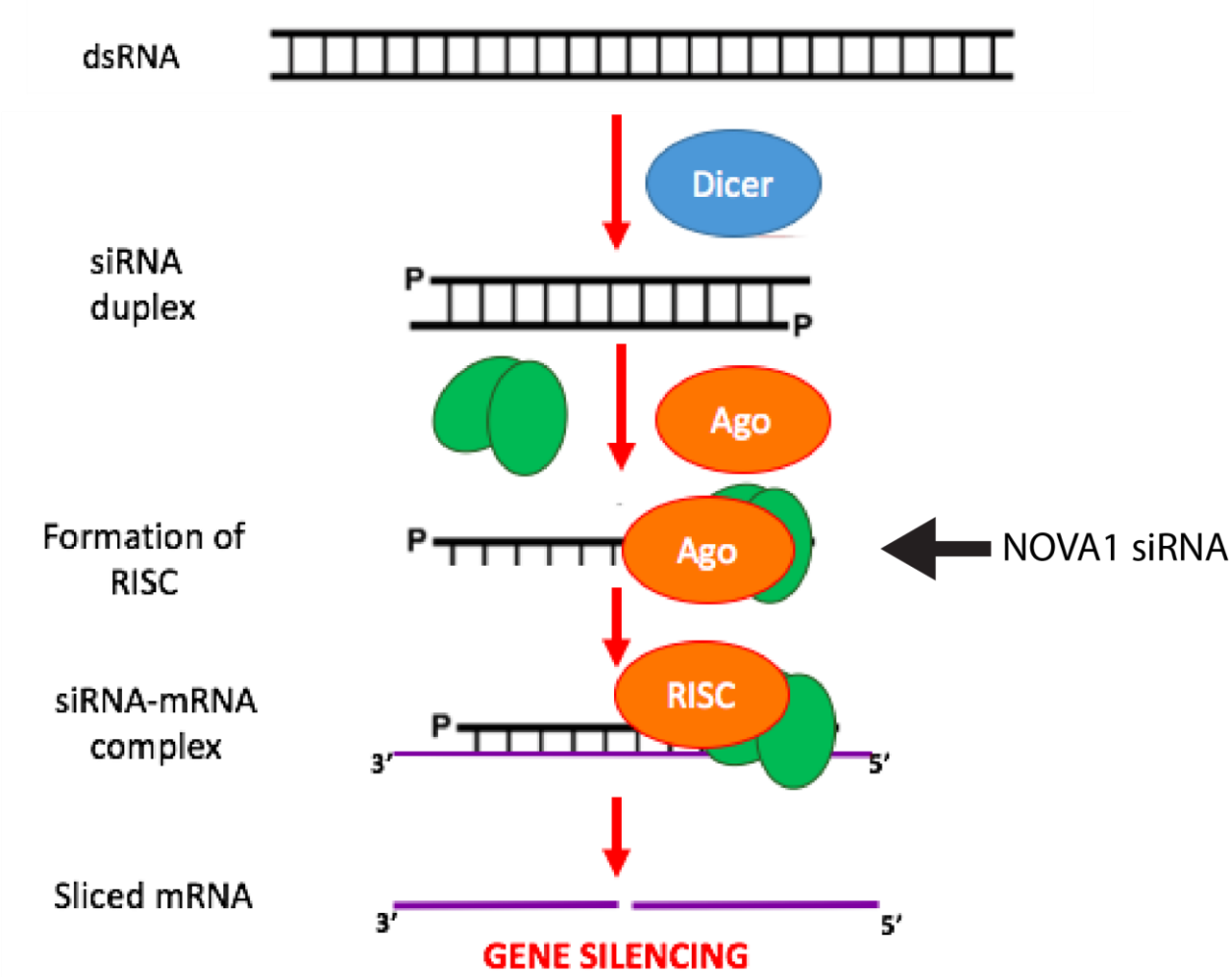


Figure 4.2: Process of gene silencing. *The arrow shows the point in process where commercial siRNA is introduced and every process before this point has been bypassed. The picture is adapted from Singh135, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons and has been modified for the thesis.*

I chose to knockdown NOVA1 gene expression using siRNA. Normally dsRNA is cleaved by DICER enzyme into siRNA. siRNA then recruits other proteins to form a RISC (RNA-induced silencing complex). Once the complex is formed, the RISC cuts down the double strand to a single strand, and an RNA strand complementary to the target mRNA is retained. Once the mRNA is transported from the nucleus to the cytoplasm to be translated, the siRNA RISC complex binds to the complementary mRNA strand, leading to mRNA strand cleavage and degradation. Thus, the process of translation from RNA to protein is blocked [166,167]. In this case I used a commercial siRNA therefore the DICER cleavage step is skipped, and the introduced siRNA directly becomes a part of the RISC which then binds to the complementary sequence and carries out the repression of the translation. Figure 4.2 shows the complete process of normal dsRNA and arrow in Figure 4.2 points to a step where the commercial siRNA was introduced bypassing any previous steps.

After siRNA depletion of specific mRNAs, it is possible to measure effects on specific downstream genes and proteins, or more globally, by monitoring gene expression pathways. After the breakthrough by Watson and Crick that elaborated the DNA structure in 1953 [168], the understanding of DNA sequences matured but progressed slowly until the advent of Sanger sequencing in the 1970s [169] that has since changed again with the introduction and evolution of NGS (Next Generation Sequencing) technology [170] which was a game changer as genome sequencing was much faster and more cost-effective, with a complete human genome sequence obtained within twenty-four hours. However, it became even more accessible to run RNA sequencing, developed fifteen years ago. Thirty samples of RNA-Seq could be run simultaneously, and results obtained in as little as five hours to yield whole transcriptome data [170,171].

Data generated by the high-throughput sequencing experiments is raw data that needs post-sequencing processing to extract meaningful information. Therefore, downstream data processing must initially go through quality control, alignment, and mapping tools. Typical quality control for RNA sequencing data is done using FASTQC [172]. Individual RNAseq reads also need to be aligned to the genome. The tool utilized for this project was HISAT2 to align sequence reads to a reference genome [173]. The quality control for the RNAseq data set and

the alignment process for this thesis was done in collaboration with Sara Luzzi, Newcastle University.

4.2 Aims of Chapter

Previously I had observed NOVA1 protein regulation in various cell lines and in response to androgens (Chapter 3). NOVA1 had the highest expression in PC3 cell lines, and it was repressed in presence of androgens. In this chapter, I aimed to achieve a successful knockdown of NOVA1 in PC3 cells and generate data for downstream analysis. I wanted to test the hypothesis that in prostate cancer cells NOVA1 would regulate splicing patterns in the opposite direction to ESRP1.

My aims for this chapter were:

1. **Achieve a successful knockdown of NOVA1 in PC3 cell lines.** I carried out the siRNA knockdown in the PC3 cell line as it has high endogenous expression of NOVA1, and so should enable a better visualization of targets that would be changing in response to NOVA1 gene.
2. **Produce a high-quality RNA of the knockdown and control samples.** Upon successful knockdown of NOVA1, RNA extraction was carried out to obtain high quality RNA.
3. **Obtain a high-quality RNA sequencing dataset for downstream analysis.** RNA was sent for RNA sequencing which would then be quality assessed to ensure it is worthy of downstream analysis. This would ensure accuracy and authenticity.
4. **Analyze known ESRP1/2 targets from the Munkley et al. 2019 study.**[48] I would utilize already available exon targets list from ESRP1/2 knockdown identified from this previous study and compare them directly to NOVA1 knockdown data using IGV to look for any similarities and differences across selected events.

4.3 Results

4.3.1 Optimizing NOVA1 knockdown.

The results in the previous chapter showed that PC3 cells express reasonably high levels of NOVA1 protein, making them a suitable cell line to monitor effects after NOVA1 depletion. A knockdown of the *NOVA1* gene was thus done to reduce the NOVA1 protein expression in PC3 prostate cancer cell lines. NOVA1 protein is 52KDa, so I monitored levels present within control cells and after siRNA depletion using western blots relative to actin as a loading control.

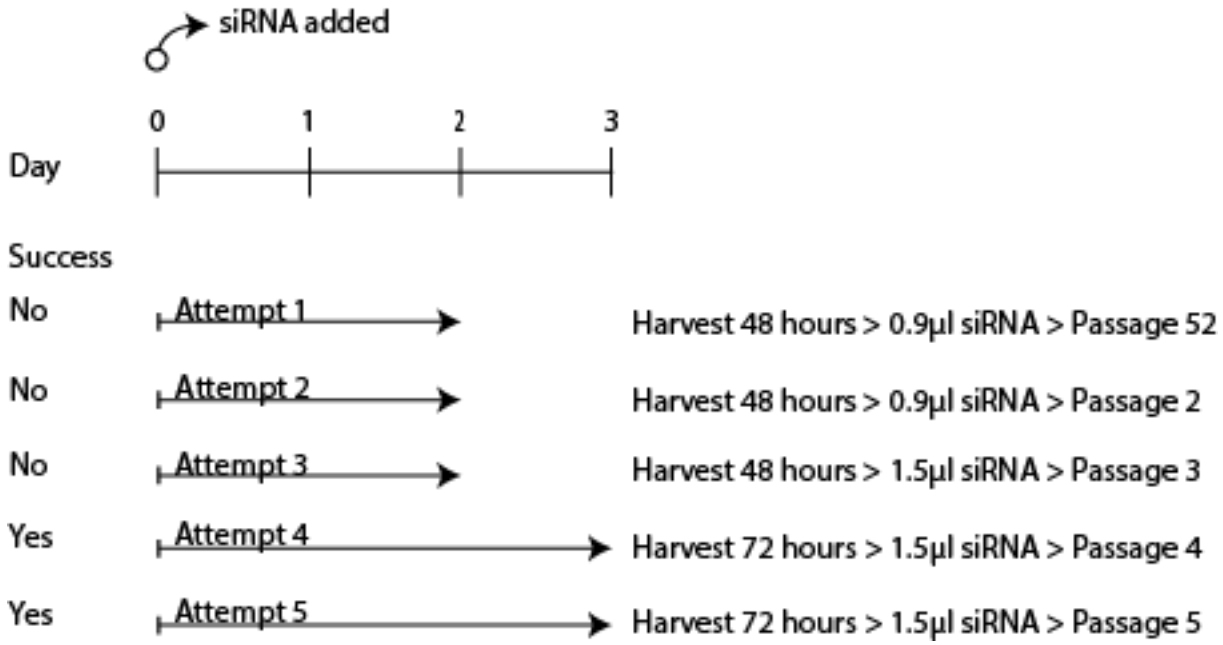


Figure 4.3: Schematic for NOVA1 siRNA knockdown. *Five attempts were made to knockdown NOVA1 in PC3 cells. Dot on the top of the figure shows the siRNA that was added on day zero in each case. On the left figure shows whether an attempt was a success or not. On the right, conditions for each attempt are mentioned along with the passage number of the PC3 cell lines.*

To investigate the role of NOVA1 in prostate cancer, a series of siRNA-mediated knockdown experiments were conducted in PC3 cells. Given that efficient and reproducible knockdown was essential for downstream analyses including RNA sequencing and protein validation, substantial effort was dedicated to optimizing the siRNA transfection protocol as shown in Figure 4.3.

To knockdown NOVA1 expression in PC3 cell lines, I first grew the PC3 cells in a tissue culture lab in RPMI media that contained five percent FBS (Fetal Bovine Serum) and one percent P/S (Penicillin/Streptomycin) antibiotics. Once the cells had optimum growth, they were split and added to a six-well plate. The cells were plated for twenty-four hours at 37°C in an incubator. Later, either a siRNA specific to NOVA1 or a control siRNA was transfected into the cells using RNA iMax lipofectamine, and cells were retained in an incubator at 37°C for seventy-two hours. Afterwards, the cells were harvested and split into two halves for each well. One-half of the cell pellet was used to extract RNA, and another half was used to make a protein sample to carry out western blot analysis to monitor whether the knockdown of NOVA1 protein had been successful.

In the first round of transfection, PC3 cells were plated and transfected using 0.9 µL of siRNA diluted in 10 µL of nuclease-free water, following a 24-hour incubation period post-seeding. Cells were incubated for 48 hours post-transfection before harvesting. Western blot analysis revealed no discernible change in NOVA1 protein band intensity between the knockdown and negative control samples (Figure 4.4A), indicating a failure in silencing NOVA1 expression.

To eliminate the possibility of using compromised cells, a second attempt was performed using a fresh batch of PC3 cells with a lower passage number, under identical transfection conditions. This also failed to produce a visible reduction in NOVA1 protein levels (Figure 4.4B), suggesting that the transfection conditions themselves, rather than cell quality, were insufficient.

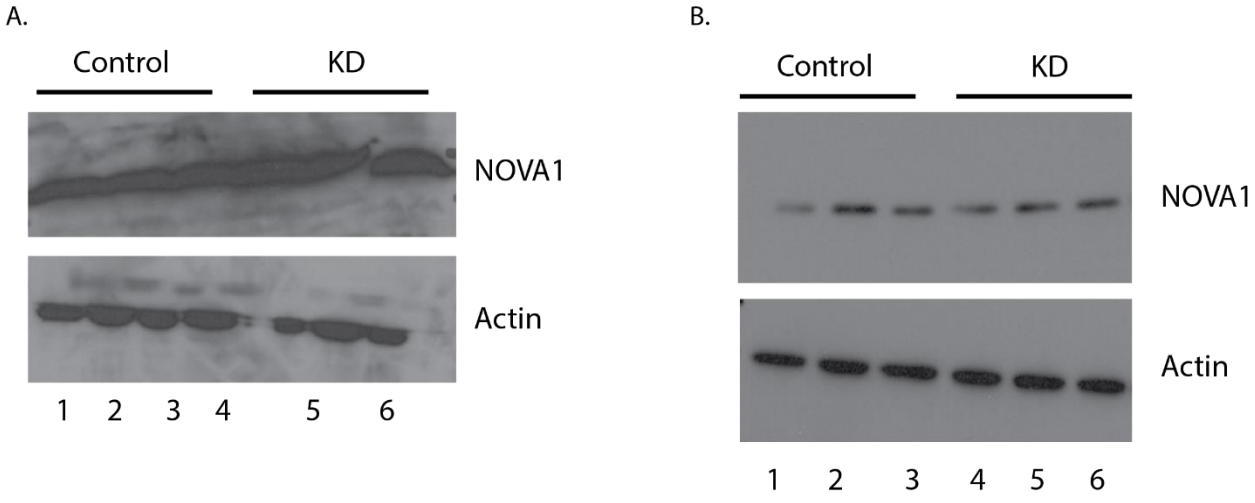


Figure 4.4: Western Blot analysis of NOVA1 protein. *Western blot **A** and **B** shows Lanes 1-3 with samples prepared from PC3 cells treated with control siRNA and showing NOVA1 protein expression. Lanes 4-6.*

In response, a third attempt increased the siRNA concentration to 1.5 µL per well, while maintaining the same volume of nuclease-free water and incubation time (48 hours). This modification resulted in a partial knockdown, evidenced by a modest reduction in NOVA1 band intensity compared to controls on the Western blot shown in Figure 4.5. Though not fully satisfactory, this result confirmed that increasing siRNA concentration had a measurable effect and informed the next stage of optimization.

A fourth transfection attempt built upon the previous findings by maintaining the 1.5 µL siRNA concentration but extending the incubation period to 72 hours, allowing more time for the siRNA to exert its effect. This yielded a substantial reduction in NOVA1 protein levels, with Western blot showing strong bands at ~52 kDa in control lanes and faint to undetectable bands in knockdown lanes.

However, although knockdown was clearly successful at the protein level, the RNA extracted using TRIzol from this attempt yielded low concentrations (<100 ng/µL) and poor purity, making it unsuitable for RNA-seq applications.

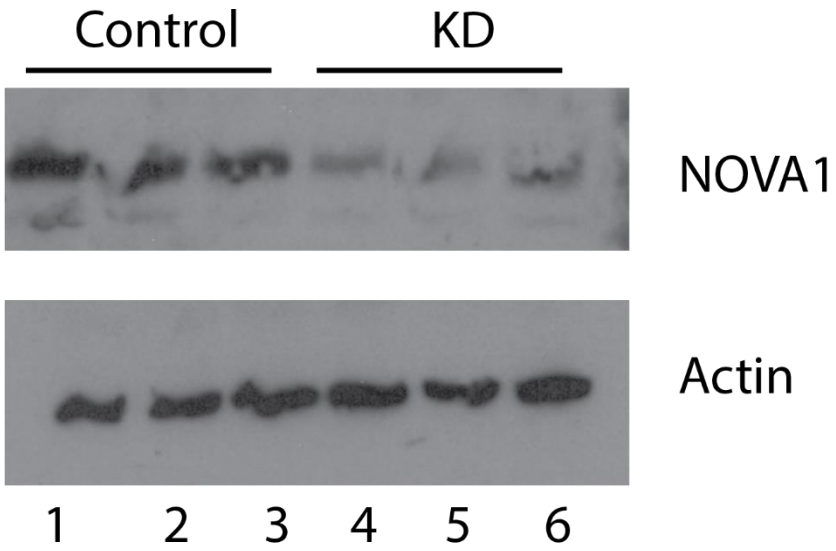


Figure 4.5: Western Blot analysis of NOVA1 protein. *Lanes 1-3 contain the NOVA1 control samples whereas, lanes 4-6 contain the knockdown NOVA1 samples. The KD vs Control samples are compared against actin loading control.*

Recognizing the need for high-quality RNA, a fifth and final attempt repeated the conditions of the fourth (1.5 µL siRNA, 72-hour incubation), this time prioritizing RNA yield and integrity. Protein lysates again confirmed effective NOVA1 knockdown via Western blot (Figure 4.6,

lanes 4–6 showed no visible NOVA1 bands). To further validate specificity, a mouse brain lysate—known to express high levels of NOVA1—was included as a positive control and showed a distinct NOVA1 band at the expected molecular weight. The RNA extracted was of sufficient quantity and quality (Nanodrop readings > 200 ng/µL, A260/280 ratio ~2.0), making it viable for downstream transcriptomic analysis..

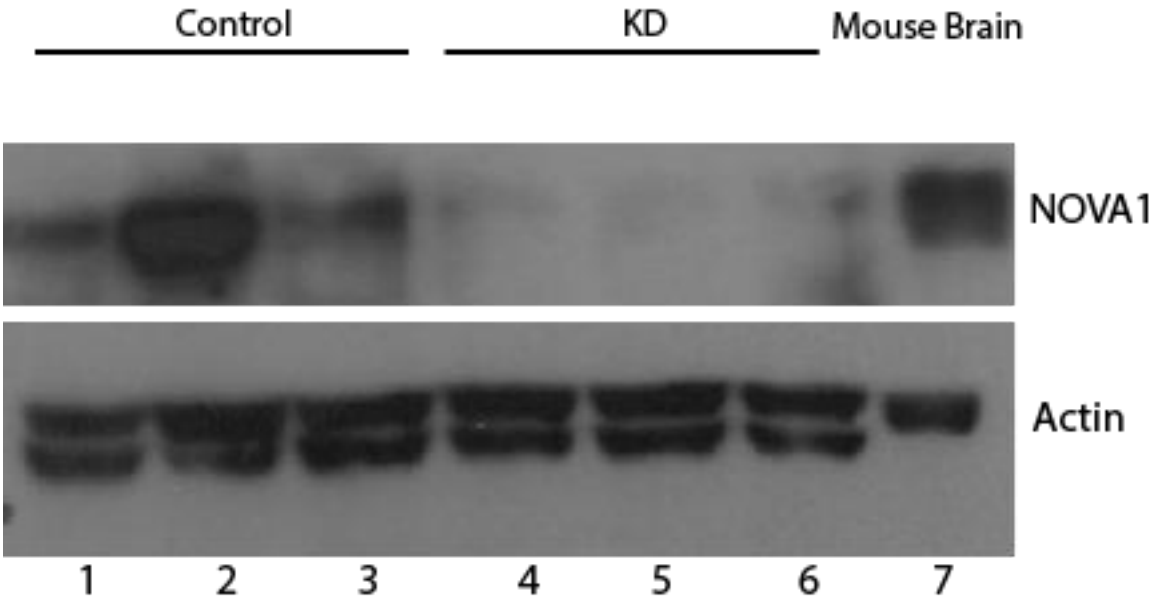


Figure 4.6: Western Blot Analysis of NOVA1 Gene Expression. *Western blot shows Lanes 1-3 with samples prepared from PC3 cells treated with control siRNA and showing NOVA1 protein expression. Lanes 4-6 show analysis of protein samples that were prepared from PC3 cells that were treated with siRNA specific to NOVA1 and indicating that NOVA1 protein has been knocked down. Lane 7 was an additional mouse brain control sample that showed NOVA1 expression.*

4.3.2 RNA extraction from PC3 cells for global RNAseq analysis

The above section showed that I had efficiently knocked down NOVA1 protein expression in triplicate samples of PC3 cells. I next planned to use these samples for parallel RNA purification and downstream analysis using RNAseq. Before high-throughput sequencing, I had to purify high-quality RNA suitable for global analysis. This was achieved using the QIAGEN RNA extraction kit as per the standard protocol provided with the kit (see Chapter 2, methods).

Table 4.1: RNA concentration obtained using Nanopore.

Sample	*Concentration (ng/ul)*
Knockdown 1	*236.8*
Knockdown 2	*288.4*
Knockdown 3	*219.0*
Control 1	*173.9*
Control 2	*156.86*
Control 3	*142.3*

Before RNA sequencing, assessing the concentration of purified RNA is a crucial step. To achieve this, I conducted an initial quality check by analyzing the RNA concentration using a nanodrop Table 4.1. The RNA concentration exceeded the recommended threshold of 2 ng, prompting us to proceed with further quality assessment. Subsequently, the RNA samples underwent further evaluation of their concentrations using a bioanalyzer. The results from the bioanalyzer, including the RNA Integrity Number (RINe), which is recommended at more than 7 to be considered of good quality and the concentration in ng/ml, are presented in Table 4.2. The RINe number of my samples was close to 10, which is considered a good quality RNA for sequencing. The samples met the quality standards and were advanced to RNA sequencing through high-throughput sequencing.

Table 4.2 Bioanalyzer results for the extracted RNA.

Sample	*GCF Number*	*RINe*	*Tng/ul*
1	*Knockdown 1*	*9.9*	*180*
2	*Knockdown 2*	*9.8*	*336*
3	*Knockdown 3*	*9.9*	*351*
4	*Control 1*	*9.9*	*161*
5	*Control 2*	*10*	*106*
6	*Control 3*	*9.9*	*193*

4.3.3 RNAseq Quality Control.

After conducting quality control on triplicate RNA samples for NOVA1 knockdown and control, the Genomic facility at Newcastle University performed the RNA sequencing. The resulting data underwent sequence quality control using FASTQC and alignment using HISAT2. These crucial steps ensure the data's usability and minimize contamination that could compromise its significance. Dr. Sara Luzzi in our lab performed these steps of quality control, alignment, and bioinformatic processes using my RNAseq data.

The FastQC analysis revealed a notably high number of unique sequences reads (Figure 4.7A), indicating minimal overlap or repetitive reads and reinforcing the overall data quality. The predominance of unique reads within the NOVA1 knockdown samples suggests a diminished presence of PCR bias or contamination. The lower incidence of duplicate reads in these samples further supports the absence of substantial artefacts or biases that typically arise during PCR amplification or sequencing. Further, both per sequence quality (Figure 4.7B) and mean quality (Figure 4.7C) line graphs displayed a green range, signifying good sequence quality. Additionally, the per base N content (Figure 4.7D) was observed to be very low, indicating superior base call quality. Higher N content typically suggests substitution due to difficulty calling the conventional base for the specified genome location. However, the low N content here attests to the high quality of base calls. The GC content (Figure 4.7E) exhibited a normal uniform distribution curve without deviations or irregularities. In contrast, the sequence length distribution (Figure 4.7F) remained consistent across all samples, with no outliers detected.

Additionally, Principal component analysis (PCA), shown in Figure 4.8 below, was performed on each knockdown and control sample. PCA showed that in each prostate cancer cell sample, either siRNA knockdown or control had three replicates. Control replicates (Red) in Figure 4.8 seen on the left-hand side of the figure and NOVA1 knockdown samples (Blue) shown on the right-side fall into a similar side of the variance that suggests that in all the cases shown in Figure 4.8 shared similar properties. This separation of samples confirms that these samples are appropriate for further analysis.

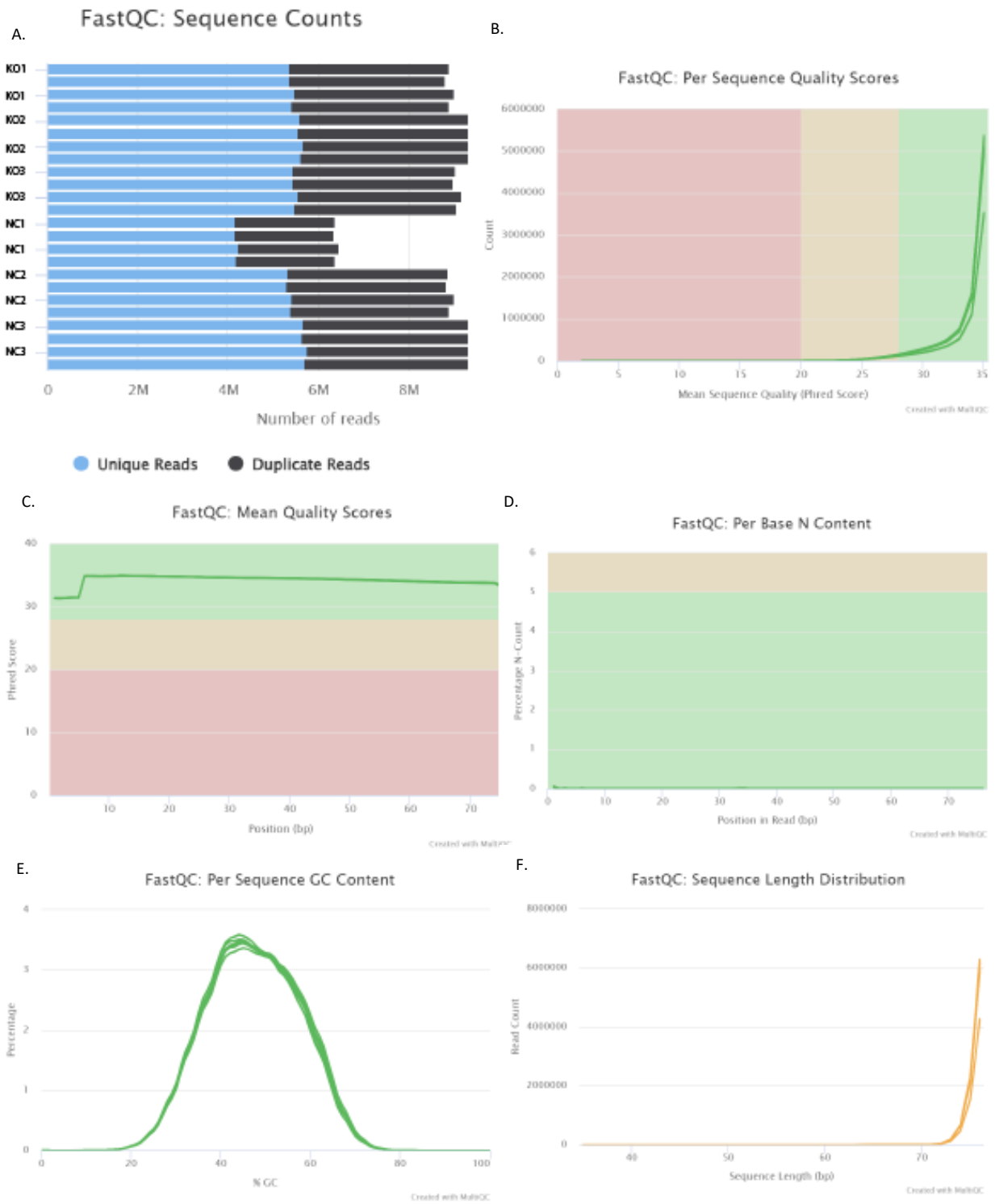


Figure 4.7: Quality control of RNAseq data. *A. Bar graph shows unique reads obtained (Blue) and duplicate reads (grey). The low number of duplicate reads indicate a high quality of sequenced data. B. The line graph shows number of reads associated with mean quality score for each read. Three shaded areas can be seen representing poor (red), average (yellow) and good (green) quality of data. C. The line graph again is used to see the mean quality score of the sequence reads and green indicates a good quality data. D. Graph shows N calls for the sequence that has been aligned. If the N calls were high that would decrease the sequence reliability. E. Graph shows GC content with normal distribution curve and no outliers observed which is consistent for a good quality data. F. Graph shows sequence length distribution for obtained reads. The distribution through-out the reads is uniform.*

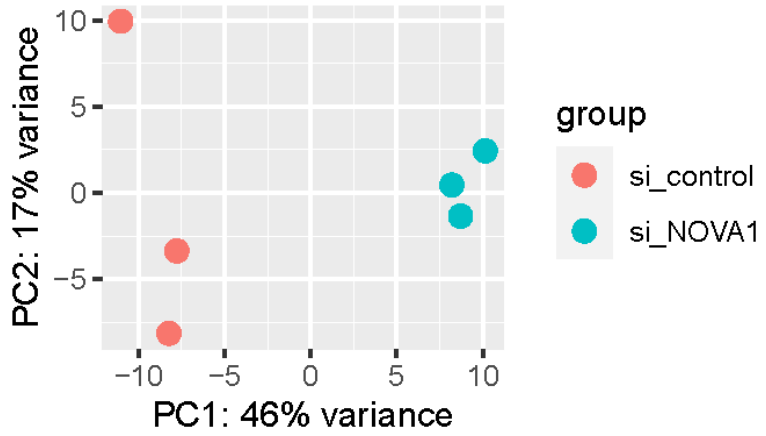


Figure 4.8*: Principal component analysis of RNAseq samples.* *Control (Red) cells and knockdown (Blue) cells performed on NOVA1 and negative control knockdown samples on PC3 cell lines. PCA was performed by Dr. Sarah Luzzi (Elliot Lab).*

4.3.4 Does NOVA1 have an opposite pattern of splicing control compared to ESRP1/2?

The initial hypothesis was that the regulatory patterns of common events between NOVA1 knockdown and ESRP 1/2 knockdown would be inversely correlated. Thus, I expected them to be controlling the same target exons, albeit in an inverse manner. An example would be an exon activated by ESRP1/2 seen from ESRP1/2 knockdown in LNCaP cells, which would be repressed by NOVA1 knockdown seen in PC3 cells. Therefore, to test this hypothesis, I set out to compare and analyze the known ESRP1/2 regulated exons shown in Table 2 below observed in the context of ESRP1 and ESRP2 knockdown, as documented in Munkley et al. 2019 to my NOVA1 knockdown RNAseq data using IGV.[48] An exon with higher expression in the control sample of NOVA1 compared to the knockdown sample is activated by NOVA1. Whereas, if an exon has higher expression in the NOVA1 knockdown sample and has lower expression in the control sample, it is repressed by NOVA1.

Exons that behave the same on ESRP1/ESRP2 depletion and NOVA1 depletion.

Some exons behaved similarly for *ESRP1/2* and *NOVA1* genes. Notably, *DOCK7* was repressed in the ESRP1/2 knockdown, mirroring its repression in NOVA1 knockdown. This similar pattern was also observed in other downstream genes, including *CEACAM1*, *FNIP1*, *KIF13A*, *MLPH*,

OSBPL3, and *ZNF207*. Whereas other genes exhibited activation for both *ESRP1/2* and *NOVA1* genes. These gene exons include *APLP2*, *MINK1*, *NUMB*, *RALGPS2*, and *TUFT1*.

Exons that have an opposite pattern of behavior after NOVA1 depletion versus ESRP1/2 depletion

It was also observed that some gene exons behaved inversely when compared to *ESRP1/2* gene-regulated exons to *NOVA1* gene-regulated exons. These gene exons include ACSF2, *ARFGAP2*, *ARHGEF11*, *ARHGEF12*, *CTNND1*, *DNM1L*, *EIF4A2*, *ENAH*, *FLNB*, *FN1*, *GRHL1*, *INF2*, *ITGA6*, *MAP3K7*, *MBNL1*, *MYO1B*, *RAC1*, *RPS24*, *SLC37A2*, *SLK*, *TCIRG1*, *TRIP10*, *UAP1*, *VPS39*, *WNK1*. Out of these genes, for instance, *ESRP1/2* activated *FLNB* and *FN1* but were repressed by *NOVA1*. Similarly, *MYO1B* and *RPS24* were repressed by ESRP1/2 and were activated by *NOVA1*.

Exons that have an independent pattern of behavior after NOVA1 versus ESRP1/2 depletion

Surprisingly, certain gene exons showed regulation changes in ESRP1/2 knockdown but remained unaffected in NOVA1 knockdown, including *ADAM15*, *ADAM7*, *MYH10*, *MAGI1*, *EPN3*, *EXOC7*, and *MYO6*. Consequently, despite being inversely regulated to ESRP1/2 and androgen controlled, NOVA1 impacts the same exons inversely. It can also regulate them in the same direction as ESRP1/2, potentially contributing to a synergistic effect in cancer progression. Additionally, as NOVA1 does not uniformly influence events comparable to ESRP1/2 knockdown, it may affect other events unique to the NOVA1 gene, possibly linked to similar biological and molecular pathways as ESRP1/2 or distinct pathways specific to NOVA1.

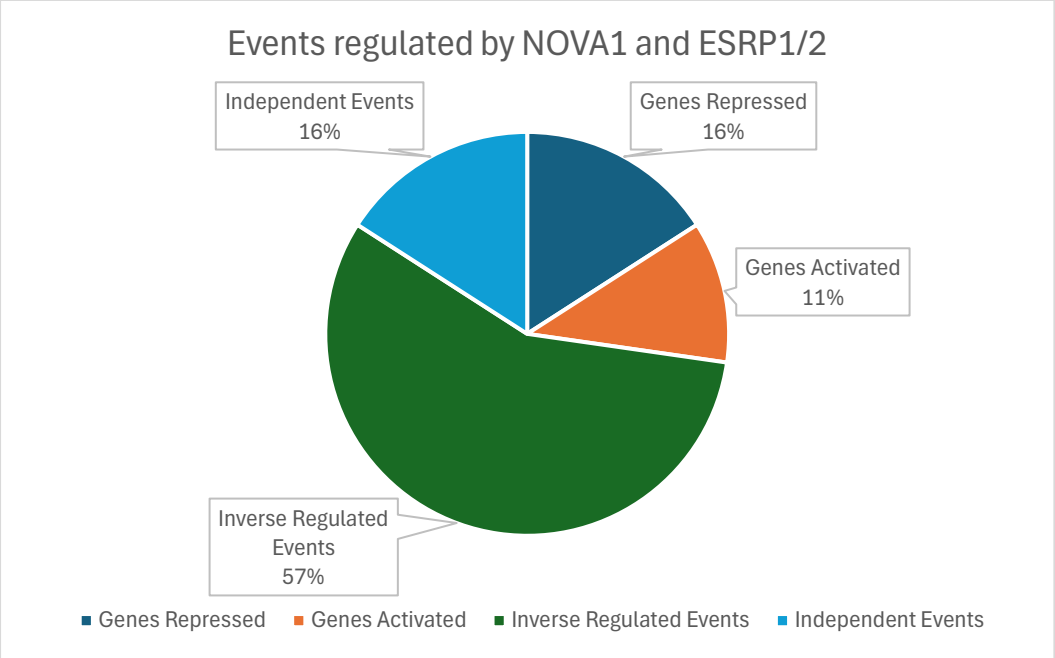


Figure 4.9: Pie chart for overlapping events between NOVA1 and ESRP1/2.

Comparison of NOVA targets and ESRP1/2 targets

Above, I compared how ESRP1/ESRP2 splicing targets were affected by NOVA1 depletion. I saw that 57% of events analyzed were inversely regulated as was expected based on initial prediction, as seen in figure 4.9. Surprisingly, it was not the case with all events, as seen in the pie chart above that 27% of the events observed were regulated similarly by both the genes, where 16% of the exon events were seen to be repressed by ESRP1/2 as well as NOVA1 and 11% of exons were activated by both genes. However, it was also seen that 16% of events were independent events that were controlled by ESRP1/2 but were unchanged in NOVA1 gene knockdown. The events ESRP1/2 targets that were investigated in NOVA1 knockdown dataset are listed in Table 4.3.

This gives an insight into the correlation between two genes that are otherwise seen to oppose each other and yet regulated by androgens. Consequently, despite being inversely regulated to ESRP1/2 and androgen controlled, NOVA1 can impact the same exons inversely. It can also regulate them in the same direction as ESRP1/2, potentially contributing to a synergistic effect in cancer progression. Additionally, as NOVA1 does not uniformly influence

events comparable to ESRP1/2 knockdown, it may affect other events unique to the *NOVA1* gene, possibly linked to similar biological and molecular pathways as ESRP1/2 or distinct pathways specific to NOVA1.

Table 4.3: List of known ESRP1/2 targets and their regulation in NOVA1 knockdown.

Genes	*Exon No.*	*Regulation in ESRP1 & 2 KD*	*Regulation in NOVA1 KD*	*Coordinates HG38*
ACSF2	Exon 3/4	Activated	Repressed	chr17:50461242-50461370
ADAM15	Exon 19/20	Repressed	--	chr1:155061904-155062117
ADAM7	Exon 20/21	Activated	--	chr8:24506091-24506156
APLP2	Exon 7	Activated	Activated	chr11:130123612-130123779
ARFGAP2	Exon 6 and 13	Activated	Repressed	chr11:47172710-47172751
ARHGEF11	Exon 38/39	Activated	Repressed	chr1:156939149-156939285
ARHGEF12	Exon 4	Activated	Repressed	chr11:120409394-120409450
CEACAM1	Exon 2	Repressed	Repressed	chr19:42511576-42511628
CTNND1	2/3	Repressed	Activate	chr11:57789037-57789155
DNM1L	Exon 13/15	Activated	Repressed	chr12:32738264-32738296
DOCK7	Exon 21/23	Repressed	Repressed	chr1:62544947-62545039
EIF4A2	Exon 10/11	Activated	Repressed	chr3:186788310-186788420

ENAH	Exon 11A	Activated	Repressed	chr1:225504991-225505053
EPN3	Exon 4	Activated	-	chr17:50538884-50538964
EXOC7	Exon 6	Repressed	-	chr17:76091143-76091235
FLNB	Exon 30	Activated	Repressed	chr3:58141858-58141929
FN1	Exon 25	Activated	Repressed	chr2:215392931-215393203
FNIP1/RAPGEF6	Exon 7	Repressed	Repressed	chr5:131710578-131710661
GRHL1	Exon 5	Activated	Repressed	chr2:9962455-9962531
INF2	Exon 22	Activated	Repressed	chr14:104715284-104715340
ITGA6	Exon 25	Activated	Repressed	chr2:172501772-172501901
KIF13A	Exon 25	Repressed	Repressed	chr6:17789872-17789910
MAGI1	Exon 7	Repressed	-	chr3:65448022-65448057
MAP3K7	Exon 12	Activated	Repressed	chr6:90544552-90544632
MBNL1	Exon 8	Activated	Repressed	chr3:152455542-152455577
MINK1	Exon 18	Activated	Activated	chr17:4892402-4892512
MLPH	Exon 9	Repressed	Repressed	chr2:237534564-237534647

MYO6	Exon 32	Activated	-	chr6:75911672-75911698
MYH10	Exon 6	Repressed	-	chr17:8576643-8576672
MYO1B	Exon 23	Repressed	Activated	chr2:191400749-191400835
NUMB	Exon 6	Activated	Activated	chr14:73316390-73316422
OSBPL3	Exon 8	Repressed	Repressed	chr7:24863200-24863292
RAC1	Exon 3B	Activated	Repressed	chr7:6398662-6398718
RALGPS2	Exon 15	Activated	Activated	chr1:178892230-178892307
RPS24	Exon 5	Repressed	Activated	chr10:78040204-78040225
SLC37A2	Exon 18	Activated	Repressed	chr11:125086204-125086260
SLK	Exon 13	Activated	Repressed	chr10:104010816-104010908
TCIRG1	Exon 19	Activated	Repressed	chr11:68050487-68050664
TRIP10	Exon 10	Activated	Repressed	chr19:6746029-6746196
TUFT1	Exon 2	Activated	Activated	chr1:151562091-151562165
UAP1	Exon 9	Activated	Repressed	chr1:162592732-162592782
VPS39	Exon 3	Activated	Repressed	chr15:42192066-42192098

WNK1	Exon 11	Repressed	Activated	chr12:879573-880031
ZNF207	Exon 8	Repressed	Repressed	chr17:32366665-32366757

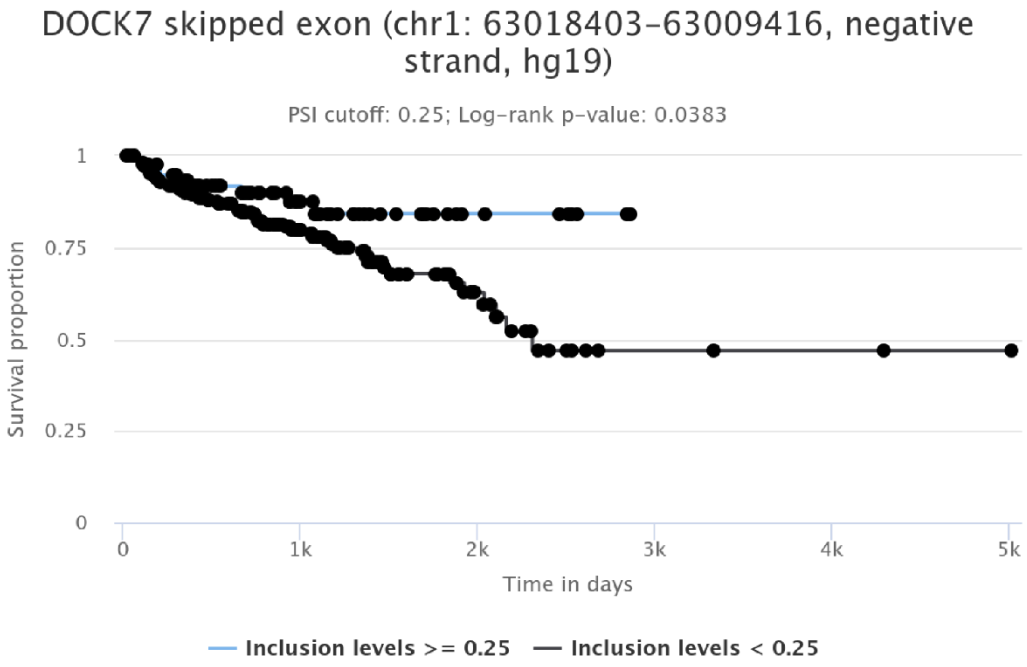
4.4 Discussion

In this chapter, NOVA1 transfection protocol was optimized to achieve effective knockdown of NOVA1 in PC3 cell lines that were found to have the highest NOVA1 expression. Optimization of the protocol involved refining the concentrations in the standard protocol. After successful knockdown of NOVA1, RNA was extracted using Qiagen RNA extraction kit as mentioned in chapter 2 of this thesis. In order to ensure the extracted RNA was of the required quality for RNA sequencing, first quality check was done using nanodrop machine and then using bioanalyzer. These steps confirmed the RNA was of the quality standards required for RNA sequencing. Subsequently, RNA sequencing was carried out and the quality of the sequenced data was assessed.

Then, bioinformatic analysis of this NOVA1 knockdown dataset was carried out by Dr. Sara Luzzi of NUBI. I then analyzed the data to investigate whether ESRP1 and ESRP2 splicing targets were influenced by depletion of *NOVA1* gene expression. The ESRP1 and ESRP2 splicing targets utilized in this study were taken from 2019 Munkley et al. study.[162] The analysis revealed that 57% of the ESRP1/2 splicing events were inversely regulated by NOVA1 knockdown as shown in Figure 4.9 and Table 4.3. Additionally, 11% of the splicing events were activated by both the gene datasets whereas 16% of the events were repressed by both. Lastly, about 16% of the events were specific to ESRP1 and ESRP2 and remained unaffected by NOVA1.

A noteworthy finding from this analysis was the identification of *DOCK7* and *FN1* as splicing targets of *NOVA1*. DOCK7 was repressed by both the NOVA1 and ESRP1/2. However, FN1 was activated by ESRP1 and ESRP2 while NOVA1 repressed FN1. When the splicing events for these splicing events were investigated further, it was found that low inclusion levels of *DOCK7* exon 23 led to decreased time to biochemical recurrence as shown in Figure 4.10A.

A.



B.

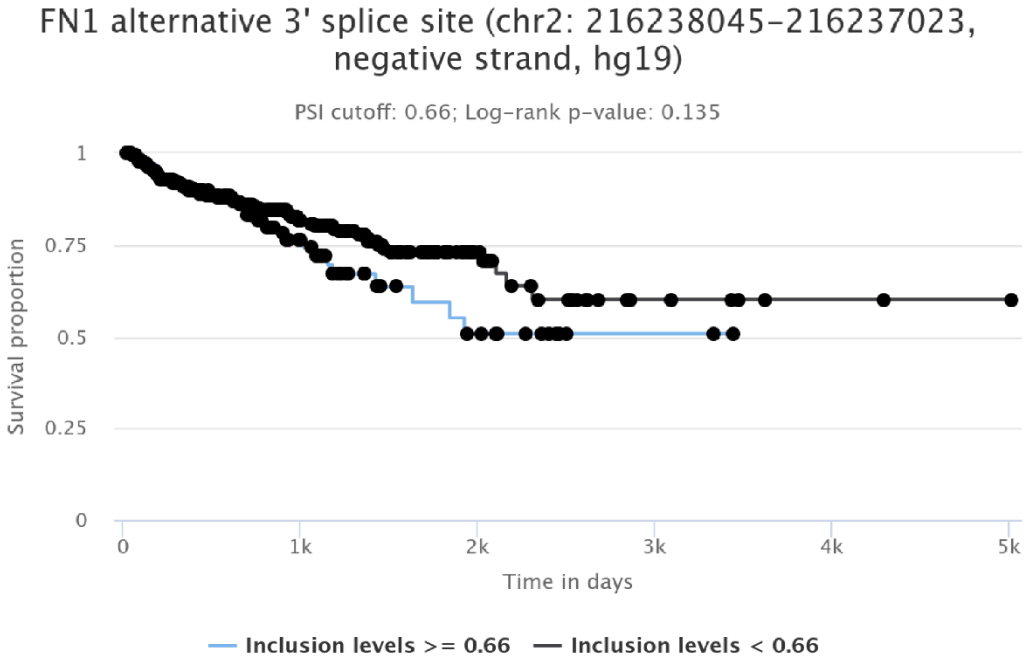


Figure 4.10: Survival graphs showing time to first biochemical recurrence. A. *DOCK7 blackline shows the event controlled by ESRP1/2 and NOVA1 and has decreased time to biochemical recurrence.* ***B.*** *Shows FN1 splicing target where blackline is ESRP1/2 regulated event and blue line is NOVA1 regulated event and blue line shows decreased time to biochemical recurrence of prostate cancer.*

Conversely, higher inclusion levels of FN1 exon 25 were found to be correlated with decreased time to biochemical recurrence of prostate cancer Figure 4.10B. This suggested that androgens would be regulating NOVA1 to synergize the effect of prostate cancer aggressiveness.

These results put together suggest that NOVA1 functions as an additional splicing factor regulated by androgens influencing progression of prostate cancer as an additional pathway rather than just a supporting event for *ESRP1* and *ESRP2* genes. This explains the complexity of prostate cancer and importance of understanding the functions of NOVA1 as it may prove to be a diagnostic marker or a therapeutic target for prostate cancer treatment.

Chapter 5: Investigating NOVA1's Global Impact on Gene Expression in Prostate Cancer

5.1 Introduction.

In previous chapters, I have established that NOVA1 expression changes in the presence and absence of androgens, with NOVA1 having a higher expression in the absence of androgens. I have also shown that NOVA1 has higher expression in the androgen negative PC3 prostate cancer cell line compared to other available cell lines. I optimized *NOVA1* gene knockdown in PC3 cell lines, and on confirmation of a good quality RNA extraction, it was sent for RNA sequencing. Initial analysis of this RNAseq data gave me an insight into how NOVA1 influences already known splicing targets regulated by ESRP protein. In this chapter, I describe how I used this RNAseq data to investigate more broadly how NOVA1 controls global gene expression and alternative splicing in prostate cancer.

Alternative gene expression and its global analysis from RNAseq data

Alternative splicing is a fascinating mechanism where a limited number of genes, approximately 20,000 in a human, give rise to more than 100,000 proteins.[174–180] Initially, the phenomenon of splicing was identified in adenovirus transcripts in 1977.[181–183] Alternative splicing can lead to the generation of different mRNA isoforms from a gene. Alternative splicing can happen in coding and non-coding regions of a gene.[183–186] Thus, the process of alternative splicing is a phenomenon increasing the complexity of the transcriptome. Alternative splicing is tightly regulated by splicing factors like NOVA1, which can either promote or inhibit the inclusion of specific exons during mRNA processing. During the process of exon-to-exon splicing, exons can be joined in an alternative sequence or in a way that some exons are skipped while others are retained, as shown in Figure 1.8 for *ESRP1* and *ESRP2* genes. Alternative splicing produces diverse mRNAs with different structures and functions, sometimes leading to the production of diverse proteins. There are seven basic types of known alternative splicing types, called skipped exons (SE), alternative 5’ splice site (A5), alternate 3’ splice site (A3), mutually exclusive exon (MX), intron retention (RI), alternative first exon (AF), and alternative last exon (AL).[187]

NOVA1 appears to exert downstream control on genes through alternative splicing. There are multiple tools that can be used for global analysis of differential splicing from RNAseq data, such as MAJIQ, rMATS, DEXSeq, and SUPPA2.[188–191] To investigate how NOVA1 knockdown affects alternative splicing, we used SUPPA2, a sophisticated bioinformatics tool designed for the analysis of splicing events from RNAseq data.[191] Initially, RNAseq data is handled through Salmon or Kallisto to estimate the transcript abundance by alignment of RNAseq data to a reference transcriptome, and then these transcript abundance estimates are used by SUPPA2 to calculate Percentage Splicing Inclusion (PSI) values between different sample conditions (in this case RNAseq data NOVA1 control and NOVA1 knockdown PC3 cells). SUPPA2 offers several strengths over other splicing analysis tools, such as MAJIQ, including its ability to handle large datasets efficiently and at faster speeds. SUPPA2 has a high detection rate even with shorter reads at low depth.[191] Along with the high speed performance of SUPPA2, the reported accuracy of PSI values for identified events is similar to or even surpassing that of rMATS and MAJIQ, in analyses presented by Trincado et al. where SUPPA2 was able to identify higher number of novel events not previously detected.[191] Additionally, SUPPA2 has a lower false positive rate in comparison to the other three tools.[191] Therefore, SUPPA2 was used in the analyses described in this Chapter to identify differential splicing events controlled by NOVA1.

I also monitored patterns of gene expression in response to NOVA1 depletion. Differential gene expression is a process in which different cells express different sets of genes that have a specialized function. Differential gene expression analysis is important in understanding how knockdown of gene function affects cellular patterns of transcription. Through differential expression analysis, NOVA1 knockdown samples were compared against control samples to identify the genes that were significantly upregulated or downregulated. To carry out differential expression analysis, DESeq2 was used. DESeq2 is a widely used and powerful statistical package designed to identify gene regulation from RNAseq data sets. DESeq2 has many advantages over other differential expression analysis tools, including handling varying library sizes and determining the fold changes. The downregulated and upregulated genes identified by differential expression analysis can enable me to understand how NOVA1 might be affecting prostate cancer.

Gene Ontology (GO) analysis [192,193] is a resource that can be used to analyse and interpret biological datasets and describe the function of the respective gene. GO functions are characterized into three domains, i.e. biological process, cellular component, and molecular function. Performing gene ontology on my dataset was a useful tool to understand key processes and pathways affected by the knockdown of NOVA1 in prostate cancer.[193]

Dr. Sara Luzzi, in our lab, analysed differential expression in my RNAseq sample data using SUPPA2 and DESeq2, and I analysed the resulting data. Global analyses can help to identify the genes or variants that have a potential importance in disease progression. Psichomics is an R programming based graphical interface tool that was useful for this analysis. Psichomics library can be downloaded from Bioconductor and loaded on to R to initiate a graphical user interface where TCGA data can be loaded and used to perform analysis where genes and splice variants can be correlated to the disease survival.

5.2 Aims

The aims of this chapter were to use the existing RNAseq data described in the last chapter to do the following:

1. Perform differential expression analysis using DESeq2 to find out all the genes upregulated or downregulated in the dataset.
2. Perform Gene Ontology (GO) and Gene Set Enrichment Analysis (GSEA) on the genes down and upregulate to identify the pathways that are regulated by NOVA1 expression changes.
3. Perform functional assay on cell to observe effect of gene regulation.
4. Perform differential splicing analysis using SUPPA2 to identify any splicing events.
5. Visually confirm splicing events found using IGV (Integrative Genome Viewer).
6. Compare any novel targets of NOVA1 to ESRP1 & ESRP2 to find out if they are affected by ESRP genes.
7. Use Psichomics to find a correlation between NOVA1 targets and their disease correlation in prostate adenocarcinoma samples available on TCGA.

5.3 Results

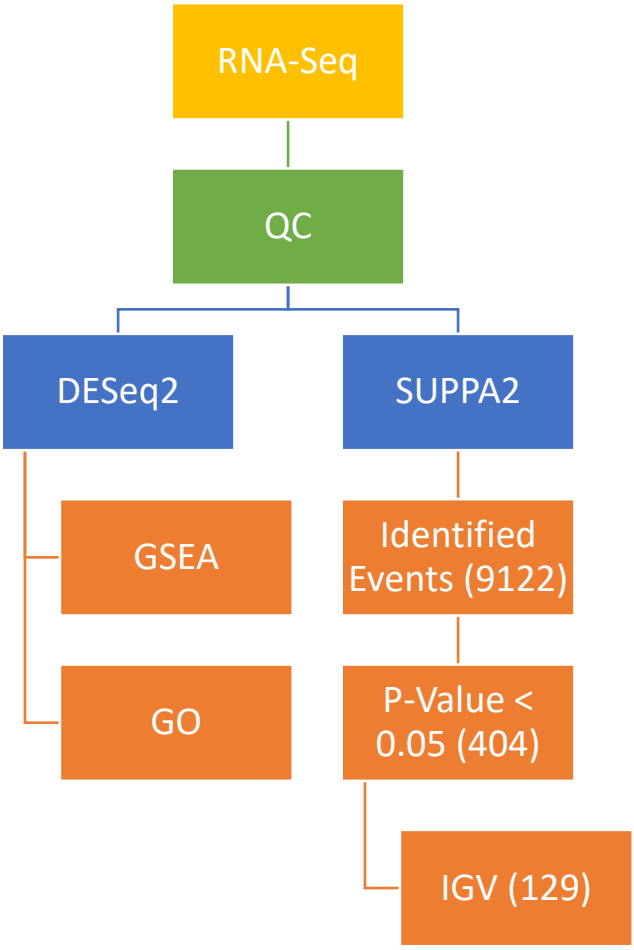


Figure 5.1: RNA-Seq flow diagram. *The flow diagram shows the steps for bioinformatics analysis where RNA-Seq was performed initially following that a quality control was carried out. The data was then used to perform two different analyses. Differential expression analysis using DESeq2 tool and differential splicing analysis using SUPPA2 tool.*

5.3.1 Global analysis of gene expression patterns after NOVA1 depletion using RNA sequencing.

I sought to globally analyse the RNA sequencing data from Chapter 4 to identify patterns of differential gene expression and differential gene splicing after NOVA1 knockdown in the PC3 (Androgen independent) prostate cancer cell line. Differential gene expression analysis was done bioinformatically by Dr. Sara Luzzi in the Elliott lab, using my RNAseq data. The overall pathway of this bioinformatics pipeline is shown in Figure 5.1. The DESeq2 tool was used to identify the genes that were upregulated and downregulated with p-value less than 0.05 as shown in flow diagram in Figure 5.1. The parameters used to filter differentially expressed

genes in DESeq2 were log2 fold change of greater than 1 and less than -1 with a p-value of less than 0.05. Following the differential expression analysis using DESeq2 tool, gene ontology was performed on the genes to identify the pathways enriched by those genes. To identify the differentially spliced genes in response to NOVA1 knockdown, the SUPPA2 tool was utilised. The SUPPA2 identified events were then visualized using IGV (Integrated Genome Viewer), the psichomics programme, and in some cases experimentally validated.

5.3.2 Patterns of differential gene expression after NOVA1 depletion.

Differential expression analysis using DESeq2 identified that the total number of differentially expressed genes after NOVA1 depletion was approximately 16000. When this list was filtered according to adjusted p-value of < 0.05, 4188 genes were found to be significantly differentially expressed. Out of these 4188 genes, 2135 genes were upregulated, and 2053 genes were downregulated. This pattern of overall gene expression change is shown in the

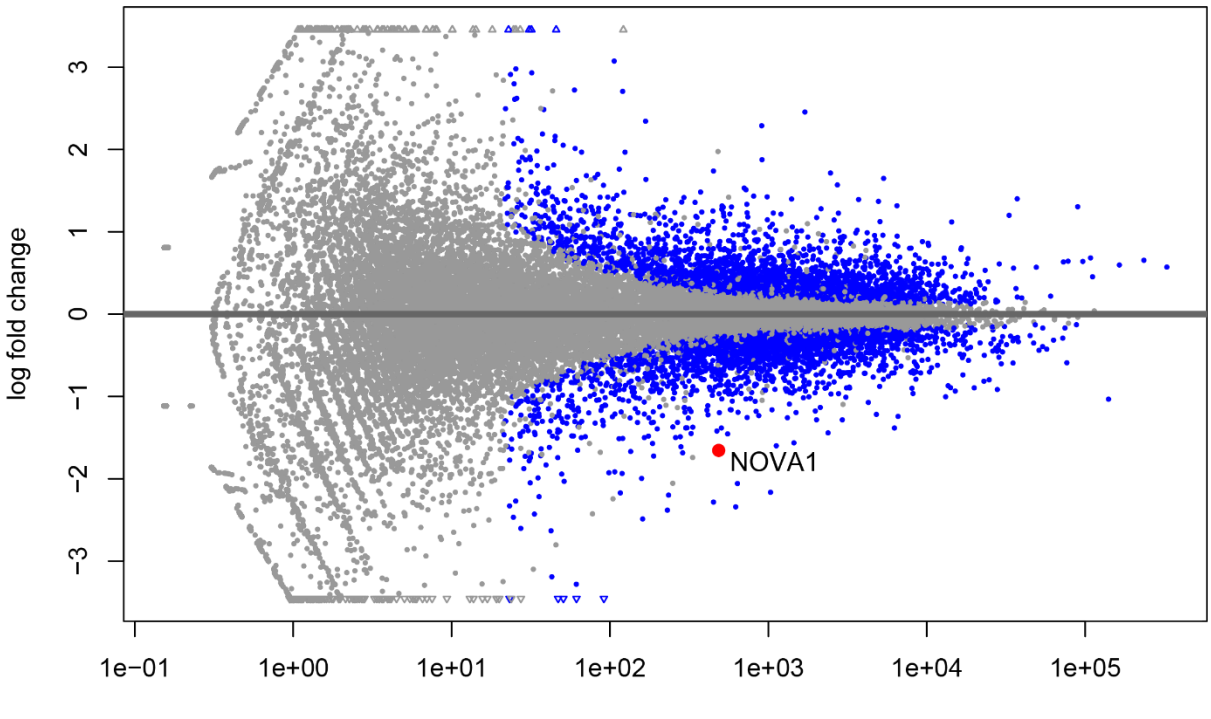


Figure 5.2: MA plot for the RNA sequencing data from NOVA1 depletion. *This plot shows average gene expression level and fold change in knockdown vs control conditions for NOVA1. Grey dotted area represents the genes that were not significant. Blue dots show the events that had p-value < 0.05. The genes below the zero-point line are the genes downregulated and genes above the line are upregulated genes. Red dot among downregulated genes is NOVA1*

MA plot in Figure 5.2 and heatmap Figure 5.3. Importantly, the *NOVA1* gene was among the genes downregulated as can be seen in Figure 5.2.

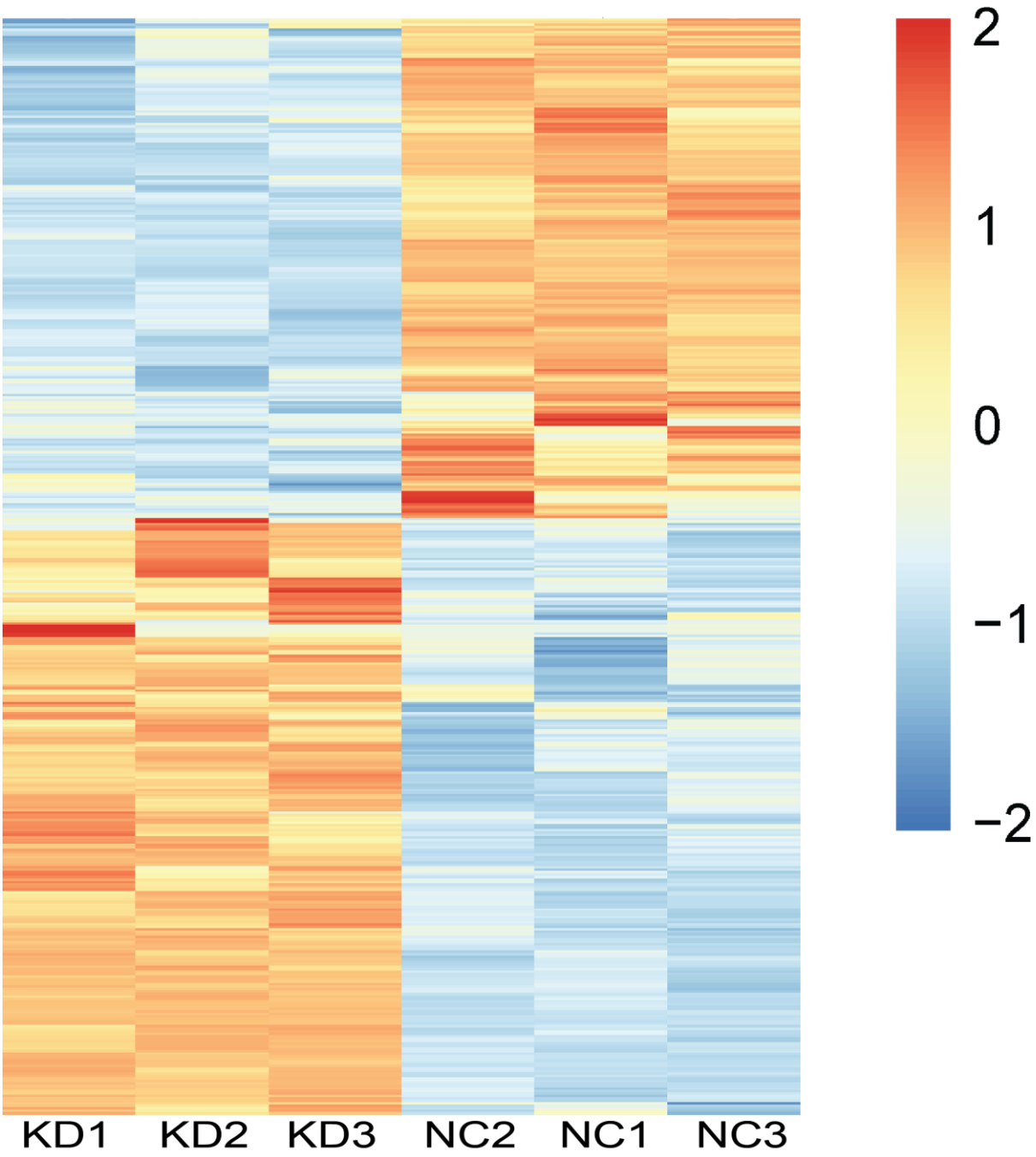


Figure 5.3: Heatmap of the genes with upregulation and downregulation after NOVA1 depletion.

The volcano plot shown in Figure 5.4 also confirms downregulation of the *NOVA1* gene in the PC3 prostate cancer cell lines after siRNA-mediated depletion, with significant p-value. The volcano plot Figure 5.3 shows that the *NOVA1* gene signal is detected on the left side of the plot along with the downregulated genes, whereas the genes on the right side of the volcano plot indicate the genes that have been upregulated. I also confirmed this by examining aligned RNAseq reads on the IGV genome browser. Consistent with its depletion, the *NOVA1* gene had a lower aligned read count in the *NOVA1* knockdown cells compared to the control PC3 cells (Figure 5.5A). In comparison, a close paralog of *NOVA1* known as *NOVA2* (Figure 5.5B) does not seem to have much of a change in aligned reads, which indicates that *NOVA1* and *NOVA2* might not depend on each other for expression within PC3 cells.

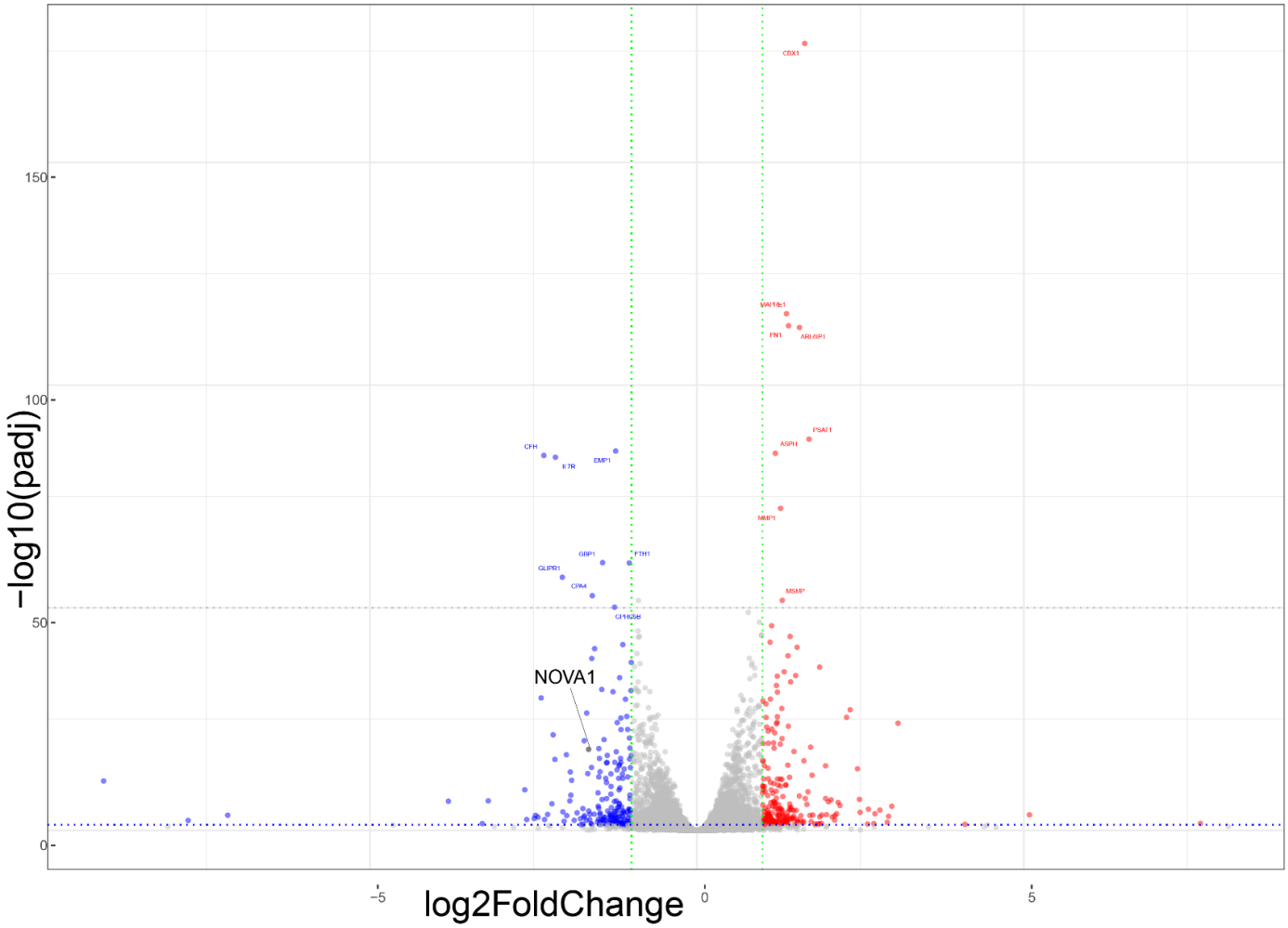


Figure 5.4: Volcano plot from DESeq2. *Volcano plot shows the genes in NOVA1 knockdown vs control RNA-Seq dataset. Genes that are seen on the right in red are the upregulated genes. Blue dots on the left are the downregulated genes. NOVA1 can be seen in this volcano plot as downregulated. The most significant genes are towards the top of the plot.*

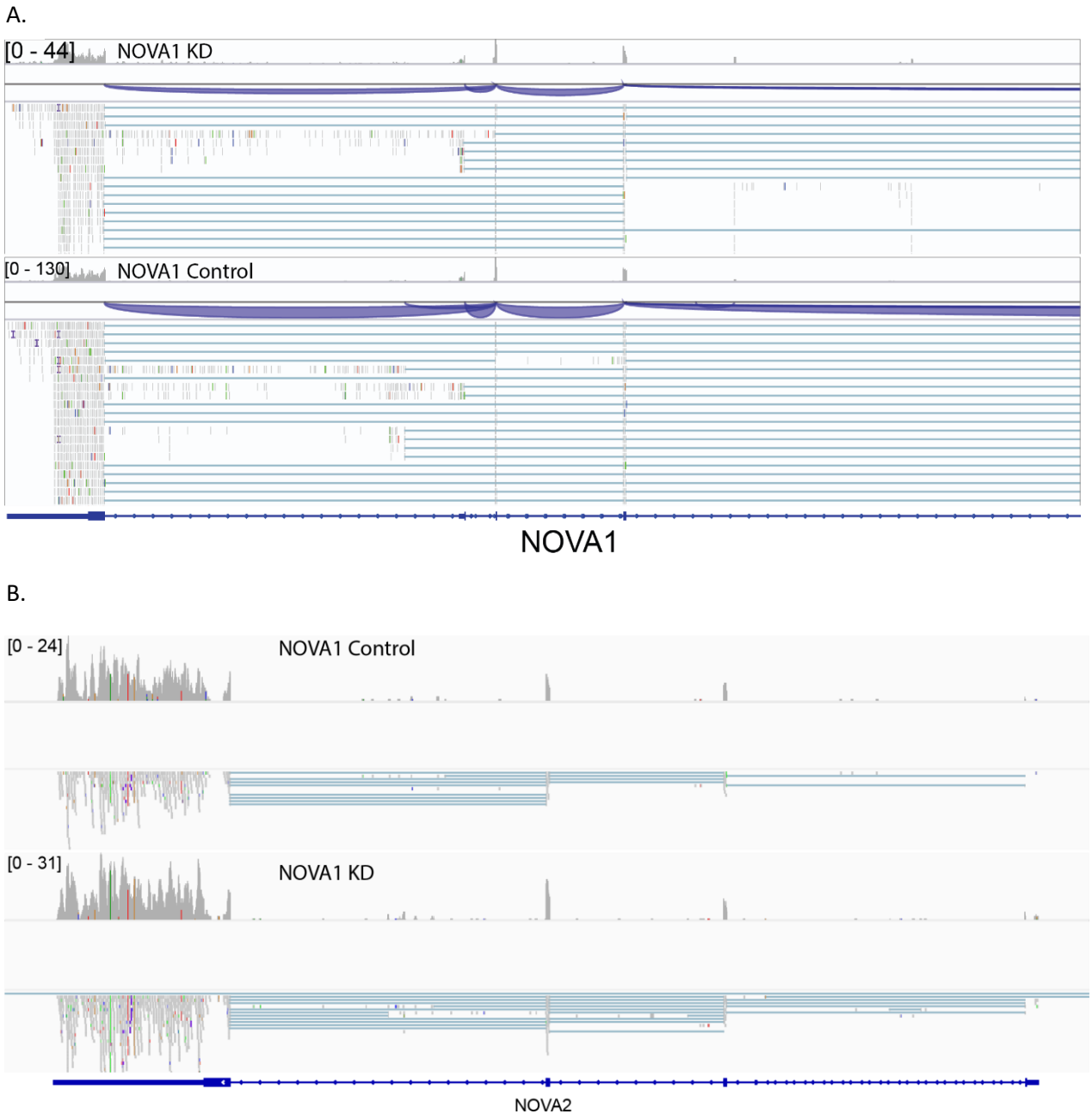


Figure 5.5: Screenshot taken from integrative genome viewer. A. *NOVA1 IGV screenshot shows NOVA1 KD (knockdown) tracks on the top and NOVA1 control track at the bottom. Comparing both tracks it can be seen that NOVA1 gene has low read counts within the depleted sample indicating NOVA1 gene knockdown whereas control track shows higher read count for NOVA1 gene.* ***B.*** *For comparison, I also used the RNAseq data to compare the pattern of NOVA2 expression. This shows similar levels after NOVA1 depletion compared to control.*[203]

In addition to NOVA1, the Volcano plot in Figure 5.6 highlighted other genes that have been up or down regulated in response to the NOVA1 knockdown. Amongst these genes there was an isolated event at the top right, which shows strong upregulation of a highly significant

event. This event corresponds to the *CBX1* gene. The *CBX1* gene (also called Chromobox 1) is a protein coding gene that encodes a nonhistone protein that is highly conserved and is a member of heterochromatin protein family. CBX1 protein is associated with centromeres and plays a role in epigenetic control of chromatin structure and gene expression.[194] Visualization of my RNAseq data on the IGV browser in Figure 5.6 confirmed that the *CBX1* gene is repressed by *NOVA1*. The changes in expression of *CBX1* may indicate that the resulting downstream gene expression changes observed after *NOVA1* depletion might be due to an overall change in chromatin. On the contrary, one of the most downregulated genes following *NOVA1* knockdown is *CFH* (Complement Factor H). *CFH* is a member of a gene cluster that play a vital role in Regulation of Complement Activation (RCA) and is associated with diseases such as haemolytic-uremic syndrome and chronic hypocomplementemic nephropathy. The *CFH* gene is involved in pathways such as complement activation, proteolysis, and in regulation of complement dependent cytotoxicity.[195–197]

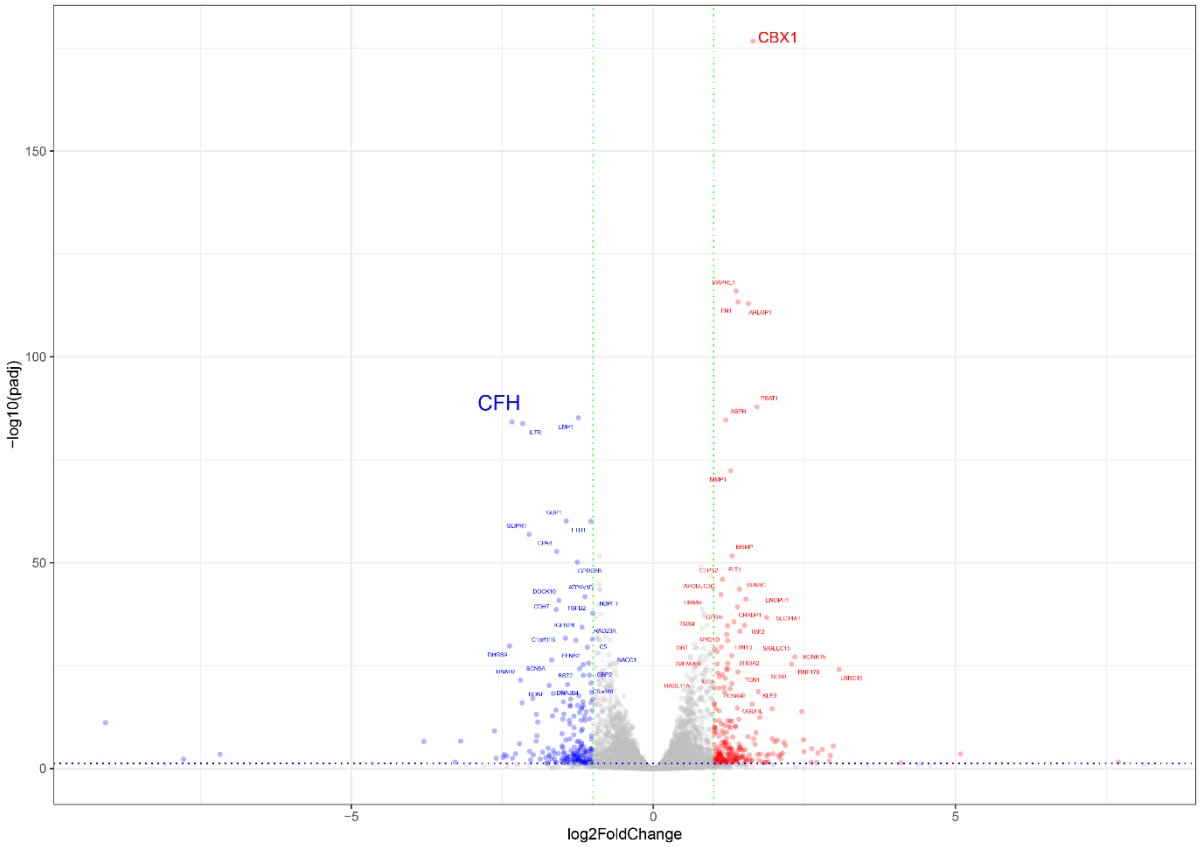


Figure 5.6: Volcano plot from DESeq2. *Volcano plot shows the most significant upregulated gene CBX at the top right whereas the most significantly downregulated gene CFH can be seen on the left.*

Further, analysing the impact of genes such as *CBX1* and *CFH* using Psichomics show that higher expression of *CBX1* (repressed by NOVA1, activated by ESRP1 and ESRP2) correlates with low patient survival whereas, lower *CFH* (activated by NOVA1) gene expression correlates with poor survival. This indicates that elevated levels of NOVA1 lead to better chances of patient survival for these two genes.

Further to the differential gene expression analysis, GSEA and gene ontology was done to identify the molecular functions of the genes that were either up or down regulated as a result of NOVA1 depletion. The molecular function GO terms associated with genes that significantly change expression after *NOVA1* knockdown included cold-induced thermogenesis, RNA splicing, and regulation of alternative mRNA splicing via spliceosome.

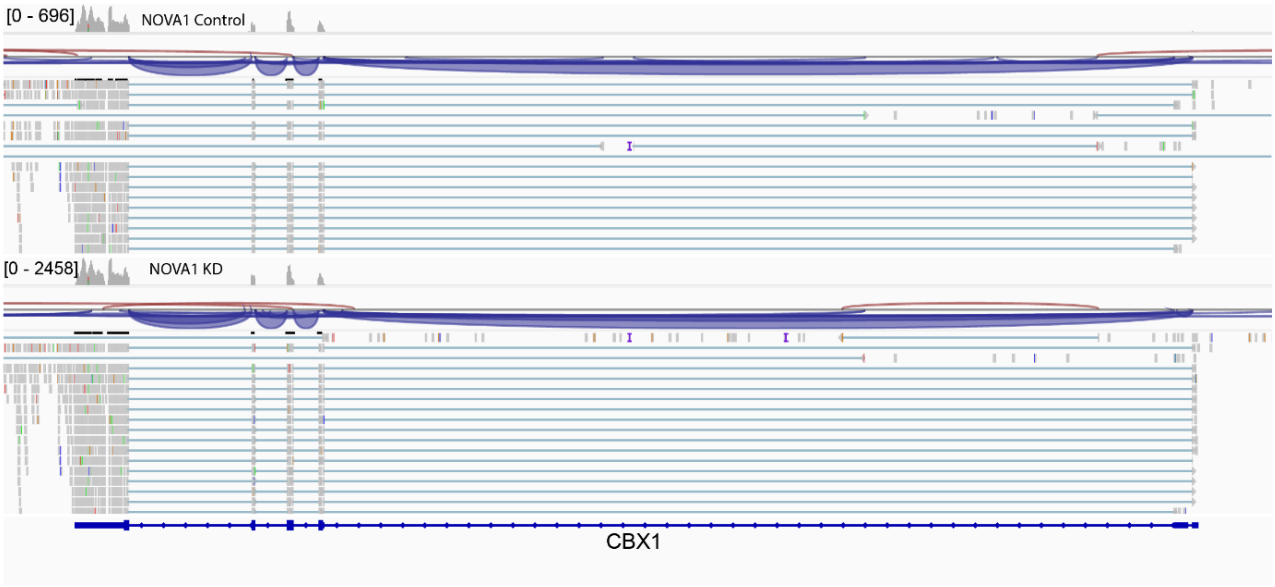


Figure 5.7: IGV Screenshot of CBX1 gene. *The figure shows IGV screenshot for NOVA1 RNA-Seq dataset with NOVA1 control tracks at the top and NOVA1 KD tracks at the bottom for CBX1 gene. It can be seen that CBX1 has high read count in NOVA1 knockdown indicating NOVA1 gene expression leads to the repression of CBX1 (Chromobox1 gene).*[204]

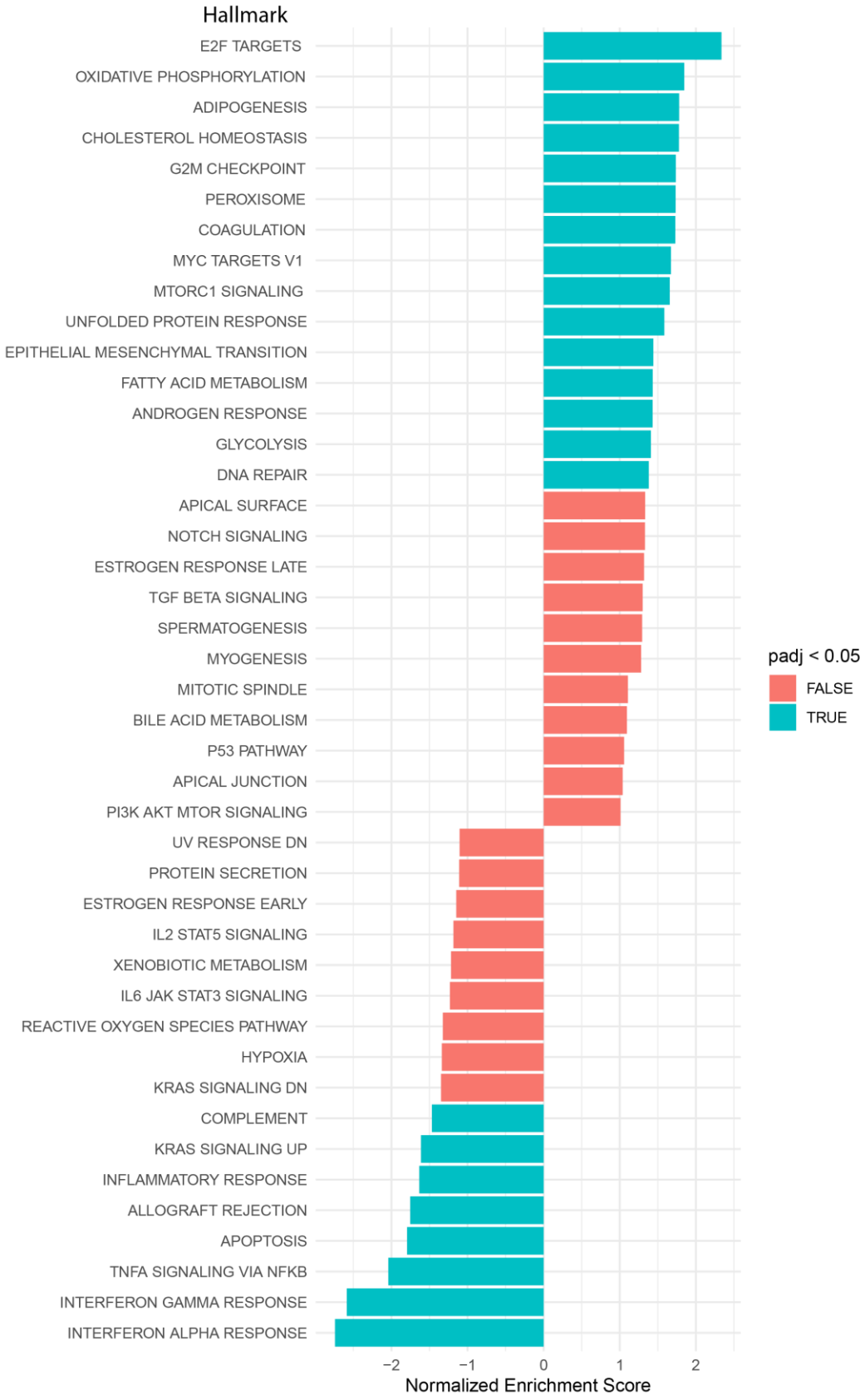


Figure 5.8: Gene Set Enrichment Analysis (GSEA). *GSEA shown in the figure is based on the gene list obtained following the DESeq2. The red colour shows hallmarks that were associated with the genes differentially expressed but had p adjusted value of > 0.05 whereas blue colour shows the genes related to hallmarks that were < 0.05 thus highly significant.*

Table 5.1: List of genes correlated to EMT and regulated by NOVA1.

S. No	Genes	NOVA1 Regulation	ESRP1/2 Regulation
1.	*FN1*	Repressed	Activated
2.	*MMP1*	Repressed	--
3.	*SPARC*	Repressed	--
4.	*NT5E*	Repressed	--
5.	*COL4A1*	Repressed	--
6.	*LRRC15*	Repressed	--
7.	*COL4A2*	Repressed	Repressed--
8.	*PTHLH*	Repressed	--
9.	*ITGB5*	Repressed	Activated
10.	*CTHRC1*	Repressed	--
11.	*SDC1*	Repressed	Activated
12.	*CALU*	Repressed	--
13.	*COL6A3*	Repressed	--
14.	*LOX*	Repressed	Activated
15.	*MMP14*	Repressed	--
16.	*COL5A1*	Repressed	--
17.	*ITGB3*	Repressed	--
18.	*COLGALT1*	Repressed	--
19.	*FBN1*	Repressed	Repressed
20.	*HTRA1*	Repressed	Activated
21.	*TGFB1*	Repressed	--
22.	*MMP3*	Repressed	Repressed
23.	*P3H1*	Repressed	--
24.	*APLP1*	Repressed	Activated
25.	*ANPEP*	Repressed	--
26.	*FBLN1*	Repressed	--
27.	*ITGB1*	Repressed	Repressed
28.	*FZD8*	Repressed	--

29.	*CCN2*	Repressed	--
30.	*MCM7*	Repressed	Activated
31.	*ITGA2*	Repressed	--
32.	*COL6A2*	Repressed	--
33.	*COL1A2*	Repressed	--

In a broader picture, GSEA reveals a link between NOVA1 regulated genes and critical pathways such as G2M, EMT, and E2F targets. GSEA revealed thirty-three genes listed in Table 5.1 related to epithelial-to-mesenchymal (EMT) pathway. All of these 33 EMT-related genes were repressed by NOVA1, whereas seven of these genes were activated by ESRP1 and ESRP2 and four were found to be repressed by ESRP1 and ESRP2. Moreover, GSEA in Figure 5.7 also revealed a correlation between NOVA1 regulated eighty genes to hallmarks such as G2M checkpoint. This suggests NOVA1 expression influences a checkpoint in cell cycle that does not allow mitosis to happen in a cell that has a DNA damage until DNA damage has been repaired. Interestingly, the *CBX1* gene shown in Figure 5.5 and Figure 5.6 was one of the genes associated with the G2M checkpoint, and it was also the most upregulated gene in NOVA1 knockdown. Additionally, ninety-six of the *NOVA1* regulated genes were linked to E2F targets, which has a role in cell proliferation. G2M, EMT, and E2F targets have links with DNA damage, DNA repair, metastasis, and cell proliferation, important in cancer since their dysregulation ultimately leads to cancer development and progression. It is also interesting that NOVA1 regulated genes associate to the similar pathways such as G2M and EMT that have been previously noted for involvement of ESRP1 and ESRP2 shown in a review by Advani et al.[198] Additionally, gene ontology analysis revealed a high number of genes correlated to the pathways such as chemotaxis, leukocyte migration, leukocyte proliferation, immune response, cell adhesion, cytokine production, bacterial response, compound transport, blood vessel development & angiogenesis, and extracellular structure as shown in Figure 5.9 and Figure 5.10.

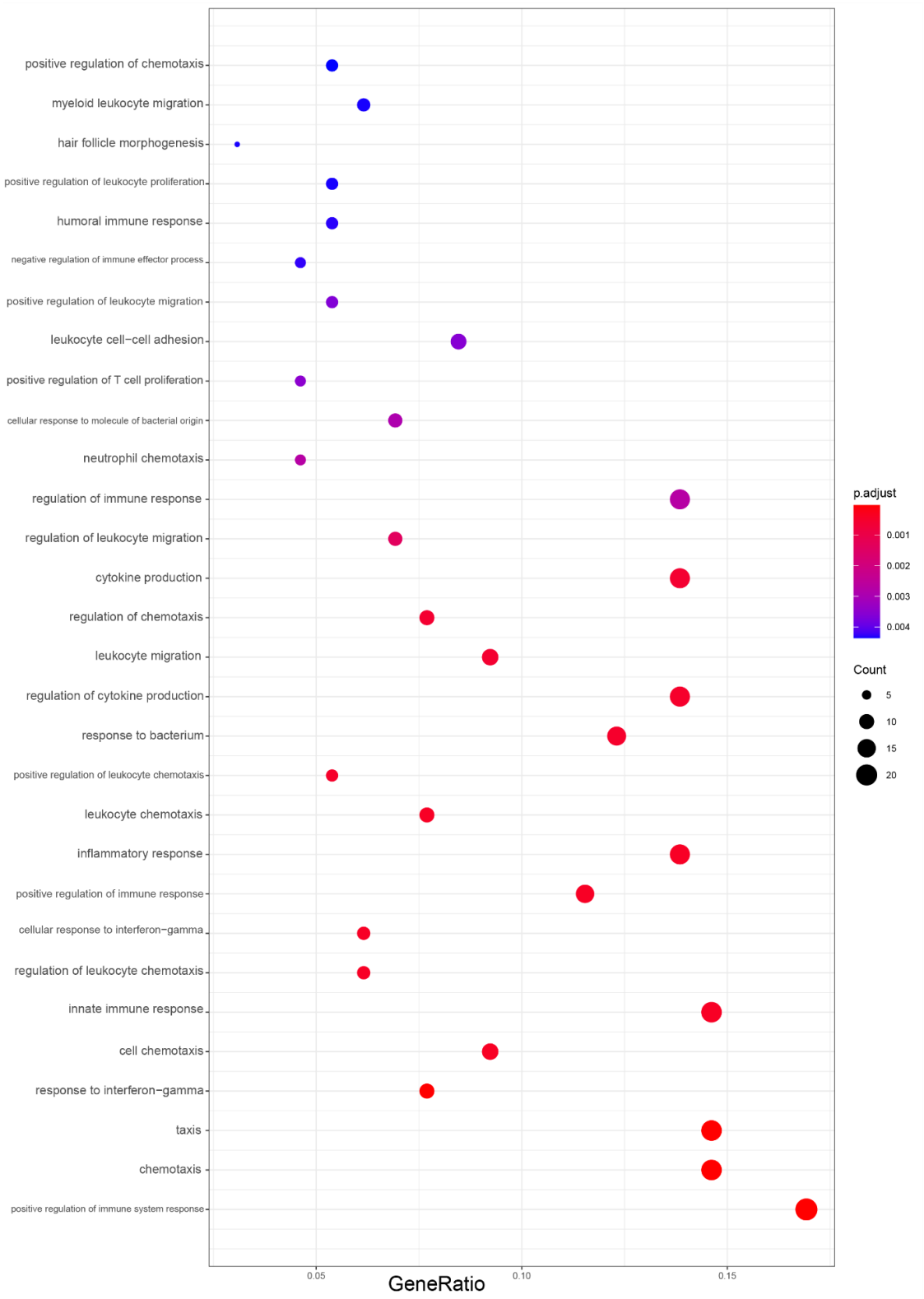


Figure 5.9: Gene Ontology of downregulated genes after NOVA1 depletion.

Figure 5.10: Gene Ontology of up regulated genes after NOVA1 depletion.

5.3.3 Cell Colonization Assay to test prediction that NOVA1 influences cell proliferation.

To test predictions from the differential gene expression analyses I did further experiments. GSEA predicted that *NOVA1* regulated the expression of E2F target genes that are linked to cell proliferation. To thus assess whether NOVA1 knockdown has consequences on the proliferation of prostate cancer cells, I performed a cell behaviour assay and looked at cell proliferation. To investigate the prostate cancer cells growth comparing control and NOVA1 knockdown samples, a cell colonization assay was performed. This proliferation experiment was done on PC3 cells that had been knocked down for NOVA1 gene expression for 72 hours using siRNA. The cells were then seeded on six different well plates. Three 6-well plates for knockdown and three 6-well plates for control cells were prepared. These well plates were incubated for 14 days while changing media weekly. Then the cells were treated with crystal violet dye on the fourteenth day to visualize the colonies that had formed.

The findings, depicted in Figure 5.11, revealed a significant difference in cell growth. NOVA1 knockdown prostate cancer cells exhibited significantly less growth compared to the control cells.

5.3.4 Analysis of splicing patterns after NOVA1 depletion.

SUPPA2 analysis identified 9122 differential splicing events, which were then filtered to reveal more significant hits based on a p-value less than 0.05. This filtration of events based on p-value parameters returned 404 gene events. Thereafter, splicing events regulated by NOVA1 were also looked for in the *ESRP2* gene target list, to reveal 14 events that were common between two sets listed in Table 5.2. The 404 events with significant differential splicing were then visualized using the IGV genome browser, and 127 gene events were seen to be changing with sufficient amplitude to be obvious. Only one of these strongly changing genes was also found in the previously published ESRP splicing target list [199]. This indicates that a total of 126 novel targets are regulated by NOVA1 that were not in the list of ESRP target list from Munkley et al. 2019.[199] These 127 novel targets are shown in Table 5.3. These splicing target genes included the following examples of particular interest.

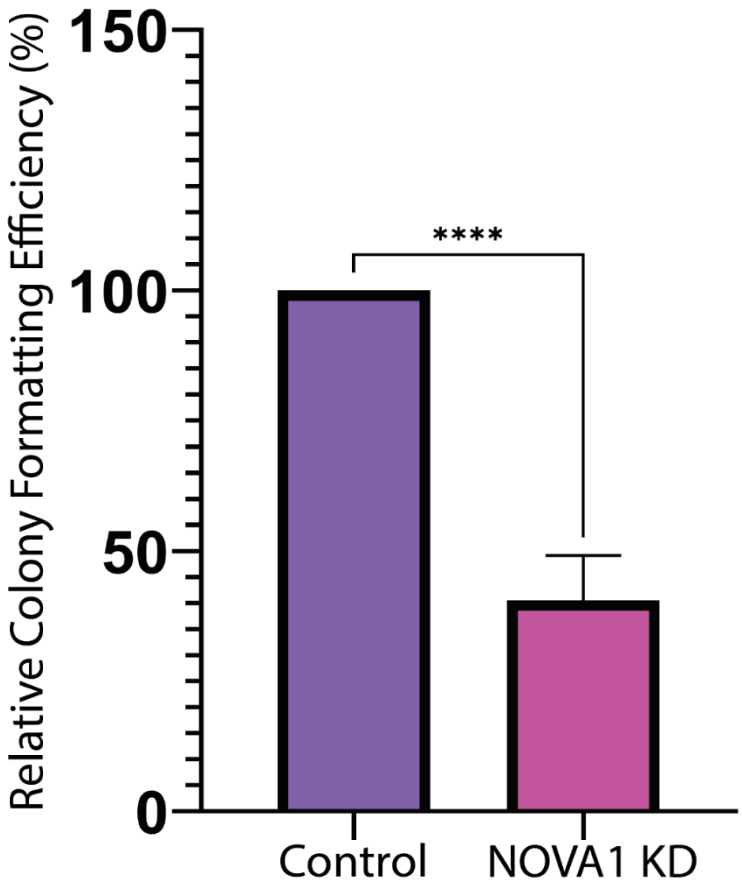


Figure 5.11: Cell Colony Formation in Control and Knockdown Conditions. *Bar graph in this figure represents the cell colony formation where cells were allowed to grow for 2 weeks after knockdown of an initial equal number of cells. As shown in graph there were higher number of cells grown in colonies in control samples compared to NOVA1 gene knockdown samples. The results were statistically significant P < 0.05. The graph was made using Prism GraphPad 9.*

Table 5.2: Common splicing events between NOVA1 and ESRP1/2 knockdown.

S. No.	Event	Event Type	Symbol
1.	ENSG00000089902.9;SE:chr14:102721084-102721320:102721377-102722187:+	SE	*RCOR1*
2.	ENSG00000126903.15;RI:chrX:154487306:154488496-154488584:154489082:-	RI	*SLC10A3*

3.	ENSG00000163257.10;AL:chr4:17804312:17804756-17810447:17809425:17809692-17810447:-	AL	*DCAF16*
4.	ENSG00000120925.15;AF:chr8:42887871-42896484:42896588:42887871-42897240:42897290:-	AF	*RNF170*
5.	ENSG00000120925.15;AF:chr8:42887871-42896484:42896706:42887871-42897240:42897290:-	AF	*RNF170*
6.	ENSG00000120925.15;AF:chr8:42887871-42896484:42896721:42887871-42897240:42897290:-	AF	*RNF170*
7.	ENSG00000120925.15;AF:chr8:42887871-42896484:42896723:42887871-42897240:42897290:-	AF	*RNF170*
8.	ENSG00000164442.9;AF:chr6:139373952-139374132:139374213:139373952-139374264:139374362:-	AF	*CITED2*
9.	ENSG00000164442.9;AF:chr6:139373952-139374264:139374362:139373952-139374413:139374620:-	AF	*CITED2*
10.	ENSG00000185885.15;AF:chr11:313853:314356-314922:314418:314456-314922:+	AF	*IFITM1*
11.	ENSG00000003436.15;AF:chr2:187503770-187513616:187513653:187503770-187554200:187554492:-	AF	*TFPI*
12.	ENSG00000003436.15;MX:chr2:187503770-187513616:187513646-187554200:187503770-187529364:187529485-187554200:-	MX	*TFPI*
13.	ENSG00000147676.13;AF:chr8:119207949:119208093-119221587:119208868:119209200-119221587:+	AF	*MAL2*
14.	ENSG00000089902.9;SE:chr14:102721084-102721320:102721377-102722187:+	SE	*RCOR1*

NOVA1 controls splicing of the *FGFR1OP2* gene. On analysis of differential splicing using SUPPA2, one gene with a significant visible change was *FGFR1OP2* (Fibroblast Growth Factor Receptor 1 Oncogene Partner 2). Using the junction reads on a sashimi plot (Figure 5.12), the PSI of *FGFR1OP2* exon 5 was calculated (using the equation PSI = exon 5 inclusion read junctions / exon 5 inclusion + exon 5 exclusion read junctions). This analysis showed that *FGFR1OP2* had a PSI of 79 in control cells, compared to lower PSI of 59 following *NOVA1* knockdown. To additionally validate *FGFR1OP2* exon 5 as a splicing target controlled by NOVA1, capillary gel electrophoresis Figure 5.13A, was performed after RT-PCR analysis. Unfortunately, this did not give a clear signal from 2/3 of the controls, but I was still able to quantitate a signal shown in Figure 5.13B. This confirmed that NOVA1 depletion causes significant repression of exon 5 inclusion, indicating that NOVA1 protein activates this exon.

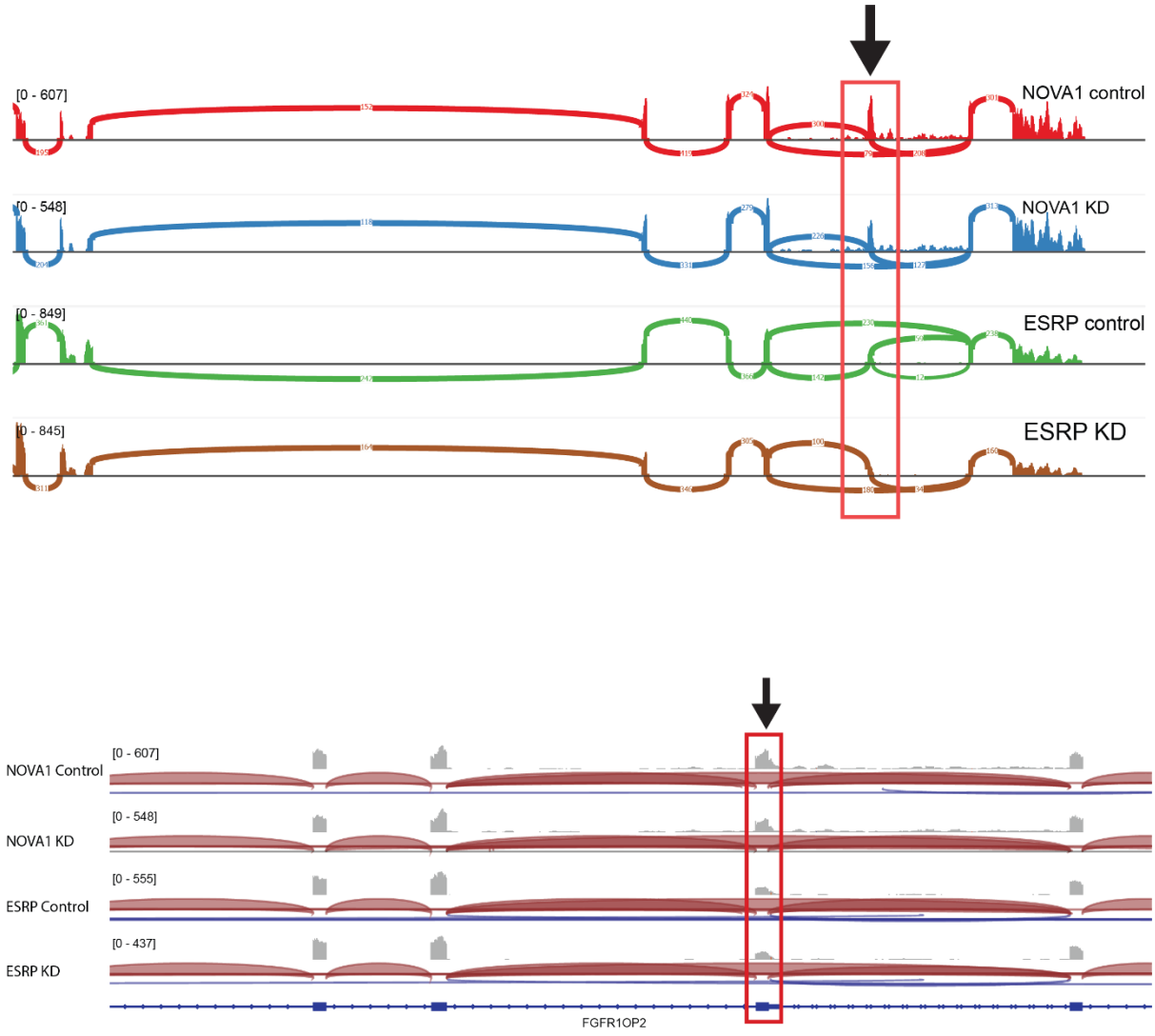


Figure 5.12: Sashimi plot (Top) and IGV screenshot (Bottom) for FGFR1OP2 gene. *The red box shows exon 5 of FGFR1OP2 gene in NOVA1 control and knockdown and in ESRP1/2 knockdown.*

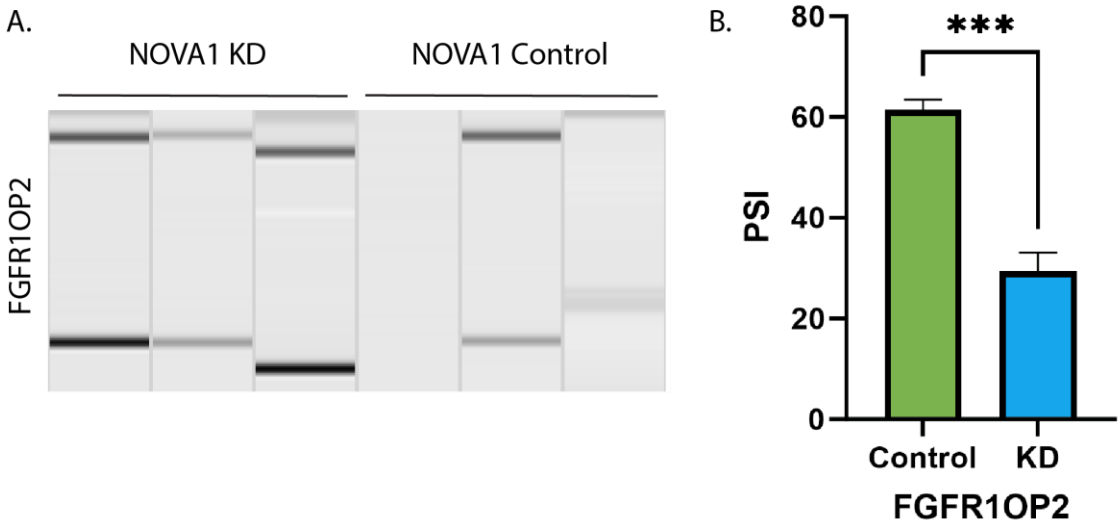


Figure 5.13: RT-PCR results and PSI calculated. A. *Capillary gel electrophoresis image shows bands for FGFR1OP2 in NOVA1 knockdown vs NOVA1 control cDNA RT-PCR.* ***B.*** *Shows the calculated PSI levels for FGFR1OP2. The *** indicates the significant p-value. Bar graph was created using GraphPad PRISM.*

NOVA1 controls splicing of the *MYL6* gene. Exon 6 of *MYL6* (Myosin Light Chain 6) was another event identified by SUPPA2, and amongst the list of 129 genes confirmed through visual analysis of the IGV genome browser. Splicing inclusion of exon 6 of *MYL6* was repressed by NOVA1 (shown in Figure 5.14 A and B). Quantitative analysis of *MYL6* exon 6 splicing inclusion was done using the read counts from the sashimi plot. The PSI level for *MYL6* exon 6 in the control samples was calculated at 51 PSI. In contrast, a 75 PSI was calculated after NOVA1 knockdown, indicating that splicing of this exon is normally repressed by NOVA1. To validate this predicted splicing change, primers were designed to be complementary to exon 5 and exon 7 of *MYL6* and used for RT-PCR analysis. The results from this RT-PCR analysis, followed by capillary gel electrophoresis are shown in Figure 5.14C. This further supported the observation made using IGV where high levels of MYL6 exon 6 were noted in NOVA1 control samples compared to NOVA1 depletion and used to calculate an experimental PSI shown in Figure 5.14D.

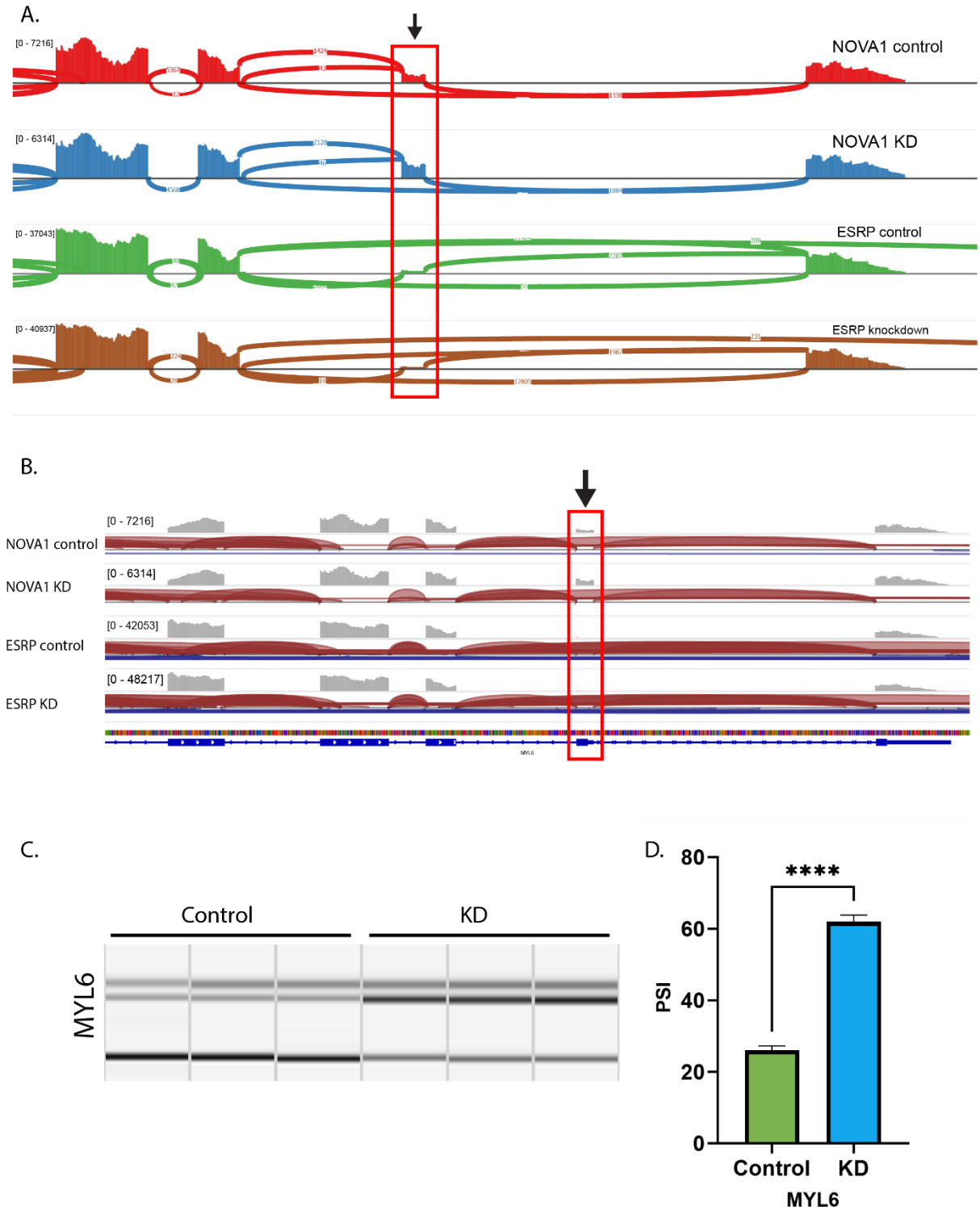


Figure 5.14: Analysis of MYL6 gene splicing. A. *Sashimi plot arrow points to exon 6.* ***B.*** *IGV screenshot arrow points to exon 6.* ***A.*** *and* ***B.*** *show the tracks for NOVA1 compared to ESRP and in this case MYL6 does not seem to be affected by ESRP1 and ESRP2 knockdown.* ***C.*** *Capillary gel electrophoresis image from QiAxcel.* ***D.*** *PSI levels of MYL6 in control vs knockdown of NOVA1 and *** indicates a significant p-value.*

NOVA1 controls splicing of the *RHBDD2* gene. The *RHBDD2* was also predicted to have splicing changes controlled by NOVA1. Exon 2 of *RHBDD2* was found to have splicing change in response to NOVA1 knockdown compared to control. RHBDD2 appears to be activated by *NOVA1* gene expression Figure 5.15 A and B. I experimentally validated this splice event using RT-PCR shown in Figure 5.15C. Figure 5.15D shows the calculated PSI level in the bar graph. Higher levels of inclusion of exon 2 of *RHBDD2* were seen in the NOVA1 control sample. *RHBDD2* exon 2 had a PSI of 82 vs 75 PSI in control. This observation suggested that NOVA1 promoted the inclusion of exon 2 in *RHBDD2.*

NOVA1 controls splicing of transcripts from the *GEN1* gene. *GEN1* was an interesting splicing target for NOVA1 control. A splicing change was predicted for an un-annotated exon between exon 4 and 5 in the *GEN1* gene after NOVA1 depletion, when I first carried out alignment of the RNAseq reads. However, this region of the genome has been now annotated, identifying this splice change as exon 4 of the GEN1 gene on IGV. This exon 4 appeared to be activated by NOVA1 shown in Figure 5.16 A and B. To validate this splicing change, I carried out an RT-PCR analysis. Capillary gel electrophoresis picture shown in Figure 5.16C and
PSI calculated shown in Figure 5.16D. A significant splicing change was seen in *GEN1* exon 4, with lower PSI levels in the NOVA 1 knockdown sample.

NOVA1 controls splicing of *ESYT2* gene. One of the initial events to be observed to have altered splicing due to NOVA1 expression was exon 14 of *ESYT2* shown in figure 5.17A and figure 5.17B. I further validated this splicing change using RT-PCR and confirmed higher exon 14 PSI in the control versus the NOVA1 knockdown PC3 cells figure 5.7C and figure 5.17D. *ESYT2* was found to have two different splicing events exons skipping and alternative first exon splicing event.

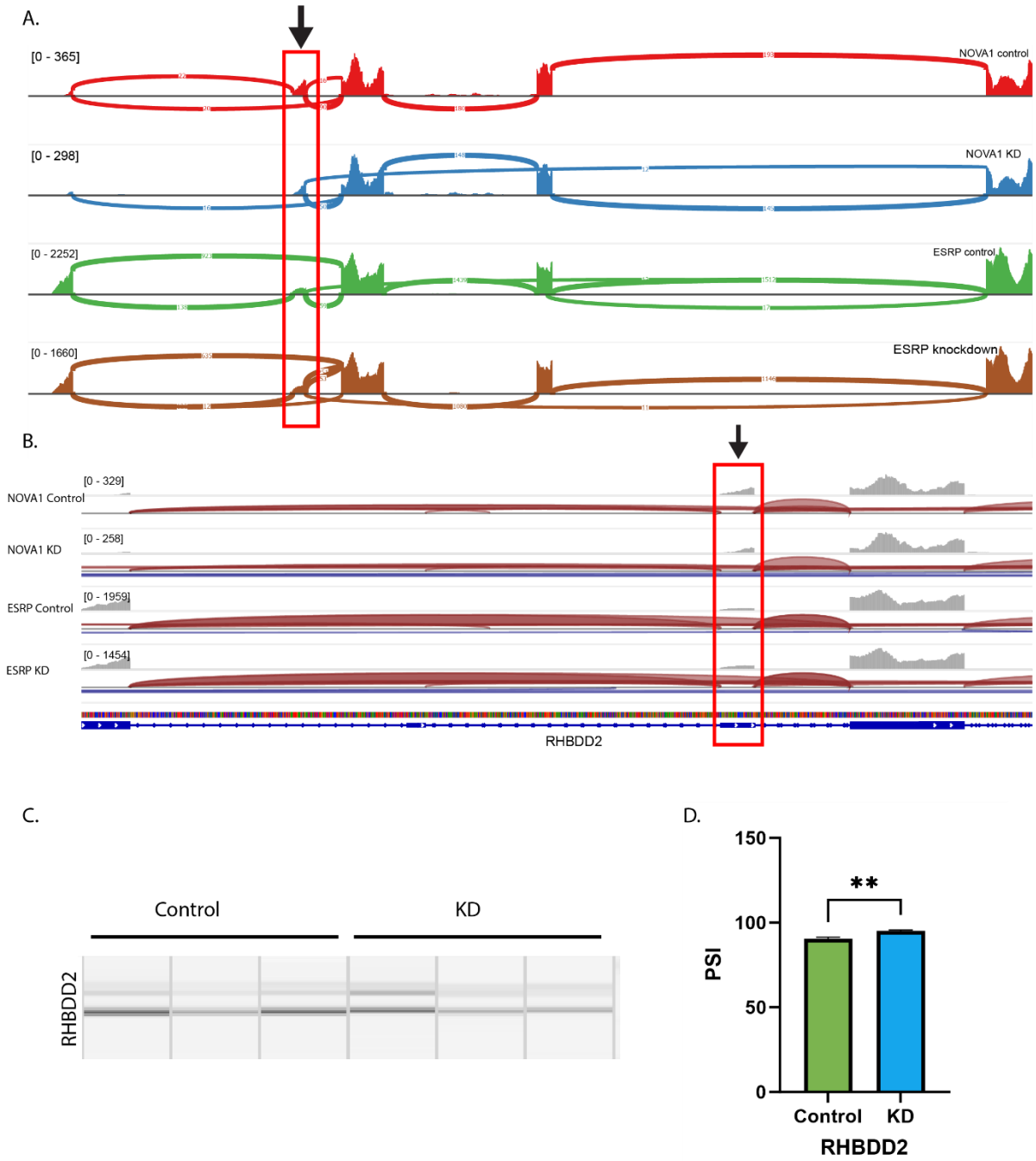


Figure 5.15: Analysis of RHBDD2 gene splicing. A. *Sashimi plot arrow points to exon 2.* ***B.*** *IGV screenshot arrow points to exon 2.* ***C.*** *Capillary gel electrophoresis image from QiAxcel.* ***D.*** *PSI levels of RHBDD2 in control vs knockdown of NOVA1 and ** indicates a significant p-value.*

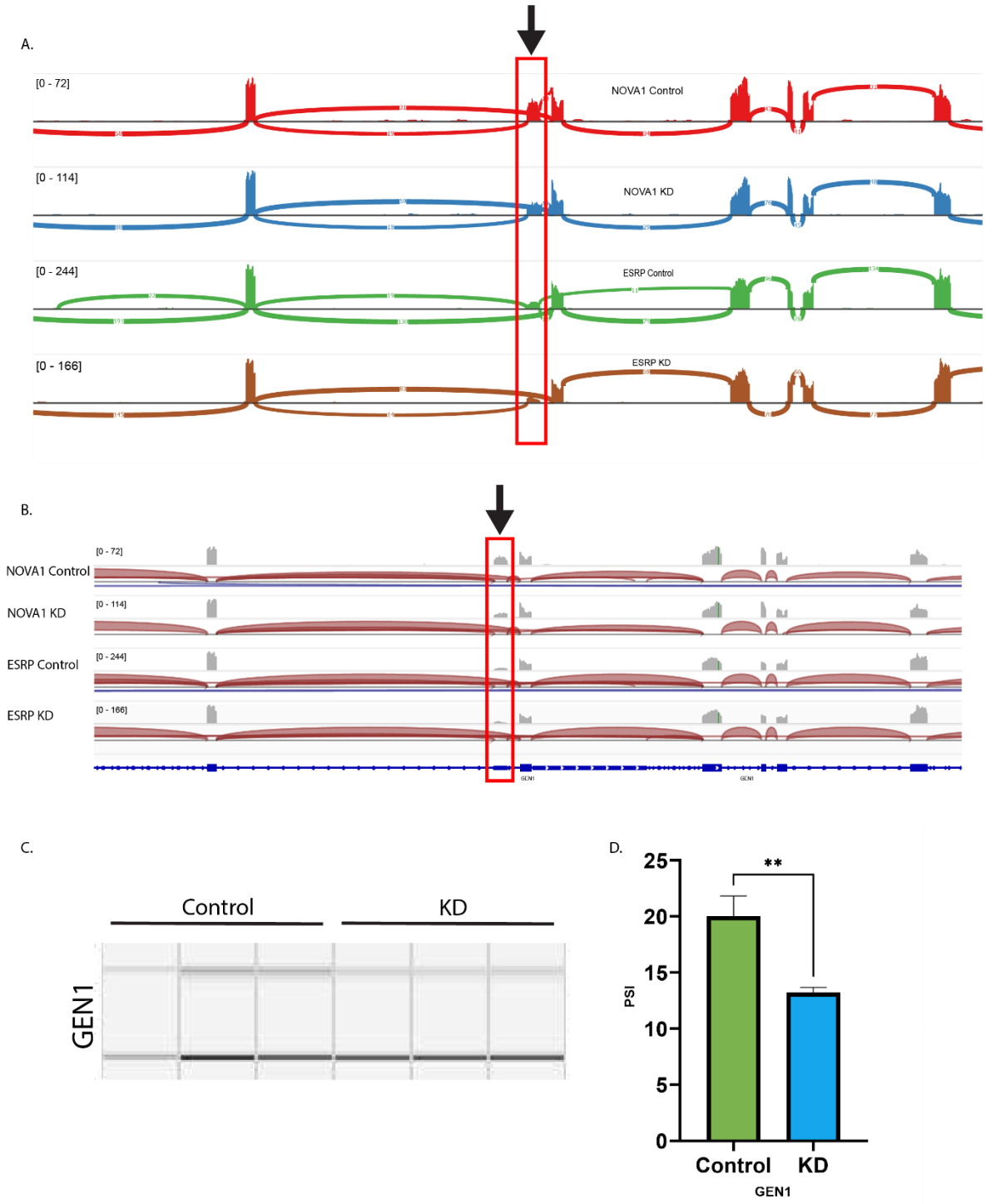


Figure 5.16: GEN1 gene. A. *Sashimi plot arrow points to exon 4.* ***B.*** *IGV screenshot arrow points to exon 4.* ***C.*** *Capillary gel electrophoresis image from QiAxcel.* ***D.*** *PSI levels of GEN1 in control vs knockdown of NOVA1 and ** indicates a significant p-value.*

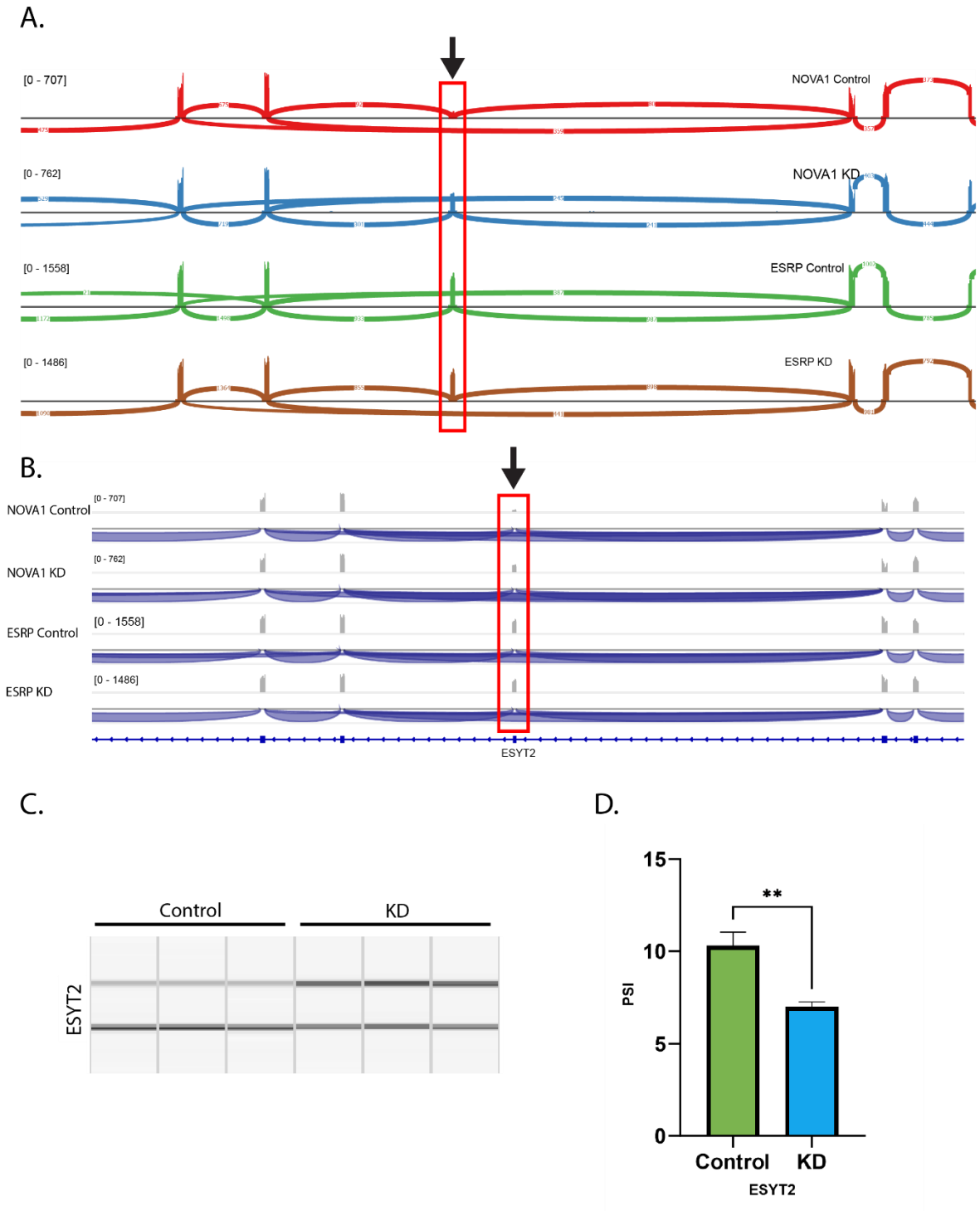


Figure 5.17: Analysis of ESYT2 gene splicing. A. *Sashimi plot arrow points to exon 14.* ***B.*** *IGV screenshot arrow points to exon 14.* ***C.*** *Capillary gel electrophoresis image from QiAxcel.* ***D.*** *PSI levels of ESYT2 in control vs knockdown of NOVA1 and ** indicates a significant p-value.*

To finally visualize how these validated NOVA1-regulated splicing events might correlate with patient prognosis, I used the psichomics analysis programme to correlate splicing patterns with the time to first biochemical recurrence figure 5.18A-D. The biochemical recurrence is of particular importance as biochemical recurrence of prostate cancer usually correlates with castrate resistant disease, and thus leading to higher treatment resistance therefore, increased mortality. I found that higher inclusion of *FGFR1OP2* exon 5 (the form driven by NOVA1 expression) correlated with less time to biochemical recurrence (with a PSI cut off 0.44 and a p-value less than 0.05). Similarly, high splicing inclusion of *MYL6* exon 6 (PSI cut off levels and a p-value) also correlated with a decreased time to biochemical recurrence. Higher splicing inclusion of *RHBDD2* exon 2 (PSI cut off levels and a p-value) correlated with a lower time to biochemical recurrence. Additionally, a correlation between *ESYT2* exon skipping and biochemical recurrence was detected, with a PSI cut off 0.82 and a p-value less than 0.05. It

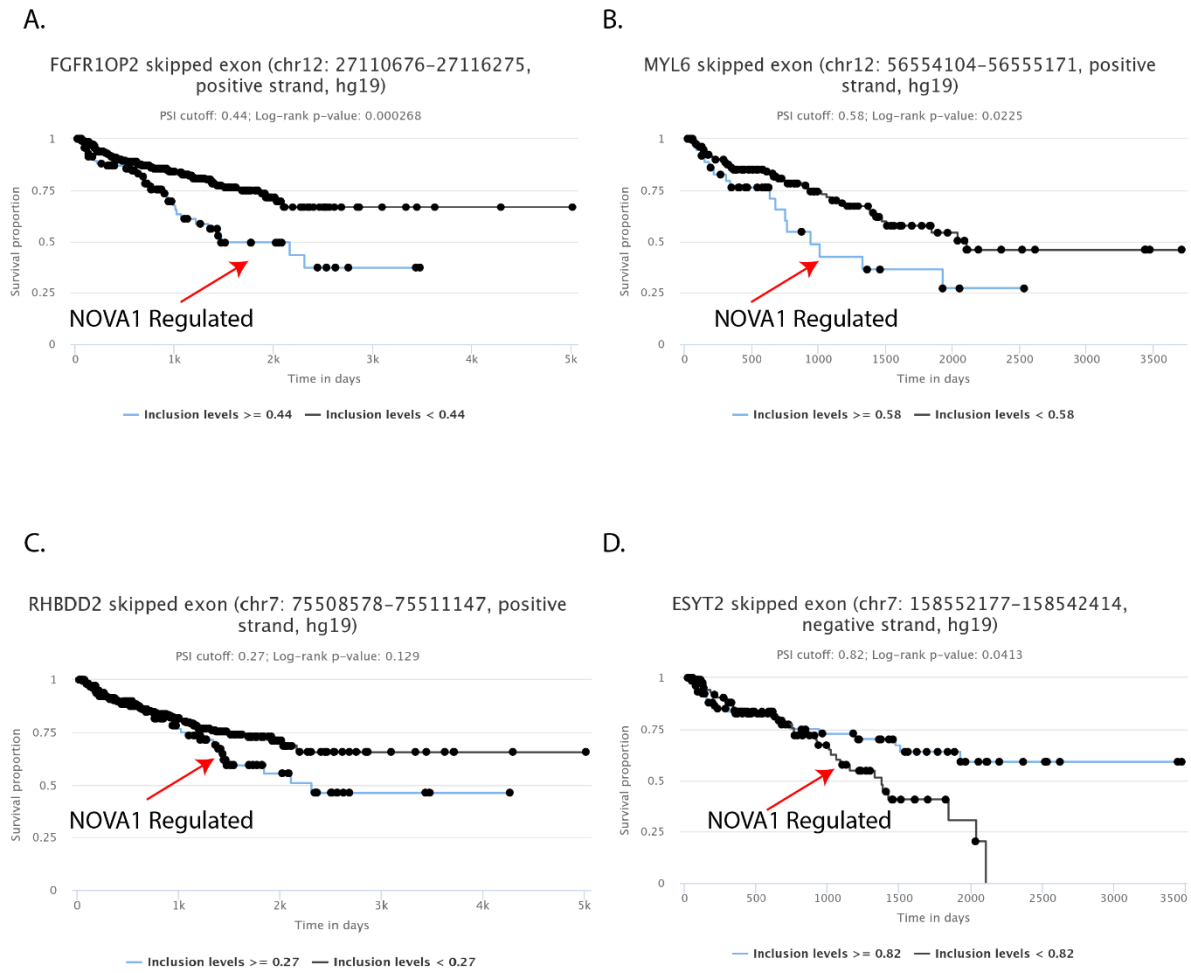


Figure 5.18: Survival graphs with time to first biochemical recurrence for specific splicing targets of NOVA1. A. *Survival graph for FGFR1OP2.* ***B.*** *Survival graph for MYL6.****C.*** *Survival graph for RHBDD2.* ***D.*** *Survival graph for ESYT2.*

was seen that higher inclusion level (the NOVA1 expression regulated event) led to an increased time to biochemical recurrence. Taken together, these NOVA1-regulated splicing events correlated with disease progress. Psichomics did not allow analysis for *GEN1*, possibly because of the previously un-annotated exon. SUPPA2 identified GEN1 having an alternative 3 prime splice site splicing.

CBX1 and *CFH* genes that were either highly up and downregulated in NOVA1 knockdown were analysed using Psichomics. The survival analysis through Psichomics in Figure 5.19 and Figure 5.20 show that higher expression of *CBX1* (repressed by NOVA1, activated by ESRP1 and ESRP2) correlates with low patient survival whereas, lower *CFH* (activated by NOVA1) gene expression correlates with poor survival.

Additionally, SET was another protein-coding gene identified from the differential splicing analysis. SET was seen to be repressed by NOVA1 and activated by ESRP1 & 2. SET is a nuclear proto-oncogene that has been associated with Precursor T-Cell Acute Lymphoblastic Leukaemia and has been linked to pathways such as cell cycle. The GO annotations related to this gene included histone binding and protein phosphatase regulator activity.

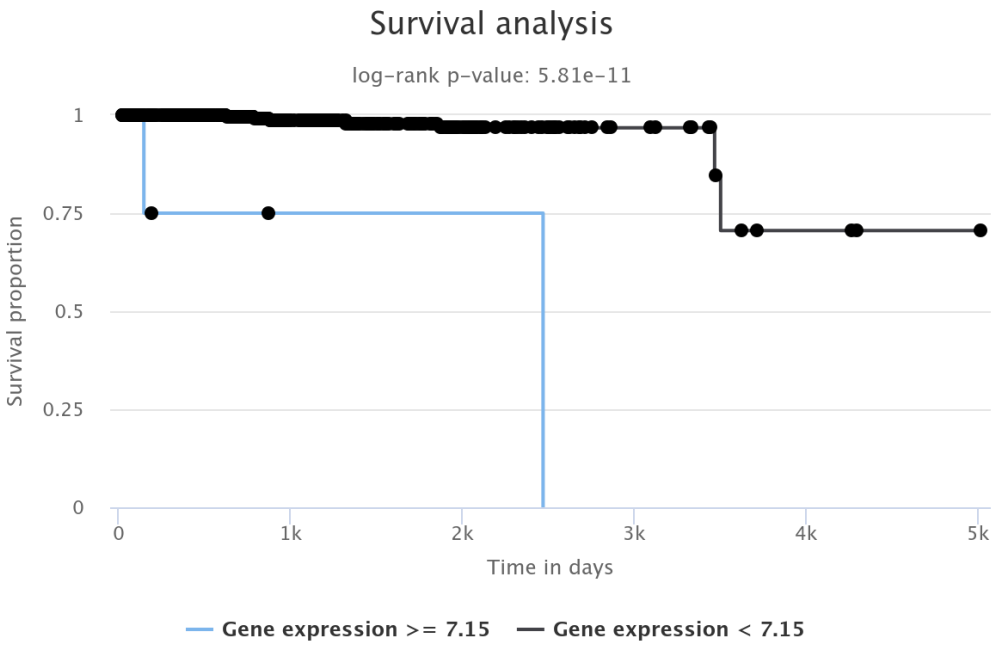


Figure 5.20: Survival analysis of patients with CBX1 splice isoform measured using Psichomics. *The survival curve shows gene expression levels over time (days) with a significant p-value. The sudden drop in blue line shows that higher expression of CBX1 gene leads to decrease in patient survival. Note only a small number of patients had high expression levels of CBX1.*

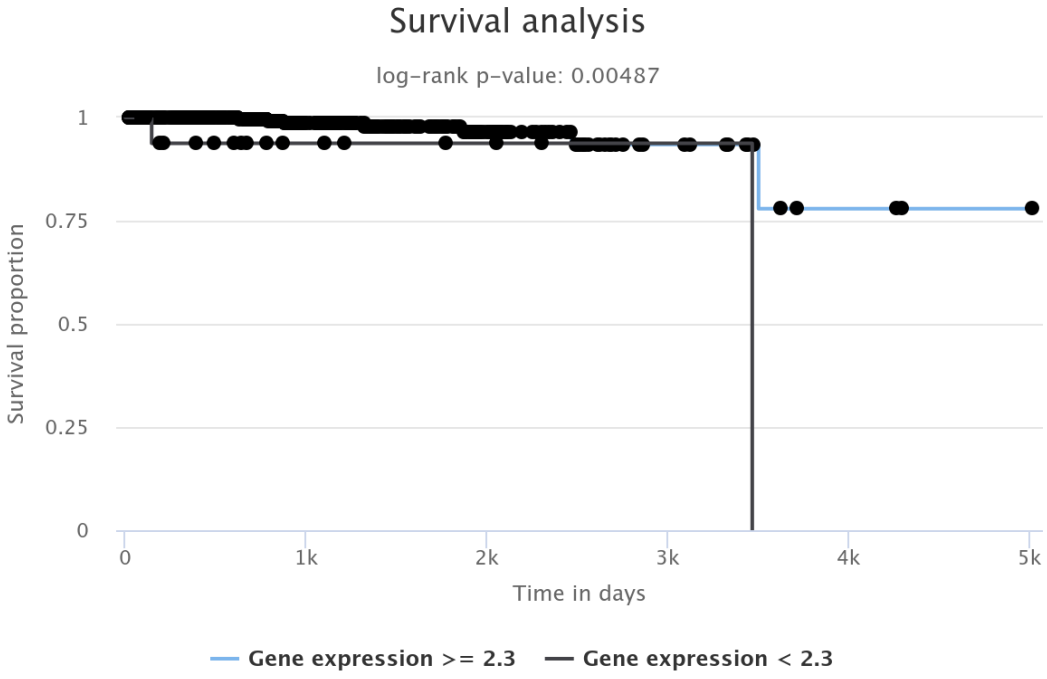


Figure 5.19: Survival analysis of patients with CFH splice isoform using Psichomics. *The survival curve shows a difference in survival for patient with different gene expression levels of CFH over time (days) with a significant p-value. The sudden drop in black line shows that lower expression of CFH gene leads to decrease in patient survival whereas, a higher CFH level correlates with better survival. Note only a small number of patients drop expression levels.*

5.4 Discussion

In this chapter I have analysed global consequences of NOVA1 expression on the prostate cancer transcriptome to identify the novel NOVA1 regulated targets. This analysis showed that there were changes in both differential gene expression and differential splicing. The changes found might be important in cancer progression.

Initially, differential gene expression analysis was done to find out the genes that were either up or downregulated. This showed that almost 16000 genes were either up or downregulated. This set of genes was used to perform Gene Set Enrichment Analysis (GSEA), which showed that NOVA1-regulated genes associated with many pathways. The genes upregulated after NOVA1 knockdown with a significant adjusted p-value associated with hallmark pathways such as E2F pathway that is linked to cell proliferation; G2M checkpoint that is related to cell cycle regulation; and EMT that is concerned with cells shifting from Epithelial to Mesenchymal morphology, and to acquire ability to move, thus enabling metastasis. The genes downregulated significantly after NOVA1 depletion were associated with pathways including inflammatory response and apoptosis. It would be interesting to see how these genes regulated by NOVA1 effect the mentioned pathways in functional assays, as this would suggest the specific impact of these pathways.

The *CBX1* gene was amongst the most strongly regulated by NOVA1. *CBX1* is very important and is involved in embryonic stem cell (ESC) biology, suppresses ESC differentiation and supports cell renewal by binding ESCs to chromatin regions of genes that promote differentiation and also maintains cell pluripotency. Well-differentiated cells have better prognosis and are not as aggressive. So, a prediction that as NOVA1 represses *CBX1* expression it stops cell differentiation and promotes aggressive cancer progression. *CBX1* was also studied in mice with a null mutation of murine in *CBX1* gene, which led to perinatal lethality.[194] *CBX1* plays a role in neurogenesis and brain development. Does it Link with cancer? CBX1 activates the Wnt/Beta-catenin pathway through interaction with HMGA2 (chromatin binding protein) in hepatocellular carcinoma.[194,200] Also a study suggests CBX1 expression and ability to regulate cell growth is controlled by miR-205-5p in pituitary

tumours.[194,201] Considering the role of CBX1 in hepatocellular carcinoma and miR-205-5p control of CBX1 in pituitary tumours, it is interesting to find CBX1 as a target of NOVA1 in a prostate cancer cell line. It would be interesting to understand how CBX1 contributes to prostate cancer biology, and if it is one of the NOVA1 controlled targets that can be used for a promotion of repression of aggressive prostate cancer. CBX proteins also control nuclear organization by regulating the formation of domains of SAHF (senescence-associated heterochromatin foci) which represent heterochromatin domains that are associated with irreversible cell cycle arrest that are selectively bound by CBX1 and CBX3 proteins during onset of cellular senescence.[194]

Another NOVA1-regulated gene of particular interest was *SPARC* (Secreted Protein Acidic and Cysteine Rich). *SPARC* has been correlated with increased metastasis and may promote cell invasion. *SPARC* expression is repressed in the presence of normal *NOVA1* gene expression levels and upregulated after *NOVA1* knockdown. This might suggest role of *NOVA1* expression in metastasis. This also correlates with the prior evidence that suggests ESRP1 and ESRP2 expression promotes EMT, and they are regulated inversely to NOVA1.

I also analysed NOVA1 regulated genes using gene ontology analysis. Downregulated genes on gene ontology analysis associated with important pathways such as regulation of chemotaxis, leukocyte migration, humoral response, inflammatory response, immune response and more whereas upregulated genes associated with blood vessel development, ion transport, organization of extra cellular matrix, T-cell differentiation and more. This further indicates a complex and diverse range of pathways that are affected by this single splicing regulator NOVA1. To functionally test the phenotypic effects of these pathways, a cell colony forming assay was performed on PC3 cell lines, comparing *NOVA1* knockdown and control cells. The results of this assay showed that the control samples had much higher colony formation compared to the *NOVA1* knockdown samples. This combined with the differential gene expression profile after *NOVA1* knockdown suggested that genes involved in biological pathways and their downstream effects affected cell division but need to be functionally evaluated further.

Additionally, SUPPA2 was used to analyse differential splicing patterns between control cells and after *NOVA1* depletion. SUPPA2 predicted approximately 9000 differentially regulated splicing events. When these were then filtered based on adjusted p-value this left only 404 significant changes. When these 404 events were compared to the known ESRP 1/2 splicing events, only 14 events were found to be common between the two datasets suggesting there are many novel splicing targets for NOVA1 rather than shared targets affected by NOVA1 and ESRP1/2 genes.

I also looked in more detail at each of these predicted splicing events on the IGV genome browser. When all 404 events were visualized using IGV genome browser, only 127 events were visually affected (and so have strong splicing changes). Among these stronger splicing events, only one event was also discovered in the identified events for ESRP1 and ESRP2, suggesting that NOVA1 works differently to the ESRP1 and ESRP2 genes and affecting different splicing targets. Among the identified splicing events some of the events such as *FGFR1OP2*, *RHBDD2*, *MYL6*, *ESYT2* and *GEN1* were some of the interesting events. One exon, *GEN1* exon 4, was seen to be a splicing target for NOVA1 but this exon was not annotated at the time of discovery. It has been since annotated. Exons within the *FGFR1OP2*, *RHBDD2*, and *GEN1* genes appeared to be activated by NOVA1 whereas *MYL6* splicing was seen to be repressed by NOVA1.

Chapter 6: Conclusion and Future Works

At the start of this thesis, it was known that androgens-controlled splicing in prostate cancer cells via the control of ESRP1 and ESRP2 expression levels.[1] However, it was unclear whether androgens solely relied on ESRP1 and ESRP2 for downstream splicing control or exerted splicing control via controlling other splicing regulator genes. Preliminary evidence suggested that *NOVA1* might be a potential gene through which androgens might modulate downstream splicing regulation control. Therefore, the main questions at the beginning of this thesis research project were about whether NOVA1 was indeed androgen regulated? Additionally, what prostate cell lines had the highest expression of NOVA1? Moreover, did NOVA1 have a regulation pattern that was opposite to that of ESRP1 and ESRP2, and different or the same target genes? Lastly, what are the global endogenous effects of NOVA1 on prostate cancer gene expression? In this thesis I focused on answering these questions.

6.1 NOVA1 is androgen regulated.

Initially, I set out to understand whether *NOVA1* was androgen regulated. To address this, I used the LNCaP cell line. The LNCaP cell line is an androgen sensitive prostate adenocarcinoma cell line with epithelial morphology. Five lines of evidence in this thesis support NOVA1 as an androgen-repressed gene. (1) LNCaP cells were grown in steroid deplete medium and then treated with androgen-containing medium. The LNCaP cells grown in steroid deplete medium had higher expression levels of NOVA1 protein compared to LNCaP cells that were later treated with androgen positive medium. This indicates that androgens repress NOVA1 expression. (2) Further evidence supporting this came from a time course experiment, where the steroid depleted and androgen positive LNCaP cells treated for 24 hours were compared. An RT-qPCR analysis showed that *NOVA1* RNA levels were higher in steroid deplete compared to androgen positive treated cells. (3) Androgen signaling activity can be blocked by drugs like casodex. To provide further evidence that NOVA1 is androgen-regulated, I quantified NOVA1 protein expression in steroid deplete and androgen positive LNCaP cells that had additional treatments with either Casodex or DMSO. This showed consistent higher levels of NOVA1 protein after Casodex treatment. (4) Androgen signaling is mediated by the androgen receptor. NOVA1 protein also had higher levels in androgen

receptor knockdown LNCaP cells compared to controls. (5) Moreover, looking at different prostate cancer cell lines, highest levels of NOVA1 protein were found in PC3 cell line, which is an androgen receptor negative cell line. Hence the androgen receptor is not available to repress the NOVA1 gene in PC3 cells. Taken together, these five lines of evidence confirmed that *NOVA1* was affected by androgen regulation, and that NOVA1 had an inverse expression pattern to the androgens compared to *ESRP2*.

Once I knew that *NOVA1* was repressed by androgens I needed to find out whether there were any androgen receptor binding sites close to the *NOVA1* gene. To achieve this, I initially used the publicly available ChIP data of the androgen receptor binding sites to identify their proximity to NOVA1.[2] The distance between androgen receptor binding sites and NOVA1 was more than expected which made me wonder whether NOVA1 was directly controlled by androgen receptor. Later, I came across a database called ReMap2022 that had updated androgen receptor binding site tracks.[3] I loaded the androgen receptor tracks and found that the androgen receptor binding sites overlapped with the *NOVA1* gene. This supported a model where *NOVA1* is androgen regulated by direct binding of the androgen receptor. Additionally, NOVA1 was downregulated in prostate adenocarcinoma compared to normal prostate tissue whereas, ESRP1 and ESRP2 were found upregulated in the Prostate adenocarcinoma cohort. This also confirmed that expression of NOVA1 was inversely regulated to expression of ESRP1 and ESRP2 within patients.

In summary, the data from both expression studies and AR binding site analysis support that NOVA1 is regulated by androgens. The inverse relationship between NOVA1 expression and androgen presence, alongside the AR binding sites near NOVA1, indicate that androgen signaling plays a critical role in the regulation of NOVA1. These findings also bring in to light the importance of NOVA1 as a splicing factor in androgen signaling pathways and suggest potential avenues for further investigation into its role in prostate cancer biology.

Since NOVA1 is highly expressed in PC3 cell lines I reasoned that the PC3 cell line would be the best model to carry out *NOVA1* gene knockdown and perform the downstream analysis. To further investigate the effect of androgen regulation of NOVA1, I examined known splicing targets of ESRP1 and ESRP2 discovered by Munkley et al to see how these behaved after

NOVA1 knockdown.[4] Interestingly, several targets were commonly repressed by both NOVA1 and ESRP, including alternative exons in *DOCK7*, *CEACAM1*, *FNIP1*, *KIF13A*, *MLPH*, and *ZNF207*. This suggested that there was a shared regulatory control between the NOVA1 and ESRP1/2 splicing targets. In addition, certain targets for ESRP1/2 such as exons within *APLP2*, *MINK1*, *NUMB*, *RALGPS2*, and *TUFT1* were activated after both depletion of ESRP1/2 and NOVA1, which additionally suggested a coordination of regulation between NOVA1 and ESRP1/2 splicing regulation. However, there was also a subset of targets that displayed inverse regulation between NOVA1 and ESRP1/2. These targets included exons within the *ACSF2*, *ARFGAP2*, *ARHGEF11*, *ARHGEF12*, *CTNND1*, *DNM1L*, *EIF4A2*, *ENAH*, *FLNB*, *FN1*, *GRHL1*, *INF2*, *ITGA6*, *MAP3K7*, *MBNL1*, *MYO1B*, *RAC1*, *RPS24*, *SLC37A2*, *SLK*, *TCIRG1*, *TRIP10*, *UAP1*, *VPS39*, and *WNK1* genes. These latter targets were upregulated by one factor while downregulated by the other factor. This inverse regulation suggested that NOVA1's effect on splicing might be more complicated than only regulating splicing targets also controlled by ESRP1/2. Moreover, there were some known ESRP1/2 targets that did not show any splicing changes in response to NOVA1 gene expression knockdown. This was an indication that NOVA1, although coordinating the regulation some does not have a universal control over all the ESRP1/2 target exons. This suggests that NOVA1 and ESRP1/2 have a likely independent functions while coordinating in splicing of some common targets.

In summary, despite the fact that NOVA1 is inversely regulated compared to ESRP1/2, it does not solely exert an opposite effect on all shared targets. Instead, NOVA1 appears to influence these targets in both shared and distinct ways, suggesting that it contributes to a broader regulatory network. Furthermore, the lack of NOVA1 influence on certain ESRP targets reveals additional regulatory layers, potentially contributing to the heterogeneity of splicing events in prostate cancer. This shared yet distinct regulatory control hints at a synergistic effect of NOVA1 and ESRP1/2 in prostate cancer biology, adding complexity to how androgen signaling may modulate downstream processes.

6.2 The effect of NOVA1 on global expression patterns.

The data in this thesis suggests that NOVA1 plays a crucial role in global gene expression and splicing patterns in prostate cancer. The analysis of NOVA1 knockdown cells revealed a

profound impact on differential gene expression and differential splicing. Specifically, NOVA1 knockdown caused differential expression of approximately 16000 genes, with around 4200 of these genes showing significant differential gene expression changes, either upregulated or downregulated. Gene set enrichment analysis and gene ontology revealed that many of the genes controlled by NOVA1 are involved in essential processes such as epithelial mesenchymal transition, cell cycle regulation, and cell proliferation. All these pathways have been implicated in cancer progression.

It was interesting to note that NOVA1 target genes within PC3 cells aligned with pathways such as G2M and E2F, which are the same pathways affected by ESRP1/2 genes shown in Advani et al.[5] This finding suggests that NOVA1 expression levels may help control cell morphology and promote the mesenchymal shift in the epithelial cells. This aligns with initial hypothesis that NOVA1 expression may facilitate transition from epithelial to mesenchymal state and ultimately metastasis. In addition to the pathways discussed, certain genes showed a very significant regulation in presence and absence of NOVA1. One such gene was *CBX1* that was highly and significantly upregulated after *NOVA1* gene expression knockdown. *CBX1* belongs to the Chromobox protein family associated with chromatin modification, and the control of *CBX1* expression by NOVA1 suggests that androgen regulation by NOVA1 might be contributing to the promotion of aggressive, metastatic prostate cancer in addition to the androgens controlling the ESRP1/2 genes. Furthermore, *SPARC* is a gene downregulated in NOVA1 knockdown, and linked with cell invasion and metastasis.

At the gene-splicing level, NOVA1 knockdown significantly impacted around 400 splicing events. Out of all these predicted significant splicing events, 127 were validated using the IGV browser and only one of these validated events was found to be a common splicing target between NOVA1 and ESRP1/2 suggesting the rest are novel targets identified for NOVA1. Moreover, survival analysis of splicing events targeted by NOVA1 using Psichomics revealed that some NOVA1 regulated splicing events correlated with decreased time to biochemical recurrence in prostate cancer patients. For example, *FGFR1OP2* exon 5 inclusion is promoted by NOVA1 which leads to a decreased time to biochemical recurrence. Similarly, *RHBDD2* exon 2 is another gene where NOVA1 promotes the inclusion and leads to a decreased time to biochemical recurrence. *FGFR1OP2* is a fibroblast growth factor that is associated with

wound healing where as *RHBDD2* belongs to rhomboid family of membrane bound proteases and overexpression of this gene family has been linked to breast cancer.

In summary, these findings position NOVA1 as a key factor in driving the aggressive behavior of prostate cancer. While ESRP1/2 plays an established role in maintaining the epithelial phenotype, NOVA1 expression may enable prostate cancer cells to transition into more lethal, mesenchymal states. NOVA1 may act synergistically with ESRP1/2 to promote disease progression at advanced stages when ESRP1/2 alone is insufficient to drive further malignancy.

6.3 Future research

In future experiments it would be useful to dissect splicing control by some of the exons directly regulated by NOVA1 and ESRP1/2 using minigene approaches, and to identify the sequences within the pre-mRNAs that are important for splicing control by either or both of these proteins, and how they interact. These experiments were planned during this thesis but I did not have time to do them. It would be exciting to test functional predictions from the splicing and gene expression analyses after NOVA1 gene expression depletion with functional properties of PC3 cells. It would be also useful to correlate *NOVA1* gene expression with some of the targets identified within this thesis in more metastatic prostate tumour samples.

Bibliography

1. Um, P. Cancer, Definition BT - Encyclopedia of Metagenomics: Environmental Metagenomics. in (eds. Highlander, S. K., Rodriguez-Valera, F. & White, B. A.) 65 (Springer US, Boston, MA, 2015). doi:10.1007/978-1-4899-7475-4_106.

2. Institute, N. C. NCI. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/cancer.

3. Abeshouse, B. S. Primary benign and malignant tumors of the ureter: A review of the literature and report of one benign and twelve malignant tumors. *The American Journal of Surgery* **91**, 237–271 (1956).

4. Hakim, D. N., Pelly, T., Kulendran, M. & Caris, J. A. Benign tumours of the bone: A review. *J Bone Oncol* **4**, 37–41 (2015).

5. Marino-Enriquez, A. & Fletcher, C. D. M. Shouldn't we care about the biology of benign tumours? *Nat Rev Cancer* **14**, 701–702 (2014).

6. UK, C. R. How cancers grow. https://www.cancerresearchuk.org/about-cancer/what-is-cancer/how-cancers-grow (2017).

7. Faguet, G. B. A brief history of cancer: Age-old milestones underlying our current knowledge database. *Int J Cancer* **136**, 2022–2036 (2015).

8. Sudhakar, A. History of Cancer, Ancient and Modern Treatment Methods. *J Cancer Sci Ther* **01**, i–iv (2009).

9. Olsen, J. Causes and Prevention. *Scand J Public Health* **19**, 1–6 (1991).

10. Hastings, K. G. *et al.* Socioeconomic Differences in the Epidemiologic Transition From Heart Disease to Cancer as the Leading Cause of Death in the United States, 2003 to 2015: An Observational Study. *Ann Intern Med* **169**, 836–844 (2018).

11. Dagenais, G. R. *et al.* Variations in common diseases, hospital admissions, and deaths in middle-aged adults in 21 countries from five continents (PURE): a prospective cohort study. *The Lancet* **6736**, 1–10 (2019).

12. International Agency for Research on Cancer. IARC. International Agency for Research on Cancer.

13. UK, B. C. Breast Cancer UK. https://www.breastcanceruk.org.uk/about-breast-cancer/facts-and-figures/?gclid=EAIaIQobChMIsbyT-52r5wIVB_lRCh1bIgPxEAAYASAAEgJe7vD_BwE.

14. UK, C. R. Cancer Research UK (Prostate). https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer/incidence.

15. Percy, C., Holten, V. van, Muir, C. S. & Organization, W. H. International classification of diseases for oncology. (1990).

16. Jack, A., Percy, C., Shanmugarathan, S. & Whelan, S. *International Classification of Diseases for Oncology: ICD-O*. (World Health Organization, 2000).

17. Institute, N. C. Cancer Staging. *National Cancer Institute, NIH* https://www.cancer.gov/about-cancer/diagnosis-staging/staging (2015).

18. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).

19. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).

20. Bradley, R. K. & Anczuków, O. RNA splicing dysregulation and the hallmarks of cancer. *Nature Reviews Cancer* vol. 23 Preprint at https://doi.org/10.1038/s41568-022-00541-7 (2023).

21. Tolkach, Y. & Kristiansen, G. The Heterogeneity of Prostate Cancer: A Practical Approach. *Pathobiology* **85**, 108–116 (2018).

22. Jin, K. Modern biological theories of aging. *Aging Dis* **1**, 72–74 (2010).

23. Heidi Chial. Tumor Suppressor (TS) Genes and the Two-Hit Hypothesis. *Nature Education* 1(1):177 https://www.nature.com/scitable/topicpage/tumor-suppressor-ts-genes-and-the-two-887/ (2008).

24. Knudson, A. G. Mutation and cancer: statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences* **68**, 820–823 (1971).

25. Boyland, E. Tumour initiators, promoters, and complete carcinogens. *Br J Ind Med* **42**, 716–718 (1985).

26. Romer, Alfred Sherwood; Parsons, T. S. *The Vertebrate Biology*. (Saunders, Philadelphia, 1977).

27. Aaron, L., Franco, O. & Hayward, S. W. Review of Prostate Anatomy and Embryology and the Etiology of BPH. *Urol Clin North Am* **43**, 279–288 (2016).

28. Aaron, L., Franco, O. & Hayward, S. W. Review of Prostate Anatomy and Embryology and the Etiology of BPH. *Urol Clin North Am* **43**, 279–288 (2016).

29. Lee, C. H., Akin-Olugbade, O. & Kirschenbaum, A. Overview of Prostate Anatomy, Histology, and Pathology. *Endocrinology and Metabolism Clinics of North America* vol. 40 Preprint at https://doi.org/10.1016/j.ecl.2011.05.012 (2011).

30. Verze, P., Cai, T. & Lorenzetti, S. The role of the prostate in male fertility, health and disease. *Nature Reviews Urology* vol. 13 Preprint at https://doi.org/10.1038/nrurol.2016.89 (2016).

31. McNeal, J. E. Anatomy of the prostate and morphogenesis of BPH. *Prog Clin Biol Res* **145**, 27–53 (1984).

32. Lowsley, O. S. The development of the human prostate gland with reference to the development of other structures at the neck of the urinary bladder. *American Journal of Anatomy* **13**, 299–349 (1912).

33. Mike Bath. The Prostate Gland. *https://teachmeanatomy.info/pelvis/the-male-reproductive-system/prostate-gland/* (2025).

34. Wang, G., Zhao, D., Spring, D. J. & Depinho, R. A. Genetics and biology of prostate cancer. *Genes Dev* **32**, 1105–1140 (2018).

35. Berquin, I. M., Min, Y., Wu, R., Wu, H. & Chen, Y. Q. Expression signature of the mouse prostate. *J Biol Chem* **280**, 36442–36451 (2005).

36. Advani, R. *et al.* Epithelial specific splicing regulator proteins as emerging oncogenes in aggressive prostate cancer. *Oncogene* (2023) doi:10.1038/s41388-023-02838-9.

37. American Cancer Society. Key Statistics for Prostate Cancer. *https://www.cancer.org/cancer/types/prostate-cancer/about/key-statistics.html*.

38. Torre, L. A. *et al.* Global cancer statistics, 2012. *CA Cancer J Clin* **65**, 87–108 (2015).

39. Globocan. Internation Agency for Research on Cancer (Prostate). *World Health Organization* https://gco.iarc.fr/today/data/factsheets/cancers/27-Prostate-fact-sheet.pdf (2018).

40. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, (2021).

41. Ferlay, J. *et al.* Cancer statistics for the year 2020: An overview. *Int J Cancer* **149**, (2021).

42. Cancer Research UK. Prostate cancer statistics. *https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer#heading-Zero*.

43. Davey, R. A. & Grossmann, M. Androgen Receptor Structure, Function and Biology: From Bench to Bedside. *Clin Biochem Rev* **37**, (2016).

44. Tietz, K. T. & Dehm, S. M. Androgen receptor variants: RNA-based mechanisms and therapeutic targets. *Human Molecular Genetics* vol. 29 Preprint at https://doi.org/10.1093/hmg/ddaa089 (2020).

45. MacLean, H. E., Warne, G. L. & Zajac, J. D. Localization of functional domains in the androgen receptor. *Journal of Steroid Biochemistry and Molecular Biology* vol. 62 Preprint at https://doi.org/10.1016/S0960-0760(97)00049-6 (1997).

46. Heinlein, C. A. & Chang, C. Androgen receptor (AR) coregulators: An overview. *Endocrine Reviews* vol. 23 Preprint at https://doi.org/10.1210/edrv.23.2.0460 (2002).

47. Munkley, J. *et al.* Glycosylation is an Androgen-Regulated Process Essential for Prostate Cancer Cell Viability. *EBioMedicine* **8**, 103–116 (2016).

48. Munkley, J. *et al.* Androgen-regulated transcription of ESRP2 drives alternative splicing patterns in prostate cancer. *Elife* **8**, (2019).

49. Becerra, M. F., Atluri, V. S., Bhattu, A. S. & Punnen, S. Serum and urine biomarkers for detecting clinically significant prostate cancer. *Urologic Oncology: Seminars and Original Investigations* **39**, 686–690 (2021).

50. Salagierski, M. *et al.* Differential expression of PCA3 and its overlapping PRUNE2 transcript in prostate cancer. *Prostate* **70**, (2010).

51. McKiernan, J. *et al.* A Prospective Adaptive Utility Trial to Validate Performance of a Novel Urine Exosome Gene Expression Assay to Predict High-grade Prostate Cancer in Patients with Prostate-specific Antigen 2–10 ng/ml at Initial Biopsy. *Eur Urol* **74**, (2018).

52. Leyten, G. H. J. M. *et al.* Identification of a candidate gene panel for the early diagnosis of prostate cancer. *Clinical Cancer Research* **21**, (2015).

53. Van Neste, L. *et al.* Detection of High-grade Prostate Cancer Using a Urinary Molecular Biomarker–Based Risk Score. *Eur Urol* **70**, (2016).

54. Kretschmer, A. & Tilki, D. Biomarkers in prostate cancer – Current clinical utility and future perspectives. *Critical Reviews in Oncology/Hematology* vol. 120 Preprint at https://doi.org/10.1016/j.critrevonc.2017.11.007 (2017).

55. Matuszczak, M., Schalken, J. A. & Salagierski, M. Prostate cancer liquid biopsy biomarkers' clinical utility in diagnosis and prognosis. *Cancers* vol. 13 Preprint at https://doi.org/10.3390/cancers13133373 (2021).

56. Tomlins, S. A. *et al.* Urine TMPRSS2:ERG Plus PCA3 for Individualized Prostate Cancer Risk Assessment. *Eur Urol* **70**, (2016).

57. Punnen, S., Pavan, N. & Parekh, D. J. Finding the Wolf in Sheep's Clothing: The 4Kscore Is a Novel Blood Test That Can Accurately Identify the Risk of Aggressive Prostate Cancer. *Rev Urol* **17**, (2015).

58. Catalona, W. J. *et al.* A multicenter study of [-2]pro-prostate specific antigen combined with prostate specific antigen and free prostate specific antigen for prostate cancer detection in the 2.0 to 10.0 ng/ml prostate specific antigen range. *Journal of Urology* **185**, (2011).

59. Uhr, A., Glick, L. & Gomella, L. G. An overview of biomarkers in the diagnosis and management of prostate cancer. *The Canadian journal of urology* vol. 27 Preprint at (2020).

60. Wojno, K. J. *et al.* Reduced rate of repeated prostate biopsies observed in ConfirmMDx clinical utility field study. *Am Health Drug Benefits* **7**, (2014).

61. Trock, B. J. *et al.* Evaluation of GSTP1 and APC methylation as indicators for repeat biopsy in a high-risk cohort of men with negative initial prostate biopsies. *BJU Int* **110**, (2012).

62. Klein, E. A. *et al.* A 17-gene assay to predict prostate cancer aggressiveness in the context of gleason grade heterogeneity, tumor multifocality, and biopsy undersampling. *Eur Urol* **66**, (2014).

63. Kim, H. L. *et al.* Validation of the Decipher Test for predicting adverse pathology in candidates for prostate cancer active surveillance. *Prostate Cancer Prostatic Dis* **22**, (2019).

64. Marascio, J. *et al.* Prospective study to define the clinical utility and benefit of Decipher testing in men following prostatectomy. *Prostate Cancer Prostatic Dis* **23**, (2020).

65. Erho, N. *et al.* Discovery and Validation of a Prostate Cancer Genomic Classifier that Predicts Early Metastasis Following Radical Prostatectomy. *PLoS One* **8**, (2013).

66. Prolaris Cell Cycle Progression Test for Localized Prostate Cancer: A Health Technology Assessment. *Ontario health technology assessment series* vol. 17 Preprint at (2017).

67. Sarwar, M. *et al.* Targeted suppression of AR-V7 using PIP5K1α inhibitor overcomes enzalutamide resistance in prostate cancer cells. *Oncotarget* **7**, (2016).

68. Scher, H. I. *et al.* Association of AR-V7 on circulating tumor cells as a treatment-specific biomarker with outcomes and survival in castration-resistant prostate cancer. *JAMA Oncol* **2**, (2016).

69. Streicher, J., Meyerson, B. L., Karivedu, V. & Sidana, A. A review of optimal prostate biopsy: indications and techniques. *Therapeutic Advances in Urology* vol. 11 Preprint at https://doi.org/10.1177/1756287219870074 (2019).

70. O'Shea, A. & Harisinghani, M. PI-RADS: multiparametric MRI in prostate cancer. *Magnetic Resonance Materials in Physics, Biology and Medicine* vol. 35 Preprint at https://doi.org/10.1007/s10334-022-01019-1 (2022).

71. Patel, U. TRUS and prostate biopsy: Current status. *Prostate Cancer and Prostatic Diseases* vol. 7 Preprint at https://doi.org/10.1038/sj.pcan.4500728 (2004).

72. Weiss, C. R., Nour, S. G. & Lewin, J. S. MR-guided biopsy: A review of current techniques and applications. *Journal of Magnetic Resonance Imaging* vol. 27 Preprint at https://doi.org/10.1002/jmri.21270 (2008).

73. Thomson, A., Li, M., Grummet, J. & Sengupta, S. Transperineal prostate biopsy: A review of technique. *Translational Andrology and Urology* vol. 9 Preprint at https://doi.org/10.21037/tau.2019.12.40 (2021).

74. Chen, R. C. *et al.* Active surveillance for the management of localized prostate cancer (Cancer Care Ontario guideline): American society of clinical oncology clinical practice guideline endorsement. *Journal of Clinical Oncology* **34**, (2016).

75. Nguyen-Nielsen, M. & Borre, M. Diagnostic and Therapeutic Strategies for Prostate Cancer. *Seminars in Nuclear Medicine* vol. 46 Preprint at https://doi.org/10.1053/j.semnuclmed.2016.07.002 (2016).

76. American Cancer Society. Radiation Therapy for Prostate Cancer. *https://www.cancer.org/cancer/types/prostate-cancer/treating/radiation-therapy.html*.

77. American Cancer Society. Hormone Therapy for Prostate Cancer. *https://www.cancer.org/cancer/types/prostate-cancer/treating/hormone-therapy.html*.

78. American Cancer Society. Chemotherapy for Prostate Cancer. *https://www.cancer.org/cancer/types/prostate-cancer/treating/chemotherapy.html*.

79. American Cancer Society. Immunotherapy for Prostate Cancer. *https://www.cancer.org/cancer/types/prostate-cancer/treating/vaccine-treatment.html*.

80. Rui, R., Zhou, L. & He, S. Cancer immunotherapies: advances and bottlenecks. *Frontiers in Immunology* vol. 14 Preprint at https://doi.org/10.3389/fimmu.2023.1212476 (2023).

81. American Cancer Society. Targeted Drug Therapy for Prostate Cancer. *https://www.cancer.org/cancer/types/prostate-cancer/treating/targeted-therapy.html*.

82. Marasco, L. E. & Kornblihtt, A. R. The physiology of alternative splicing. *Nature Reviews Molecular Cell Biology* vol. 24 Preprint at https://doi.org/10.1038/s41580-022-00545-z (2023).

83. Jakubauskienė, E. & Kanopka, A. Alternative splicing and hypoxia puzzle in alzheimer's and parkinson's diseases. *Genes* vol. 12 Preprint at https://doi.org/10.3390/genes12081272 (2021).

84. Wilkinson, M. E., Charenton, C. & Nagai, K. RNA Splicing by the Spliceosome. *Annual Review of Biochemistry* vol. 89 Preprint at https://doi.org/10.1146/annurev-biochem-091719-064225 (2020).

85. Gao, Y. *et al.* Systematic characterization of short intronic splicing-regulatory elements in SMN2 pre-mRNA. *Nucleic Acids Res* **50**, (2022).

86. Sironi, M. *et al.* Silencer elements as possible inhibitors of pseudoexon splicing. *Nucleic Acids Res* **32**, (2004).

87. Sun, H. & Chasin, L. A. Multiple Splicing Defects in an Intronic False Exon. *Mol Cell Biol* **20**, (2000).

88. Lee, Y. & Rio, D. C. Mechanisms and regulation of alternative Pre-mRNA splicing. *Annual Review of Biochemistry* vol. 84 Preprint at https://doi.org/10.1146/annurev-biochem-060614-034316 (2015).

89. Mao, M. *et al.* Modeling and Predicting the Activities of Trans-Acting Splicing Factors with Machine Learning. *Cell Syst* **7**, (2018).

90. Baralle, M. & Baralle, F. E. The splicing code. *Biosystems* **164**, 39–48 (2018).

91. Sperling, J., Azubel, M. & Sperling, R. Structure and Function of the Pre-mRNA Splicing Machine. *Structure* vol. 16 Preprint at https://doi.org/10.1016/j.str.2008.08.011 (2008).

92. Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R. & Misteli, T. Epigenetics in alternative pre-mRNA splicing. *Cell* vol. 144 Preprint at https://doi.org/10.1016/j.cell.2010.11.056 (2011).

93. Kumar, S. & Mohapatra, T. Deciphering Epitranscriptome: Modification of mRNA Bases Provides a New Perspective for Post-transcriptional Regulation of Gene Expression. *Frontiers in Cell and Developmental Biology* vol. 9 Preprint at https://doi.org/10.3389/fcell.2021.628415 (2021).

94. Wright, C. J., Smith, C. W. J. & Jiggins, C. D. Alternative splicing as a source of phenotypic diversity. *Nat Rev Genet* **23**, (2022).

95. Arning, L. *et al.* Handedness and the X chromosome: The role of androgen receptor CAG-repeat length. *Sci Rep* **5**, (2015).

96. Cao, S., Zhan, Y. & Dong, Y. Emerging data on androgen receptor splice variants in prostate cancer. *Endocrine-Related Cancer* vol. 23 Preprint at https://doi.org/10.1530/ERC-16-0298 (2016).

97. Monks, D. A. *et al.* Androgen receptor and Kennedy disease/spinal bulbar muscular atrophy. *Horm Behav* **53**, (2008).

98. Vega, S. *et al.* Snail blocks the cell cycle and confers resistance to cell death. *Genes Dev* **18**, 1131–1143 (2004).

99. Tsai, J. H., Donaher, J. L., Murphy, D. A., Chau, S. & Yang, J. Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer Cell* **22**, 725–736 (2012).

100. Chao, Y., Wu, Q., Acquafondata, M., Dhir, R. & Wells, A. Partial mesenchymal to epithelial reverting transition in breast and prostate cancer metastases. *Cancer Microenvironment* **5**, (2012).

101. Wells, A., Chao, Y. L., Grahovac, J., Wu, Q. & Lauffenburger, D. A. Epithelial and mesenchymal phenotypic switchings modulate cell motility in metastasis. *Frontiers in Bioscience* **16**, (2011).

102. Condeelis, J. S., Wyckoff, J. & Segall, J. E. Imaging of cancer invasion and metastasis using green fluorescent protein. *Eur J Cancer* **36**, (2000).

103. Chao, Y. L., Shepard, C. R. & Wells, A. Breast carcinoma cells re-express E-cadherin during mesenchymal to epithelial reverting transition. *Mol Cancer* **9**, (2010).

104. Yates, C. C., Shepard, C. R., Stolz, D. B. & Wells, A. Co-culturing human prostate carcinoma cells with hepatocytes leads to increased expression of E-cadherin. *Br J Cancer* **96**, (2007).

105. Tarin, D. The fallacy of epithelial mesenchymal transition in neoplasia. *Cancer Research* vol. 65 Preprint at https://doi.org/10.1158/0008-5472.CAN-05-0699 (2005).

106. Wells, A., Yates, C. & Shepard, C. R. E-cadherin as an indicator of mesenchymal to epithelial reverting transitions during the metastatic seeding of disseminated carcinomas. *Clinical and Experimental Metastasis* vol. 25 Preprint at https://doi.org/10.1007/s10585-008-9167-1 (2008).

107. Their, J. P. Epithelial-mesenchymal transitions in tumor progression. *Nature Reviews Cancer* vol. 2 Preprint at https://doi.org/10.1038/nrc822 (2002).

108. Hugo, H. *et al.* Epithelial - Mesenchymal and mesenchymal - Epithelial transitions in carcinoma progression. *Journal of Cellular Physiology* vol. 213 Preprint at https://doi.org/10.1002/jcp.21223 (2007).

109. Chaffer, C. L. *et al.* Mesenchymal-to-epithelial transition facilitates bladder cancer metastasis: Role of fibroblast growth factor receptor-2. *Cancer Res* **66**, (2006).

110. Turner, M. & DÍaz-Muñoz, M. D. RNA-binding proteins control gene expression and cell fate in the immune system review-article. *Nature Immunology* vol. 19 Preprint at https://doi.org/10.1038/s41590-017-0028-4 (2018).

111. Van Nostrand, E. L. *et al.* A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, (2020).

112. Corley, M., Burns, M. C. & Yeo, G. W. How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. *Molecular Cell* vol. 78 Preprint at https://doi.org/10.1016/j.molcel.2020.03.011 (2020).

113. Shah, K. *et al.* Androgen receptor signaling regulates the transcriptome of prostate cancer cells by modulating global alternative splicing. *Oncogene* **39**, 6172–6189 (2020).

114. Bebee, T. W. *et al.* The splicing regulators Esrp1 and Esrp2 direct an epithelial splicing program essential for mammalian development. *Elife* **4**, (2015).

115. Mallinjoud, P. *et al.* Endothelial, epithelial, and fibroblast cells exhibit specific splicing programs independently of their tissue of origin. *Genome Res* **24**, 511–521 (2014).

116. Yang, Y. *et al.* Determination of a Comprehensive Alternative Splicing Regulatory Network and Combinatorial Regulation by Key Factors during the Epithelial-to-Mesenchymal Transition. *Mol Cell Biol* **36**, 1704–1719 (2016).

117. Xin, Y. *et al.* Neuro-oncological ventral antigen 1 (NOVA1): Implications in neurological diseases and cancers. *Cell Prolif* **50**, (2017).

118. Buckanovich, R. J., Posner, J. B. & Darnell, R. B. Nova, the paraneoplastic Ri antigen, is homologous to an RNA-binding protein and is specifically expressed in the developing motor system. *Neuron* **11**, 657–672 (1993).

119. Darnell, R. B. & Posner, J. B. Paraneoplastic syndromes involving the nervous system. *N Engl J Med* **349**, 1543–1554 (2003).

120. Buckanovich, R. J. & Darnell, R. B. The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo. *Mol Cell Biol* **17**, 3194–3201 (1997).

121. Jensen, K. B. *et al.* Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron* **25**, 359–371 (2000).

122. Lewis, H. A. *et al.* Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. *Cell* **100**, 323–332 (2000).

123. Ule, J. *et al.* CLIP Identifies Nova-Regulated RNA Networks in the Brain. *Science (1979)* **302**, (2003).

124. Saito, Y. *et al.* NOVA2-mediated RNA regulation is required for axonal pathfinding during development. *Elife* **5**, (2016).

125. Yang, Y. Y. L., Yin, G. L. & Darnell, R. B. The neuronal RNA-binding protein Nova-2 is implicated as the autoantigen targeted in POMA patients with dementia. *Proc Natl Acad Sci U S A* **95**, (1998).

126. Graus, F., Rowe, G., Fueyo, J., Darnell, R. B. & Dalmau, J. The neuronal nuclear antigen recognized by the human anti-Ri autoantibody is expressed in central but not peripheral nervous system neurons. *Neurosci Lett* **150**, (1993).

127. Barash, Y. *et al.* Deciphering the splicing code. *Nature* **465**, (2010).

128. Buckanovich, R. J., Yang, Y. Y. L. & Darnell, R. B. The onconeural antigen Nova-1 is a neuron-specific RNA-binding protein, the activity of which is inhibited by paraneoplastic antibodies. *Journal of Neuroscience* **16**, (1996).

129. Villate, O. *et al.* Nova1 is a master regulator of alternative splicing in pancreatic beta cells. *Nucleic Acids Res* **42**, (2014).

130. Li, H. *et al.* Dynamic Expression Pattern of Neuro-oncological Ventral Antigen 1 (Nova1) in the Rat Brain after Focal Cerebral Ischemia/Reperfusion Insults. *Journal of Histochemistry and Cytochemistry* **61**, (2013).

131. Hervé, M. & Ibrahim, E. C. MicroRNA screening identifies a link between NOVA1 expression and a low level of IKAP in familial dysautonomia. *DMM Disease Models and Mechanisms* **9**, (2016).

132. Gimenez, M. *et al.* Quantitative proteomic analysis shows differentially expressed HSPB1 in glioblastoma as a discriminating short from long survival factor and NOVA1 as a differentiation factor between low-grade astrocytoma and oligodendroglioma. *BMC Cancer* **15**, 481 (2015).

133. Zhang, Y.-A. *et al.* High expression of neuro-oncological ventral antigen 1 correlates with poor prognosis in hepatocellular carcinoma. *PLoS One* **9**, e90955 (2014).

134. Zhang, Y.-A. *et al.* RNA binding protein Nova1 promotes tumor growth in vivo and its potential mechanism as an oncogene may due to its interaction with GABAA Receptor-gamma2. *J Biomed Sci* **23**, 71 (2016).

135. Kim, E. K. *et al.* Upregulated Neuro-oncological Ventral Antigen 1 (NOVA1) Expression Is Specific to Mature and Immature T- and NK-Cell Lymphomas. *J Pathol Transl Med* **50**, 104–112 (2016).

136. Yoon, S. O. *et al.* NOVA1 inhibition by miR-146b-5p in the remnant tissue microenvironment defines occult residual disease after gastric cancer removal. *Oncotarget* **7**, 2475–2495 (2016).

137. Kim, E. K. *et al.* Implications of NOVA1 suppression within the microenvironment of gastric cancer: association with immune cell dysregulation. *Gastric Cancer* **20**, 438–447 (2017).

138. Zhang, D. *et al.* Intron retention is a hallmark and spliceosome represents a therapeutic vulnerability in aggressive prostate cancer. *Nat Commun* **11**, 2089 (2020).

139. Tang, Z., Kang, B., Li, C., Chen, T. & Zhang, Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res* **47**, W556–W560 (2019).

140. Hammal, F., De Langen, P., Bergon, A., Lopez, F. & Ballester, B. ReMap 2022: A database of Human, Mouse, Drosophila and Arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments. *Nucleic Acids Res* **50**, (2022).

141. Wu, D. C., Yao, J., Ho, K. S., Lambowitz, A. M. & Wilke, C. O. Limitations of alignment-free tools in total RNA-seq quantification. *BMC Genomics* **19**, (2018).

142. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, (2014).

143. Trincado, J. L. *et al.* SUPPA2: Fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol* **19**, (2018).

144. Bonnot, T., Gillard, M. & Nagel, D. A Simple Protocol for Informative Visualization of Enriched Gene Ontology Terms. *Bio Protoc* **9**, (2019).

145. Saraiva-Agostinho, N. & Barbosa-Morais, N. L. Psichomics: Graphical Application for Alternative Splicing Quantification and Analysis. *Nucleic Acids Res* **47**, e7 (2019).

146. Saraiva-Agostinho, N. & Barbosa-Morais, N. L. Psichomics: Graphical Application for Alternative Splicing Quantification and Analysis. *Nucleic Acids Res* **47**, e7 (2019).

147. Evans, A. J. Treatment effects in prostate cancer. *Modern Pathology* **31**, (2018).

148. Gamat, M. & McNeel, D. G. Androgen deprivation and immunotherapy for the treatment of prostate cancer. *Endocrine-Related Cancer* vol. 24 Preprint at https://doi.org/10.1530/ERC-17-0145 (2017).

149. Shen, M. M. & Abate-Shen, C. Molecular genetics of prostate cancer: new prospects for old challenges. *Genes Dev* **24**, 1967–2000 (2010).

150. Wang, G., Zhao, D., Spring, D. J. & Depinho, R. A. Genetics and biology of prostate cancer. *Genes Dev* **32**, 1105–1140 (2018).

151. Miller, D. R. *et al.* Anti-Androgen Abiraterone Acetate Improves the Therapeutic Efficacy of Statins on Castration-Resistant Prostate Cancer Cells. *J Oncol Res Ther* **3**, (2017).

152. Dagvadorj, A. *et al.* Androgen-regulated and highly tumorigenic human prostate cancer cell line established from a transplantable primary CWR22 tumor. *Clin Cancer Res* **14**, 6062–6072 (2008).

153. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580–585 (2013).

154. Consortium, T. Gte. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science (1979)* **369**, 1318–1330 (2020).

155. Tomczak, K., Czerwińska, P. & Wiznerowicz, M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)* **19**, A68-77 (2015).

156. Nassar, L. R. *et al.* The UCSC Genome Browser database: 2023 update. *Nucleic Acids Res* **51**, D1188–D1195 (2023).

157. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676–682 (2012).

158. GraphPad. Welcome to the Prism 10 User Guide. *PRISM 10* https://www.graphpad.com/guides/prism/latest/user-guide/index.htm.

159. Swift, M. L. GraphPad Prism, Data Analysis, and Scientific Graphing. *J Chem Inf Comput Sci* **37**, 411–412 (1997).

160. Zhang, D. *et al.* Intron retention is a hallmark and spliceosome represents a therapeutic vulnerability in aggressive prostate cancer. *Nat Commun* **11**, 2089 (2020).

161. Chèneby, J. *et al.* ReMap 2020: a database of regulatory regions from an integrative analysis of Human and Arabidopsis DNA-binding sequencing experiments. *Nucleic Acids Res* **48**, D180–D188 (2020).

162. Munkley, J. *et al.* Androgen-regulated transcription of ESRP2 drives alternative splicing patterns in prostate cancer. *Elife* **8**, (2019).

163. Lu, Z. *et al.* Transcriptome-wide landscape of pre-mRNA alternative splicing associated with metastatic colonization. *Mol Cancer Res* **13**, 305–318 (2015).

164. Laganà, A. *et al.* Computational design of artificial RNA molecules for gene regulation. *Methods Mol Biol* **1269**, (2015).

165. Chong, Z. X., Yeap, S. K. & Ho, W. Y. Transfection types, methods and strategies: A technical review. *PeerJ* vol. 9 Preprint at https://doi.org/10.7717/peerj.11165 (2021).

166. Nakanishi, K. Anatomy of RISC: how do small RNAs and chaperones activate Argonaute proteins? *Wiley Interdisciplinary Reviews: RNA* vol. 7 Preprint at https://doi.org/10.1002/wrna.1356 (2016).

167. Mack, G. S. MicroRNA gets down to business. *Nature Biotechnology* vol. 25 Preprint at https://doi.org/10.1038/nbt0607-631 (2007).

168. Watson, J. D. & Crick, F. H. C. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature* **171**, (1953).

169. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, (1977).

170. Barba, M., Czosnek, H. & Hadidi, A. Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses* **6**, (2013).

171. Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. *Nature Reviews Genetics* vol. 20 Preprint at https://doi.org/10.1038/s41576-019-0150-2 (2019).

172. Andrews, S. FastQC: a quality control tool for high throughput sequence data. 2010. Preprint at (2017).

173. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907–915 (2019).

174. Abascal, F. *et al.* Loose ends: Almost one in five human genes still have unresolved coding status. *Nucleic Acids Res* **46**, (2018).

175. Abdellah, Z. *et al.* Finishing the euchromatic sequence of the human genome. *Nature* **431**, (2004).

176. Clamp, M. *et al.* Distinguishing protein-coding and noncoding genes in the human genome. *Proc Natl Acad Sci U S A* **104**, (2007).

177. Hatje, K., Mühlhausen, S., Simm, D. & Kollmar, M. The Protein-Coding Human Genome: Annotating High-Hanging Fruits. *BioEssays* **41**, (2019).

178. Piovesan, A. *et al.* Human protein-coding genes and gene feature statistics in 2019. *BMC Res Notes* **12**, (2019).

179. Ponomarenko, E. A. *et al.* The Size of the Human Proteome: The Width and Depth. *International Journal of Analytical Chemistry* vol. 2016 Preprint at https://doi.org/10.1155/2016/7436849 (2016).

180. Kim, M.-S. *et al.* A draft map of the human proteome. *Nature* **509**, 575–581 (2014).

181. Berget, S. M., Moore, C. & Sharp, P. A. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA*. *Proceedings of the National Academy of Sciences* **74**, 3171–3175 (1977).

182. Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* **12**, 1–8 (1977).

183. Marasco, L. E. & Kornblihtt, A. R. The physiology of alternative splicing. *Nat Rev Mol Cell Biol* **24**, 242–254 (2023).

184. Khan, M. R., Wellinger, R. J. & Laurent, B. Exploring the Alternative Splicing of Long Noncoding RNAs. *Trends in Genetics* **37**, 695–698 (2021).

185. Rogers, S. O. Integrated evolution of ribosomal RNAs, introns, and intron nurseries. *Genetica* **147**, 103–119 (2019).

186. Sekulovski, S. *et al.* Assembly defects of human tRNA splicing endonuclease contribute to impaired pre-tRNA processing in pontocerebellar hypoplasia. *Nat Commun* **12**, 5610 (2021).

187. Liu, Q., Fang, L. & Wu, C. Alternative Splicing and Isoforms: From Mechanisms to Diseases. *Genes (Basel)* **13**, (2022).

188. Shen, S. *et al.* rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S A* **111**, (2014).

189. Vaquero-Garcia, J. *et al.* A new view of transcriptome complexity and regulation through the lens of local splicing variations. *Elife* **5**, (2016).

190. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res* **22**, (2012).

191. Trincado, J. L. *et al.* SUPPA2: Fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol* **19**, (2018).

192. Blake, J. A. *et al.* Gene ontology consortium: Going forward. *Nucleic Acids Res* **43**, (2015).

193. Aleksander, S. A. *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224**, (2023).

194. van Wijnen, A. J. *et al.* Biological functions of chromobox (CBX) proteins in stem cell self-renewal, lineage-commitment, cancer and development. *Bone* **143**, (2021).

195. Maehara, N. *et al.* Circulating AIM prevents hepatocellular carcinoma through complement activation. *Cell Rep* **9**, (2014).

196. Fleury, C. *et al.* Identification of a Haemophilus influenzae Factor H–Binding Lipoprotein Involved in Serum Resistance . *The Journal of Immunology* **192**, (2014).

197. Xue, X. *et al.* Regulator-dependent mechanisms of C3b processing by factor i allow differentiation of immune responses. *Nat Struct Mol Biol* **24**, (2017).

198. Advani, R. *et al.* Epithelial specific splicing regulator proteins as emerging oncogenes in aggressive prostate cancer. *Oncogene* (2023) doi:10.1038/s41388-023-02838-9.

199. Munkley, J. *et al.* Androgen-regulated transcription of ESRP2 drives alternative splicing patterns in prostate cancer. *Elife* **8**, (2019).

200. Yang, Y. F., Pan, Y. H., Tian, Q. H., Wu, D. C. & Su, S. G. CBX1 Indicates Poor Outcomes and Exerts Oncogenic Activity in Hepatocellular Carcinoma. *Transl Oncol* **11**, (2018).

201. Hu, A., Zhang, Y., Zhao, X., Li, J. & Ying, Y. CBX1 is a direct target of miR-205-5p and contributes to the progression of pituitary tumor. *Pharmazie* **74**, (2019).

202. Shi, Y. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nature Reviews Molecular Cell Biology* vol. 18 Preprint at https://doi.org/10.1038/nrm.2017.86 (2017).

203. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform* **14**, (2013).

204. Nassar, L. R. *et al.* The UCSC Genome Browser database: 2023 update. *Nucleic Acids Res* **51**, D1188–D1195 (2023).

Appendix A: Primer Sequences

Gene Name	Forward Primer Sequence	Reverse Primer Sequence
TRPT1	CATGCGGTCCCATTGTGAAA	GCTTGGGGCTACTCTGACAC
FN1	AACCCCTACAAACGGCCAG	AGGAGAAATTGACAACTCTGCA
FGFR1OP2	AGAACATCAGTCGGCCTTGG	TCTTGTTCCTTGCAACCCTGT
BMI1	TGTTCGTTACCTGGAGACCA	TGTCTTTGTTTACTTTCCGATCCA
ESYT2	ACTTGAGACTGGAGTGGCTC	CAGGATTTGGGTTGCTGCTT
GEN1	GGAGTCCCTGGAGTTGGAAA	ACGGAACAATGAGCCAGTTT
MYL6	GTGAGAAGATGACAGAGGAAGAAGT	CAGCCATTCAGCACCATGC
RHBDD2	ATCCGCCACCTTCTTCACTG	CATCAGGCTGAAGGGGAAGG
DOCK7	GTTTGGGAGGATCAGTGCAT	AAGCGTCCCGTTAATGTTTG
FLNB	GGCGAAGAAGTAGGCTTTGTGG	GCCGTTCATGTCACTCACTGG

Appendix B: List of NOVA1 IGV validated splicing targets in prostate cancer

S. No.	Gene	Event
1.	MYL6	ENSG00000092841;SE:12:56160320-56160626:56160670-56161387:+
2.	NPIPA9	ENSG00000233024;SE:16:18358588-18359382:18359451-18359948:-
3.	EDEM2	ENSG00000088298;SE:20:35145018-35146825:35146935-35147152:-
4.	ESYT2	ENSG00000117868;SE:7:158749723-158752781:158752843-158759486:-
5.	TM7SF3	ENSG00000064115;SE:12:26976357-26979784:26979936-26980566:-
6.	TM7SF3	ENSG00000064115;SE:12:26975658-26976260:26976357-26980566:-
7.	RHBDD2	ENSG00000005486;SE:7:75879260-75881365:75881486-75881829:+
8.	LDHB	ENSG00000111716;SE:12:21654677-21657751:21657835-21737697:-
9.	PLOD2	ENSG00000152952;SE:3:146076895-146077862:146077924-146079116:-
10.	NDUFB5	ENSG00000136521;SE:3:179604939-179607758:179607812-179616983:+
11.	NXPE3	ENSG00000144815;SE:3:101779324-101779734:101779900-101782110:+
12.	NXPE3	ENSG00000144815;SE:3:101779324-101779734:101779934-101782110:+
13.	FGFR1OP2	ENSG00000111790;SE:12:26957743-26960515:26960628-26963342:+
14.	PSAP	ENSG00000197746;SE:10:71822007-71823888:71823896-71825837:-
15.	HBS1L	ENSG00000112339;SE:6:135042126-135050582:135050647-135054649:-
16.	CDC37	ENSG00000105401;RI:19:10395228:10395343-10395435:10395696:-
17.	SEPTIN2	ENSG00000168385;MX:2:241315982-241317499:241317599-241324216:241315982-241320203:241320287-241324216:+
18.	MYL6	ENSG00000092841;MX:12:56158711-56159587:56159622-56159975:56158711-56159623:56159730-56159975:+
19.	TMBIM6	ENSG00000139644;MX:12:49741611-49743417:49743565-49752464:49741611-49744413:49744542-49752464:+

20.	HINT1	ENSG00000169567;MX:5:131159611-131160656:131160844-131162572:131159611-131162029:131162477-131162572:-
21.	HINT1	ENSG00000169567;MX:5:131159611-131160656:131160844-131162572:131159611-131162064:131162477-131162572:-
22.	HBS1L	ENSG00000112339;MX:6:135039767-135042001:135042126-135054649:135039767-135050582:135050647-135054649:-
23.	TMBIM6	ENSG00000139644;MX:12:49741611-49741957:49742115-49752464:49741611-49743417:49743565-49752464:+
24.	LDHA	ENSG00000134333;AL:11:18403811-18405449:18405495:18403811-18407117:18408218:+
25.	SPCS2	ENSG00000118363;AL:11:74949399-74965034:74965169:74949399-74976857:74977056:+
26.	PABIR2	ENSG00000156504;AL:X:134769567:134771354-134781821:134771574:134772283-134781821:-
27.	PABIR2	ENSG00000156504;AL:X:134769567:134771354-134781821:134772079:134772283-134781821:-
28.	PROSER1	ENSG00000120685;AL:13:39011140:39011337-39012083:39011376:39011487-39012083:-
29.	NSA2	ENSG00000164346;AL:5:74770810-74771477:74771653:74770810-74776604:74776910:+
30.	RHBDD2	ENSG00000005486;AL:7:75881486-75881829:75882124:75881486-75887992:75888583:+
31.	RHBDD2	ENSG00000005486;AL:7:75881486-75881829:75882165:75881486-75887992:75888583:+
32.	N4BP1	ENSG00000102921;AL:16:48542880:48543261-48548007:48546182:48546254-48548007:-
33.	NDUFB5	ENSG00000136521;AL:3:179617044-179618415:179618477:179617044-179623920:179624499:+
34.	PRDX2	ENSG00000167815;AL:19:12796820:12797166-12800177:12799862:12799989-12800177:-

35.	HBS1L	ENSG00000112339;AL:6:135002782:135002842-135039573:135035857:135038026-135039573:-
36.	HBS1L	ENSG00000112339;AL:6:135002815:135002842-135039573:135035857:135038026-135039573:-
37.	MOK	ENSG00000080823;AF:14:102233789-102234957:102235063:102233789-102238318:102238446:-
38.	POMGNT1	ENSG00000085998;AF:1:46197871-46198051:46198134:46197871-46198336:46198434:-
39.	HOOK2	ENSG00000095066;AF:19:12774897-12778345:12778419:12774897-12792324:12792582:-
40.	ASAH1	ENSG00000104763;AF:8:18075587-18083981:18084091:18075587-18084672:18084817:-
41.	G3BP2	ENSG00000138757;AF:4:75662049-75675090:75675119:75662049-75724276:75724470:-
42.	G3BP2	ENSG00000138757;AF:4:75662049-75673208:75673548:75662049-75675090:75675119:-
43.	SLC16A3	ENSG00000141526;AF:17:82229042:82229106-82235983:82235774:82235842-82235983:+
44.	CUL4B	ENSG00000158290;AF:X:120558039-120559779:120559978:120558039-120560083:120560625:-
45.	CACNB3	ENSG00000167535;AF:12:48815591:48815728-48823344:48817866:48817963-48823344:+
46.	CAPZA2	ENSG00000198898;AF:7:116811070:116811102-116888127:116862600:116862650-116888127:+
47.	CAPZA2	ENSG00000198898;AF:7:116811070:116811102-116888127:116862614:116862650-116888127:+
48.	CAPZA2	ENSG00000198898;AF:7:116811070:116811102-116888127:116862587:116862650-116888127:+
49.	LDHA	ENSG00000134333;AF:11:18396818:18396968-18400837:18398630:18399548-18400837:+

50.	G3BP2	ENSG00000138757;AF:4:75662049-75673451:75673614:75662049-75675090:75675119:-
51.	G3BP2	ENSG00000138757;AF:4:75662049-75673208:75673411:75662049-75675090:75675119:-
52.	G3BP2	ENSG00000138757;AF:4:75662049-75673019:75673114:75662049-75675090:75675119:-
53.	TMBIM6	ENSG00000139644;AF:12:49738993:49739082-49752464:49743417:49743565-49752464:+
54.	TMBIM6	ENSG00000139644;AF:12:49743417:49743565-49752464:49750620:49750705-49752464:+
55.	TMBIM6	ENSG00000139644;AF:12:49741557:49741611-49752464:49743417:49743565-49752464:+
56.	CKB	ENSG00000166165;AF:14:103520311-103521073:103521101:103520311-103521263:103521443:-
57.	CKB	ENSG00000166165;AF:14:103520311-103521073:103521101:103520311-103521263:103521322:-
58.	CKB	ENSG00000166165;AF:14:103520311-103520469:103520728:103520311-103521263:103521322:-
59.	MAGED1	ENSG00000179222;AF:X:51893646:51893755-51895053:51894569:51894598-51895053:+
60.	RHOA	ENSG00000067560;AF:3:49375591-49411820:49411994:49375591-49412404:49412538:-
61.	RHOA	ENSG00000067560;AF:3:49375591-49411820:49411928:49375591-49412404:49412538:-
62.	CKB	ENSG00000166165;AF:14:103520311-103520469:103520728:103520311-103521263:103521443:-
63.	SLC16A3	ENSG00000141526;AF:17:82229068:82229106-82235983:82235774:82235842-82235983:+
64.	SLC16A3	ENSG00000141526;AF:17:82229055:82229106-82235983:82235774:82235842-82235983:+

65.	WWTR1	ENSG00000018408;AF:3:149527969-149541015:149541060:149527969-149542335:149542439:-
66.	WWTR1	ENSG00000018408;AF:3:149527969-149540216:149540266:149527969-149541015:149541060:-
67.	RHOA	ENSG00000067560;AF:3:49375591-49411820:49411970:49375591-49412404:49412538:-
68.	RHOA	ENSG00000067560;AF:3:49375591-49411820:49412097:49375591-49412404:49412538:-
69.	RHOA	ENSG00000067560;AF:3:49375591-49411820:49412000:49375591-49412404:49412538:-
70.	RHOA	ENSG00000067560;AF:3:49375591-49411820:49411976:49375591-49412404:49412538:-
71.	TFPI	ENSG00000003436;AF:2:187503770-187513616:187513641:187503770-187554200:187554435:-
72.	BMI1	ENSG00000168283;AF:10:22322106:22322170-22326431:22325657:22325728-22326431:+
73.	BMI1	ENSG00000168283;AF:10:22322106:22322170-22326431:22325933:22326076-22326431:+
74.	BMI1	ENSG00000168283;AF:10:22321585:22321696-22326431:22322106:22322170-22326431:+
75.	BMI1	ENSG00000168283;AF:10:22322106:22322170-22326431:22325670:22325722-22326431:+
76.	ZNF442	ENSG00000198342;AF:19:12353114-12363554:12363631:12353114-12365533:12365661:-
77.	MGLL	ENSG00000074416;AF:3:127781895-127791200:127791261:127781895-127799420:127799464:-
78.	RPS15A	ENSG00000134419;AF:16:18789118-18789897:18790260:18789118-18790310:18790334:-
79.	TM7SF3	ENSG00000064115;AF:12:26976357-26979784:26980271:26976357-26980566:26980608:-

80.	RHBDD2	ENSG00000005486;AF:7:75842602:75842749-75881365:75879004:75879260-75881365:+
81.	RHBDD2	ENSG00000005486;AF:7:75878091:75878186-75881365:75879004:75879260-75881365:+
82.	RHBDD2	ENSG00000005486;AF:7:75842602:75842749-75881365:75878091:75878186-75881365:+
83.	RHBDD2	ENSG00000005486;AF:7:75842602:75842749-75881365:75879172:75879260-75881365:+
84.	TMBIM6	ENSG00000139644;AF:12:49741557:49741611-49752464:49750620:49750705-49752464:+
85.	ESYT2	ENSG00000117868;AF:7:158799072-158827543:158827641:158799072-158829089:158829509:-
86.	BMI1	ENSG00000168283;AF:10:22321099:22321696-22326431:22322106:22322170-22326431:+
87.	RHBDD2	ENSG00000005486;AF:7:75878091:75878186-75881365:75879172:75879260-75881365:+
88.	EIF5A	ENSG00000132507;AF:17:7306999:7307115-7309615:7308371:7308575-7309615:+
89.	EIF5A	ENSG00000132507;AF:17:7306999:7307115-7309615:7307628:7307752-7309615:+
90.	EIF5A	ENSG00000132507;AF:17:7306999:7307115-7309615:7307326:7307409-7309615:+
91.	EIF5A	ENSG00000132507;AF:17:7306999:7307115-7309615:7307582:7307752-7309615:+
92.	R3HCC1	ENSG00000104679;AF:8:23270120:23270243-23288506:23288108:23288157-23288506:+
93.	R3HCC1	ENSG00000104679;AF:8:23270120:23270243-23288506:23287908:23288157-23288506:+
94.	PLOD2	ENSG00000152952;AF:3:146124229-146160789:146160992:146124229-146163594:146163653:-

95.	PLOD2	ENSG00000152952;AF:3:146124229-146160881:146161184:146124229-146163594:146163653:-
96.	PLOD2	ENSG00000152952;AF:3:146124229-146160881:146160991:146124229-146163594:146163653:-
97.	SLC16A3	ENSG00000141526;AF:17:82229062:82229106-82235983:82235774:82235842-82235983:+
98.	N4BP1	ENSG00000102921;AF:16:48548114-48551386:48551444:48548114-48553539:48553653:-
99.	MCM3AP	ENSG00000160294;AF:21:46285394-46285803:46285897:46285394-46286151:46286297:-
100.	NDUFB5	ENSG00000136521;AF:3:179604890:179604939-179616983:179615206:179616049-179616983:+
101.	EHBP1L1	ENSG00000173442;AF:11:65576053:65576406-65579078:65578201:65578278-65579078:+
102.	KPNB1	ENSG00000108424;AF:17:47669573:47669869-47670702:47670324:47670343-47670702:+
103.	NXPE3	ENSG00000144815;AF:3:101779211:101779324-101782110:101779495:101779934-101782110:+
104.	SET	ENSG00000119335;AF:9:128683655:128684007-128691170:128689975:128690026-128691170:+
105.	SET	ENSG00000119335;AF:9:128683831:128684007-128691170:128689975:128690026-128691170:+
106.	SET	ENSG00000119335;AF:9:128685083:128685213-128691170:128689975:128690026-128691170:+
107.	SET	ENSG00000119335;AF:9:128689222:128689655-128691170:128689975:128690026-128691170:+
108.	SET	ENSG00000119335;AF:9:128689975:128690026-128691170:128690250:128690727-128691170:+
109.	UBTF	ENSG00000108312;AF:17:44218296-44219009:44219121:44218296-44220740:44220963:-

110.	UBTF	ENSG00000108312;AF:17:44218296-44219445:44219675:44218296-44220740:44220963:-
111.	HBS1L	ENSG00000112339;AF:6:135042126-135050582:135050644:135042126-135054649:135054817:-
112.	HBS1L	ENSG00000112339;AF:6:135050647-135054649:135054884:135050647-135102928:135103056:-
113.	HBS1L	ENSG00000112339;AF:6:135050647-135054649:135054819:135050647-135102928:135103056:-
114.	ATOX1	ENSG00000177556;AF:5:151751779-151753815:151753895:151751779-151758546:151758631:-
115.	CALM2	ENSG00000143933;AF:2:47170764-47176441:47176511:47170764-47176849:47176921:-
116.	CALM2	ENSG00000143933;AF:2:47170764-47175237:47175539:47170764-47176849:47176921:-
117.	CALM2	ENSG00000143933;AF:2:47170764-47176032:47176223:47170764-47176849:47176921:-
118.	PAICS	ENSG00000128050;A5:4:56435975-56436262:56435928-56436262:+
119.	VARS1	ENSG00000204394;A5:6:31781952-31782041:31781952-31782087:-
120.	C16orf95	ENSG00000260456;A5:16:87315823-87317087:87315823-87317091:-
121.	MCM3AP	ENSG00000160294;A5:21:46285394-46285699:46285394-46285803:-
122.	STARD3	ENSG00000131748;A3:17:39653750-39656898:39653750-39656910:+
123.	MYL6	ENSG00000092841;A3:12:56159622-56159975:56159622-56160025:+
124.	MYL6	ENSG00000092841;A3:12:56158711-56159587:56158711-56159623:+
125.	ATP6V1G2-DDX39B	ENSG00000254870;A3:6:31540664-31541950:31540534-31541950:-
126.	RNF181	ENSG00000168894;A3:2:85596649-85596822:85596649-85596857:+
127.	ZNFX1	ENSG00000124201;A3:20:49275887-49278021:49275879-49278021:-

Appendix C: Publication