Precision Genomics for Prostate Cancer Patient Stratification

October 2019

George Alexander Donald Seed

Cancer Biomarkers

The Institute of Cancer Research

Thesis submitted to The University of London for the Degree of Doctor of Philosophy

Declaration

I confirm that the work presented in this thesis has been performed by me, unless otherwise acknowledged in the relevant sections.

George Alexander Donald Seed

Acknowledgements

I must first thank Professor Johann de Bono and Dr Wei Yuan for tremendous support and guidance throughout my graduate and professional training. It has been a genuine privilege to have such support.

I am hugely grateful to work in as multi-faceted a team as the de Bono laboratory, and honoured to work with such a cohort of excellent scientists and clinicians. A number of them contributed time, data, and advice to the studies presented here, and I have attempted to cite their contributions accurately throughout with the following initials. I have doubtless overlooked many.

Pathology Support: Daniel Nava Rodrigues (D.N.R.), Ruth Riisnaes (R.R.), Susana Miranda (S.M.), Mateus Crespo (M.C.), Ana Ferreira (A.F.) and Rita Pereira (R.P.).
Bioinformatic Support: Wei Yuan (W.Y.), Matt Clarke (M.C.) and Maryou Lambros (M.L.)
Sequencing Support: Suzanne Carreira (S.C.), Claudia Bertan (C.B.) and Joaquin Mateo (J.M.).
Clinical Support: Adam Sharp (A.S.), Joaquin Mateo (J.M.), Semini Sumanasuriya (S.S.), Mariane Fontes (M.F.).
Supervision and Guidance: Johann De Bono (J.D.B.), Wei Yuan (W.Y.), and Bissan Al-Lazikani (B.A.).

I benefited hugely from discussions with: Alan Mackay and Matthew Clarke from the ICR Glioma Team; James Campbell from the ICR Big Data Team and Harry Parr from the ICR Clinical Trials and Statistics Unit. I should also thank Gerhardt Attard, Carmel Pezaro and Gunther Boysen for their generous help with my early work at the ICR. My friends and family have, of course, been a great help.

In addition, this work would not be possible without the support from Prostate Cancer UK, or the brave patients who take part in studies at the Royal Marsden Hospital.

Abstract

Introduction

Several unmet needs currently persist in advanced prostate cancers. Despite recent discoveries, such as identification of deficient DNA repair pathways as a key driver of prostate tumour development, relatively few precision medicine approaches are available. Robust genomics is critical to characterise and identify such subsets, but there are key limitations that must be addressed such as tumour purity and heterogeneity. Additionally, for men without a genomic stratification, treatment options may be limited to aggressive chemotherapies such as taxanes, which currently have no genomic biomarkers to improve patient stratification.

Results

In this study, I first show how key DNA-repair subtypes of prostate cancers (~35% of cases) can be characterised using copy-number and microsatellite analysis of targeted panel sequencing data. Both data types were well correlated with orthogonal methods. I found that tumour heterogeneity (observable at the single cell-level *via* pathological analysis) limited the detection of copy-number events from bulk sequencing. I went on to explore tumour purity estimation in cell-free DNA sequencing of mCRPCs treated with taxanes, and found that baseline purity measurements were strongly associated ($p<0.001$) with survival in both univariable and multivariable analyses. Additionally, I found significant longitudinal shifts ($p<0.001$) in tumour purity between responders and non-responders to therapy.

I sought to test specific genes and genomic loci in this cohort of taxane-treated samples, and found that copy-changes of several genes including members of the tubulin gene family had associations with changes in survival. Additionally, I identified several genomic loci with strong associations with dug response status. As these studies are limited by tumour heterogeneity, I analysed a cohort of single cells, which allowed for a high-resolution examination of tumour sub-clones. These data displayed variable genomic heterogeneity, with clinically relevant alterations identified at the sub-clonal level. Importantly, I found changes in sub-clonal proportions could be observed longitudinally at the copy number level.

Conclusions

In summary, I show here methods for characterising the 30-40% of mCRPC that may respond to DNA-repair targeting therapies targeted sequencing, before going on to illustrate similar approaches, and potential future biomarkers, for taxane chemotherapies. Future validation of these will be key in improving taxane response rates. Additionally, while tumour purity correction is a limiting factor in tumour NGS, it also presents as a potentially useful clinical marker of tumour burden. Key to genomic studies of mCRPC is sample selection, and I showed that single-cell analyses and micro-dissection can be employed to bypass issues of tumour heterogeneity and purity by directly revealing clinically relevant alterations present at levels undetectable by traditional bulk sequencing methods. In the future, adoption of these methods will improve patient stratification and monitoring of cancer progression.

Publications

Derived From This Thesis

- Chapter 3 features work that was published in two separate publications:
    - Seed *et al*, Gene copy number estimation from targeted next-generation sequencing of prostate cancer biopsies: analytic validation and clinical qualification, CCR 2017 (Seed et al., 2017).
    - Rodrigues *et al*, Immunogenomic analyses associate immunological alterations with mismatch repair defects in prostate cancer, JCI 2018 (Nava Rodrigues et al., 2018).
- Chapters 4, 5 and 6 form part of a work in progress manuscript.
- Chapter 7 features work that was published:
    - Lambros, Seed and Sumanasuriya *et al*, Single-Cell Analyses of Prostate Cancer Liquid Biopsies Acquired by Apheresis, CCR 2018 (Lambros et al., 2018).
    -

Other Relevant Work

- During my studies, I contributed to the following review:
    - Rodrigues *et al*, The molecular underpinnings of prostate cancer: impacts on management and pathology practice, J Pathology 2017.
- And my work is also featured in the following original research articles:
    - Welti *et al*, Targeting bromodomain and extra-terminal (BET) family proteins in castration-resistant prostate cancer (CRPC), CCR 2018.
    - Boysen *et al*, SPOP-mutated/CHD1-deleted lethal prostate cancer and abiraterone sensitivity, CCR 2018.
    - Sharp *et al*, Androgen receptor splice variant-7 expression emerges with castration resistance in prostate cancer, JCI 2019.
    - Sharp *et al*, Clinical utility of circulating tumour cell androgen receptor splice variant-7 status in metastatic castration-resistant prostate cancer, Eur Urol 2019.
    - Paschalis *et al*, Prostate-specific membrane antigen heterogeneity and DNA-repair defects in prostate cancer, Eur Urol 2019.

Table of Contents

Table of Abbreviations

ACRONYM	DEFINITION
(m)CRPC	(Metastatic) Castration-Resistant Prostate Cancer
ADT	Androgen Deprivation Therapy
IHC	Immunohistochemistry
FISH	Fluorescence In-Situ Hybridisation
cfDNA	Cell-Free DNA
FFPE	Formalin-Fixed Paraffin-Embedded
HR	Homologous Repair
MMR	Mismatch Repair
NHEJ	Non-Homologous End Joining
NER	Nucleotide Excision Repair
BER	Base Excision Repair
SNA/V	Single Nucleotide Alteration/Variation
CNA	Copy Number Alteration
SV	Structural Variation
MSI	Microsatellite Instability
NGS	Next Generation Sequencing
(lp)WGS	(Low-Pass) Whole Genome Sequencing
WES	Whole Exome Sequencing
aCGH	Array-Comparative Genome Hybridisation
LDH	Lactate Dehydrogenase
HB	Haemoglobin
ALB	Albumin
ALP	Alkaline Phosphatase
PSA	Prostate-Specific Antigen
OS	Overall Survival
RPFS	Radiographic Progression-Free Survival

Table of Figures

Table of Tables

1 Introduction

1.1 CRPC Background

1.1.1 Epidemiology of the Disease

Prostate tumours are the most common male cancer, accounting for around 10% of all cancer diagnoses worldwide. Around 80% of men over the age of 80 in Europe will experience some form of the disease (Kirby et al., 2011; Litwin and Tan, 2017; Marta et al., 2013). The reported incidence of prostate cancer differs, however, due to variable implementation of screening and early detection procedures.

There is a significantly higher likelihood of developing prostate cancer in individuals with positive family histories of the disease, suggesting a strong inherited genetic component (Goldgar et al., 1994). In addition the prevalence of the disease seems to be linked with environmental risk factors, as suggested by the rise in prostate cancer risk among immigrants to western countries (Marta et al., 2013; Marugame and Katanoda, 2006).

While most cases are relatively slow growing and non-invasive, a minority of tumours (~25%) present a faster growing and lethal phenotype and this combined with the overall high frequency of the disease presents a challenge for physicians in both identification and treatment (Attard et al., 2016; Rodrigues et al., 2017).

1.1.2 Clinical Overview of the Disease

The prostate is a component of the male reproductive system; a hormone-controlled gland located in the pelvis of men adjacent to the bladder and rectum. Tumours that develop in the prostate are initially confined to the gland itself.

These locally confined, low-grade tumours are often monitored by ‘active surveillance’ – sporadic examinations and biopsies – due to their indolent nature. Aggressive therapies are therefore unappealing and unnecessary (Filson et al., 2015; Rodrigues et al., 2017). Androgen deprivation therapy (ADT) is used to control prostate tumour growth by chemical castration and offers substantial

benefit (Mottet et al., 2017). Surgical resection or radiotherapy also offer benefits to men with higher grade (more active) and improve outcomes, albeit with associated morbidities (Litwin and Tan, 2017; Tosoian et al., 2015). In many men, these measures will be sufficient to control the tumour, and frequently effectively cure the disease itself (Mottet et al., 2017).

Cancer that has metastasized from the prostate is treated initially with hormone modulation – more androgen deprivation – until resistance inevitably emerges, becoming castration-resistant prostate cancer (CRPC) (Kirby et al., 2011; Lorente et al., 2015). It is important to note that while metastatic spread (as a part of tumour growth) and the emergence of castration resistance (in response to ADT) are separate processes, almost all castration-resistant tumours feature some metastatic spread (Kirby et al., 2011).

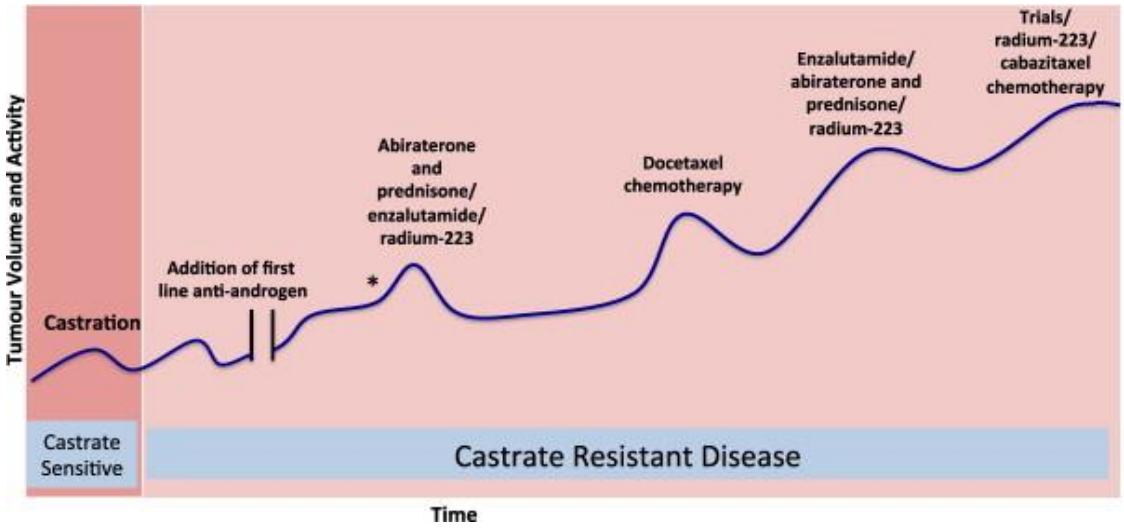


Figure 1-1 Emergence of resistance to therapies in CRPCs.

Despite temporary responses to anti-androgen therapies and other chemotherapeutic agents, prostate tumours that become castration-resistance inevitably progress on treatment. Temporary tumour shrinkage is a common feature of the disease, with durable responses of many months observed in some cases, but true curative therapy is unusual when the disease has developed castration resistance.

The most common metastatic sites include nearby pelvic bone marrow (~80% of cases), distant lymph nodes (~10%), and liver (~10%); visceral disease (metastatic spread to organs) has a significantly worse prognosis than cases restrained to bone or lymph modes (Gandaglia et al., 2014, 2015; Pezaro et al., 2014). Second line therapies including more complete androgen receptor signalling blockade (abiraterone and enzalutamide), immunostimulation (sipleucel-T) and tubulin-binding drugs (taxane chemotherapies) may slow

disease progression and improve survival (Litwin and Tan, 2017; Rodrigues et al., 2017).

Nearly all individuals with mCRPC will eventually develop resistance to these therapies (Figure 1-1). The optimal sequence of treatments for CRPC is not yet well understood, due to the limited set of approved therapies, a lack of comparative trials, and an absence of predictive biomarkers (Lorente et al., 2015; Sartor and de Bono, 2018). Only recently have true 'precision medicine' approaches been successfully applied in the care of men suffering from advanced prostate tumours (Mateo et al., 2015). Cancers are fundamentally genomic diseases, and developing the future of prostate cancer therapeutics will require an understanding of the driving molecular and genomic forces behind each individual tumour.

1.2 The Genomic Basis of CRPC

1.2.1 The Primary Prostate Cancer Genome

Inherited genetic factors contribute significantly to prostate cancer initiation: inherited factors can be linked to up to 50% of prostate cancer risk (Mucci et al., 2016; Pritchard et al., 2016). Around a hundred common variants in the genome have been identified in genome-wide association studies as conferring a mild increase in risk of prostate cancer when mutated (Cheng et al., 2017; Eeles et al., 2014). In addition, aberrations in a small subset of genes (such as *BRCA1/2*) radically increase prostate cancer risk for those affected by them and increase risk of developing severe disease. Around 8% of lethal metastatic prostate tumours feature an inherited defect in a small number of genes involved in DNA-repair, an enriched frequency compared to ~1% in non-lethal primary tumours (Attard et al., 2016; Pritchard et al., 2016).

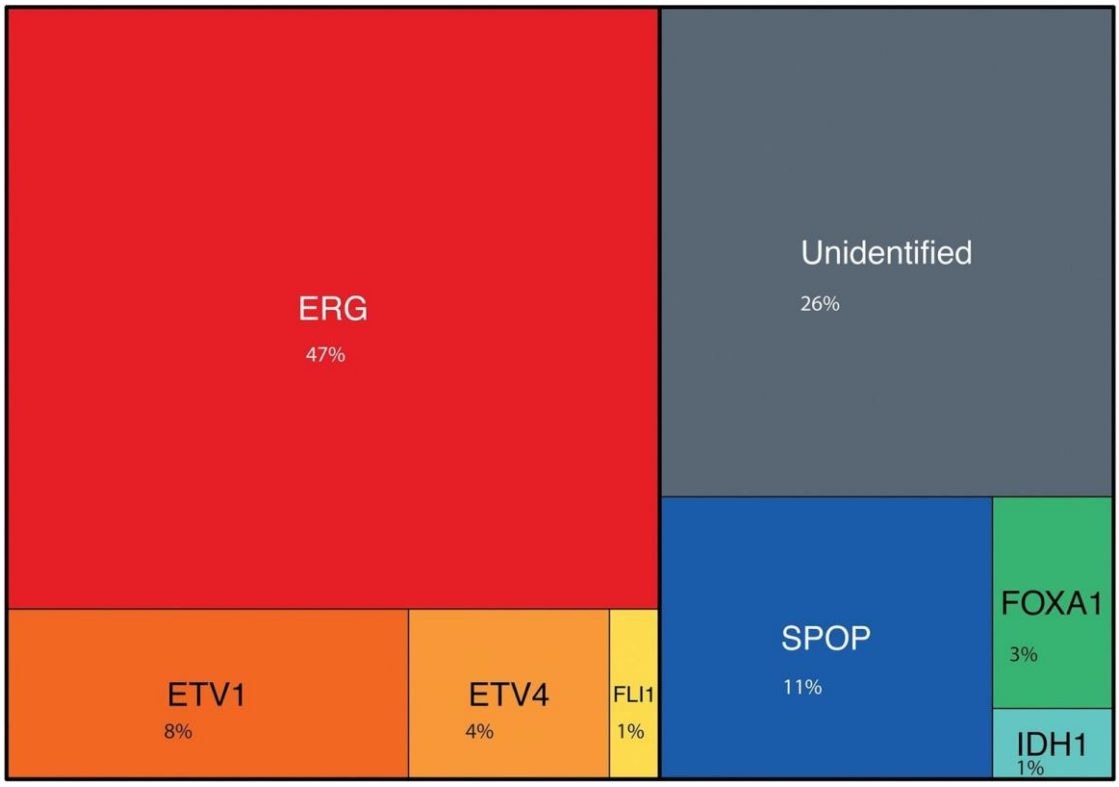


Figure 1-2 Molecular subtypes of primary prostate cancers.

Primary tumours can be grouped into ETS fusion-positive (red-yellow) and fusion-negative (blue-green). In around a quarter of cases a clear subtype is not easily identified. Area shown for each subtype relates to estimated proportion of tumours in that group (Cancer Genome Atlas Research Network, 2015; Rodrigues et al., 2017).

The Cancer Genome Atlas study analysed 333 primary prostate cancers and identified a number of recurrent aberrations that could be implicated in early tumour development, defining seven broad subclasses (Cancer Genome Atlas Research Network, 2015). The largest subclass is characterised by gene fusions of E26 transformation-specific family genes (*ERG*, *ETV1*, *ETV4*, or *FLI1*) with androgen receptor-regulated promoters (such as *TMPRSS2*) that occur in a mutually exclusive pattern. The fusions that characterise this group of tumours are not limited to single genes, and may occur at multiple sites during a simultaneous event (Berger et al., 2011).

The remaining three subtypes are defined in turn by the presence of variants in *SPOP*, *FOXO1* and *IDH1*. However, there remain prostate tumours that do not fit any of these criteria (Figure 1-2) (Cancer Genome Atlas Research Network, 2015; Rodrigues et al., 2017). Mutations in other key oncogenic pathways (cell

cycle, growth, migration) have also been frequently observed in primary tumours (Fraser et al., 2017; Teply and Antonarakis, 2017).

Tumours that develop into CRPC typically retain mutations acquired during earlier development and acquire a wide array of additional changes. Compared to other cancer types, CRPCs generally feature relatively modest numbers of somatic point mutations and small insertions/deletions (a mutation rate of ~4-5 per megabase), and a greater number of structural variants, including fusions and large copy-number events (one recent study observed ~230 structural events per sample) (Baca et al., 2013; Greenman et al., 2007; Robinson et al., 2015; Viswanathan et al., 2018).

1.2.2 Androgen Signalling and the Androgen Receptor

The most frequently identified genomic feature of CRPCs is an alteration of the androgen receptor signalling pathway. Detectable aberrations are present in ~70% of tumours and promote downstream functions of the pathway. In contrast, primary tumours have very few of these alterations (Robinson et al., 2015). Events in this pathway largely occur due to the selective pressure of initial ADT, highlighting the lengths that CRPCs go to 'escape' androgen blockade. Artificial reactivation of the AR pathway in hormone-sensitive prostate tumours is sufficient to convert the tumour to a castration-resistant phenotype (Chen et al., 2004). Prostate tissue requires androgen signalling to grow and thrive: and so do prostate tumours.

The androgen receptor (AR) is a nuclear receptor of androgen steroid hormones (such as testosterone) and functions as a DNA-binding transcription factor that regulates transcription of a wide variety of genes (Tan et al., 2015). The protein is made up of three main domains: the C-terminal ligand binding domain (LBD), the DNA binding domain (DBD), and the N-terminal domain (NTD). Following binding of androgen molecules to the LBD, a conformational change in the protein occurs leading to translocation to the nucleus. The AR then binds to specific promoter and enhancer elements (a hormone-response element) in the genome by the DBD, resulting in transcriptional modulation by the NTD (Cutress et al., 2008; Tan et al., 2015). The truncated NTD is constitutively active; with the LBD regulating full length AR through the presence of bound androgens. This 'classical pathway' – the ability to function as a transcription factor - is modulated

by coregulators and itself regulates the expression of numerous genes (Davey and Grossmann, 2016; De Gendt et al., 2014).

Genomic modifications to the AR pathway all have one goal: activation of downstream AR partners despite limited supplies (or complete absence) of activating androgenic hormones (Coutinho et al., 2016). Increased expression of the AR is commonly observed in CRPCs, and acts to hyper-sensitise the tumour cells to androgen stimulation (Grasso et al., 2012; Kawata et al., 2010; Robinson et al., 2015). Overexpression can be mediated by gene copy-number changes of the AR, with numerous extra copies (ie. gene amplifications) occurring in 30-50% or CRPC cases: more copies of AR usually correlate to higher mRNA and protein expression (Edwards et al., 2003; Grasso et al., 2012). Structural events may also rearrange the AR gene, with imperfect copying of the exons leading to aberrant protein production. Data is also emerging that suggests that extra-chromosomal DNA structures may contain many amplified genes which are hidden from standard transcriptional regulation (Turner et al., 2017).

Activating hotspot mutations of the AR have been observed in around 15% of cases. These mutated variants also confer a selective advantage to CRPC cells in an androgen-depleted environment, and cluster in the ligand-binding or n-terminal domains (Carreira et al., 2014; Chen et al., 2015; Robinson et al., 2015). Ligand-binding domain mutated variants can modify AR function by increasing promiscuity for range of ligands binding to the receptor (such as progesterone, oestrogen and glucocorticoids). This facilitates AR activation, and in some cases leads to differential coactivator recruitment with knock-on effects for AR regulation (Brooke et al., 2008; Coutinho et al., 2016; Zhao et al., 2000). AR LBD mutations can also account for antagonist-agonist switching: a change in the response to drug binding that allows drugs that traditionally inhibit the receptor (eg. bicalutimide, flutamide) to act as activating androgens (Hara et al., 2003; Lallous et al., 2016; Tan et al., 1997). These mutations are enriched in heavily-pretreated (ie. with anti-androgen therapies) cohorts and are likely the result of strong treatment induced selective pressures. Mutations of the N-terminal domain tend to play more structural roles, facilitating protein stability, the positioning of the AR in the nucleus, and the binding of coactivators and transcriptional partners (Buchanan et al., 2004; Steinkamp et al., 2009).

A third key type of alterations target AR signalling by modulating exon splicing mechanisms to create protein products capable of ligand-independent transcription factor activity (Dehm et al., 2008). AR-variants (AR-V) (as opposed to full-length AR) can be generated from exon-skipping or the addition of non-canonical exons, which may abrogate the function of the regulatory LBD and thus lead to a constitutively active molecule. Numerous AR-Vs have been identified and are expressed at variable levels; however the various roles and mechanisms have not been fully elucidated (Wadosky and Koochekpour, 2017; Watson et al., 2010). For example, while the expression of AR-V7 is elevated in CRPC compared to hormone sensitive disease, expression levels remain relatively low when compared to the full length receptor (Guo et al., 2009). Some data suggests that expression of AR-Vs may offer activation of alternative pathways, driving a more aggressive phenotype. Overall, it is reported that AR-Vs associate with poor clinical outcomes from CRPC - perhaps due to potent signalling activity even in low concentrations (Efstathiou et al., 2015; Hu et al., 2012; Welti et al., 2016). It is important to note that alternative splicing is not a mechanism unique to tumours, and that under normal circumstances offer the cell another level of fine-grained regulation of protein activity (Zhang and Manley, 2013).

While the AR gene is a key part of the overall androgen metabolism pathway, and modifications to it are a hallmark of prostate tumour development (and castration resistance in particular), other pathway members also exhibit alterations (Robinson et al., 2015). AR coactivators such as the SRC-genes (which directly bind the AR to support transcription factor activity) may be overexpressed in prostate tumours and have been associate with more aggressive disease and poor outcomes (Agoulnik and Weigel, 2009; Grasso et al., 2012; Qin et al., 2014).

Similarly, aberrations of the pioneer factor FOXA1 have also been shown to enhance AR transcriptional activity, although some data shows similar effects for loss-of-function mutations as well (Barbieri et al., 2012; Jain et al., 2011; Robinson et al., 2015; Yang and Yu, 2015). In contrast, AR repressors such as NCOR1/2 (which compete with AR coactivators to repress signalling), and SPOP (which binds to and degrades AR) may be mutated or deleted outright, allowing uninhibited AR to have a longer half-life and persist in the cytosol (Barbieri et al., 2012; Boysen et al., 2018; Yoon and Wong, 2006).

1.2.3 DNA-Repair Defects

Recent work has highlighted a critical role for DNA-repair defects in the development of CRPC (Mateo et al., 2017; Pritchard et al., 2016). DNA in cells is continuously under the threat of damage (from radiation, chemical and molecular agents, DNA replication errors, etc), and cellular pathways constantly monitor for and react to this damage.

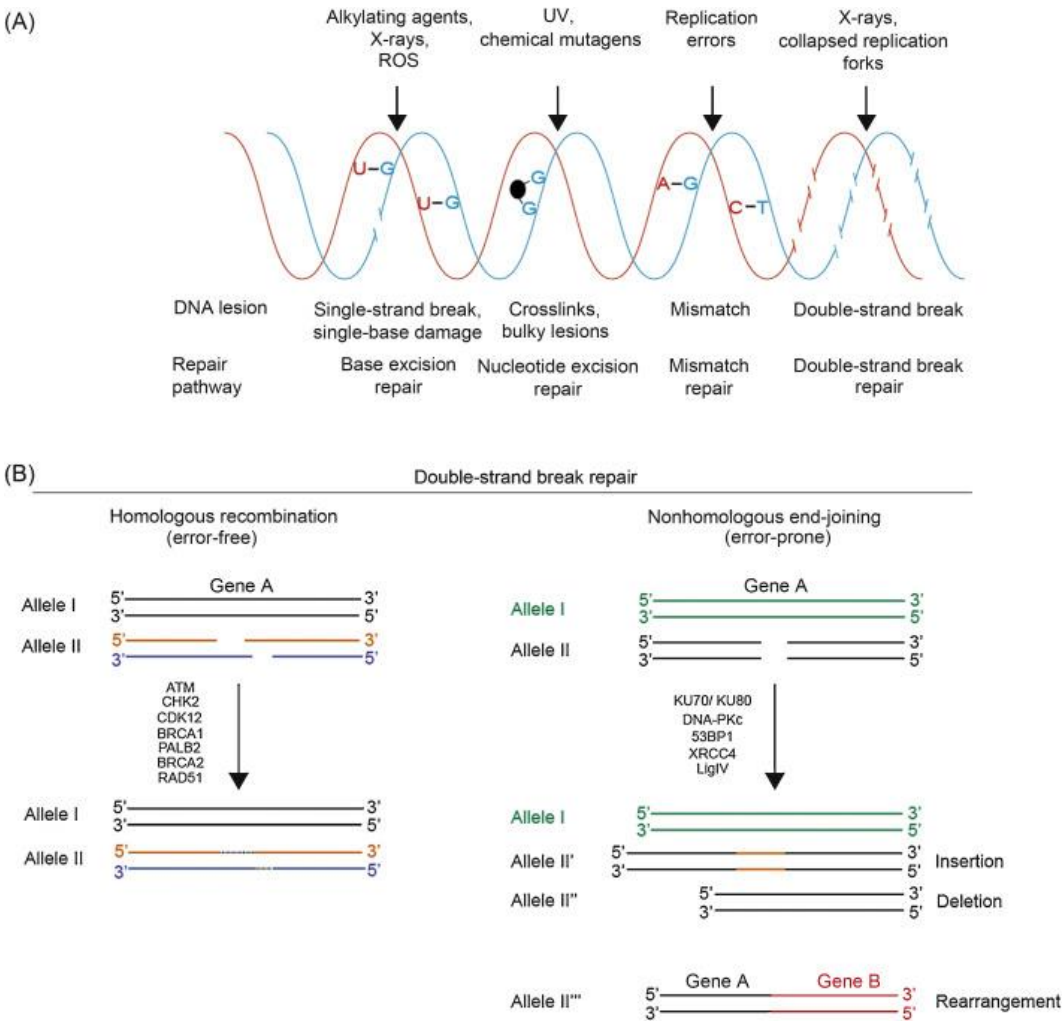


Figure 1-3 Types of DNA damage and associated repair mechanisms.

DNA can be damaged via various mechanisms, and to different extents. This damage may be repairable, with a suite (A) of cellular pathways tailored to specific lesions. Double strand breaks (DSB) (B) require complex solutions and may lead to genomic rearrangements if not adequately solved. Homologous Recombination offers a (theoretically) perfect repair to a DSB, while Non-Homologous End-Joining has wider repair capabilities but is more error prone (Mateo et al., 2017).

Following detection of damage, cell cycle arrest is enacted while repairs are attempted; if successful, the cell cycle is resumed. In cases where DNA-repair is not possible, programmed cell-death or cellular senescence pathways are

activated. Dysfunction in these processes can allow for genome instability - an enabling characteristic of tumourigenesis (Hanahan and Weinberg, 2011; Mateo et al., 2017). DNA-repair defects can therefore be thought of as a source of genomic instability, fuelling the ability of tumours to adapt to the environment.

Depending on the type of damage, different molecular processes will be called upon to perform the required repair (Figure 1-3). Damage impact the base sequence of DNA may be repaired using base-excision repair (typically single nucleotide damage), nucleotide excision repair (for molecules inappropriately adhered to the DNA), and mismatch repair (correcting small errors from DNA replication) (Mateo et al., 2017; Pritchard et al., 2016).

Under normal conditions the mismatch repair (MMR) system is capable of recognising mismatched nucleotides (and insertion-deletion loops), with single base events recognised by a complex of MSH2 and MSH6, and longer (2 to 8 nucleotides) mismatches addressed by MSH2 and MSH3. Once recognised, repair of the lesion is initiated by MLH1-PMS2 or MLH1-PMS1 complexes (Poulogiannis et al., 2010). In addition, there is some cross-talk with the nucleotide excision repair pathway. MMR pathway members are altered in some CRPC cases (3-10%), and this deficiency leads to insufficient detection of small variants, particularly at short stretches of repeats termed micro-satellites.

Double-strand breaks are significant events that, if allowed to persist through cell division, can result in large chromosomal abnormalities. Repair of these is of critical importance to the cell, and is controlled primarily by two main processes, homologous recombination (HR) and nonhomologous end-joining (NHEJ) (Figure 1-3) (Mateo et al., 2017). Homologous recombination is a high-fidelity system that relies on the presence of a matching ‘sister’ chromatid, which acts as a template.

Double stranded breaks are recognised by the MRN complex, tagged with RPA, and then bound by Rad51/BRCA1/BRCA2, which mediate repair via invasion and replication of the sister strand (Brochier and Langley, 2013). Some HR pathway genes such as *BRCA2* and *ATM* are commonly deleteriously altered in CRPCs (~30%), although other pathway members exhibit alterations at much reduced frequencies. (Mateo et al., 2015; Robinson et al., 2015). HR is predominantly active during mid-S to G2 phase of the cell cycle as it requires an available homologous template, whereas NHEJ may be active throughout. NHEJ directly

ligates double strand breaks, and while it functions as a ‘guardian’ of the genome by repairing breaks unresolved by HR, it is also more prone to erroneous repair, and allows more unrepaired breaks to persist through replication (Chang et al., 2017). Unrepaired breaks lead to genome instability, and disrupted NHEJ (through aberrations in genes such as XRCC5, XRCC6 and PRKDC) increases genome instability still further. The AR directly regulates the activity of the NHEJ pathway, and castration has been shown to impair NHEJ activity, which may explain observations of tumour radiosensitisation as a result of castration (Tarish et al., 2015)

In some tumours, environmental damage is a driving force behind the acquisition of mutations. The base excision repair (BER) pathway exists to repair single-base lesions - either spontaneous or generated by chemical damage (oxidised, alkylated and deaminated nucleotides), which are recognised by a range of DNA glycolsylases which go on to initiate base removal and repair (Krokan and Bjørås, 2013). UV radiation generates helix-distorting bulky adducts to the DNA, and the nucleotide excision repair (NER) pathway members monitor the DNA for these large structural events. Two main NER subpathways monitor the DNA for damage: global genomic repair (GCR) which acts on all DNA strands, and transcription-coupled repair (TCR) which focuses on transcriptionally active strands (Lockett et al., 2005). The subpathways have distinct methods of damage recognition but broadly similar repair processes. There is an association between inherited polymorphisms in BER and NER pathway genes (such as *XRCC1* and *ERCC2*) and prostate cancer risk (Mandal et al., 2012).

1.2.4 Cell Cycle Aberrations

Aberrations in genes that control cell cycle progression also feature in the CRPC genome, with a radical difference in the number of events targeting these genes from primary (5%) to late-stage disease (21-55%) (Grasso et al., 2012). Germline events in this pathway have also been implicated in increasing the risk of developing severe disease (Kibel et al., 2003). When prostate tumour cells are deprived of AR pathway activation, cell-cycle progression is halted, and avoiding this is a key element of CRPC development (Schiewer et al., 2012). Under normal conditions, cell-cycle progression is enabled by cyclin and cylin-dependent kinase (CDK) protein complexes, the activity of which is controlled by the availability of cyclins and the presence of CDK inhibitors (Malumbres and

Barbacid, 2007). In advanced prostate tumours however, and following AR pathway activation, these processes are aberrantly promoted. Cell-cycle activation and proliferation is induced initially by MTOR (mammalian target of rapamycin), a serine-threonine kinase which promotes CDK4/6 activation by increasing translation of D-Cyclins, leading to subsequent inactivation of RB (retinoblastoma protein) (Figure 1-4) (Comstock et al., 2013; Schiewer et al., 2012).

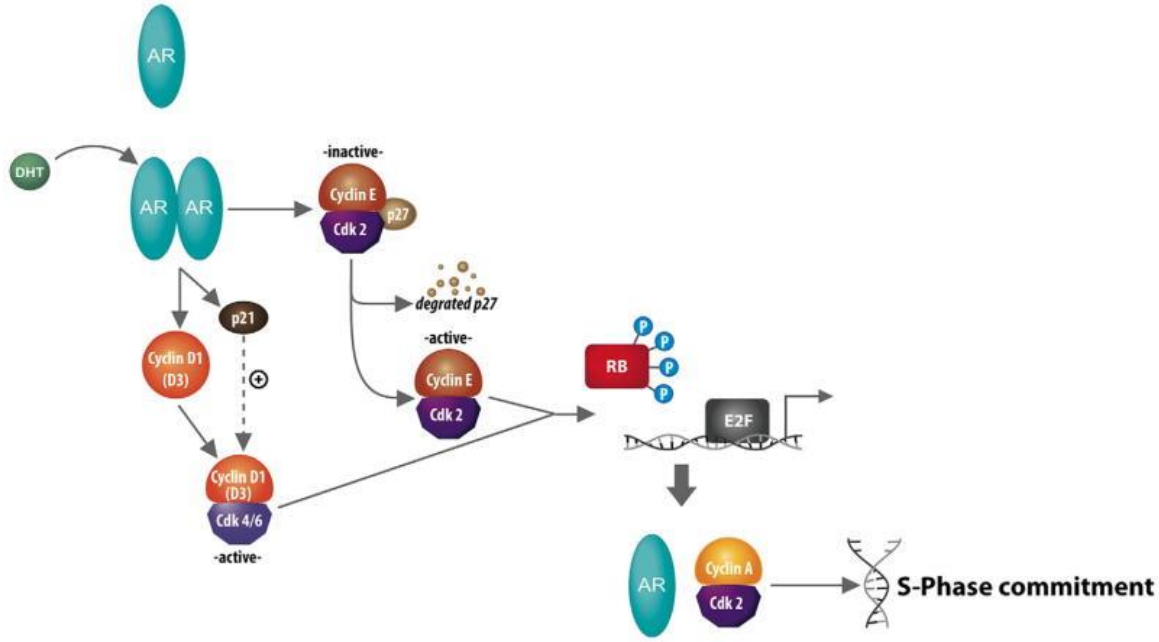


Figure 1-4 Activation of cell-cycle processes by the Androgen Receptor.

Activated AR promotes cell-cycle progression via activation of CDK4/6 and subsequent inactivation of RB, which allows for E2F gene translation and initiates pathways required for transition to S-phase and subsequent cell-cycle progression (Schiewer et al., 2012).

Active RB1 acts to limit cell-cycle progression by reversibly binding to E2F-family transcription factors, which thereby limits progression of the cell into S-phase and mitosis. Loss of the RB1 tumour-suppressor gene is a common (~20-30% of cases) event in the development of CRPC, and it has been implicated in modulating the effect of AR-pathway genes and perhaps sensitising cells to taxanes (de Leeuw et al., 2015; Robinson et al., 2015; Sharma et al., 2010). In addition, data has emerged suggesting that cell cycle deregulation, in particular mediated by RB1 complete loss, is a defining characteristic of CRPC tumours exhibiting neuroendocrine differentiation (Beltran et al., 2016; Tan et al., 2014). Mutations and copy-number changes have also been observed in numerous cancers, including CRPC, in the tumour-suppressor CDKN2A gene, which

produces cell-cycle regulatory proteins p16 and p21 that regulate cyclin and subsequent RB1 activity (Cao et al., 2018; Reis et al., 2015; Zeng et al., 2018).

In addition to AR activity promoting cell-cycle progression, there are crosstalk mechanisms that can feedback and modulate AR activity (Schiewer et al., 2012). Cyclin-dependent kinases may interact with AR directly; CDK6 can enhance ligand dependent transcriptional activity, and both CDK1 and CDK5 increase AR stability (and thus signaling) via phosphorylation (Chen et al., 2006; Hsu et al., 2011; Lim et al., 2005). Additionally, both CDK7 and CDK9 play a role in promoting AR activity by regulation of transcription factor activity, with inhibition of these proteins acting to limit AR pathway activation (Rahaman et al., 2016; Rasool et al., 2019).

A recent study has shown that CRPCs may exhibit functional loss of the cyclin-dependent kinase CDK12. Previously implicated in PARP-inhibitor sensitisation in ovarian tumours, it has now been described as a novel source of genomic instability in CRPC (Bajrami et al., 2014; Ekumi et al., 2015; Wu et al., 2018). CDK12-deficient CRPCs are associated with a genome characterised by frequent copy number events and gene translocations and very few point mutations. In contrast to HR-deficient tumours the structural events are small in size, with an overall diploid genome.

1.2.5 PI3K-AKT pathway

Half of CRPC cases (compared to around a quarter of primary tumours) bear a somatic alteration in the PI3K-AKT pathway, with deletions of *PTEN* and amplifications and hotspot mutations in *PIK3CA* and *AKT1* seen most frequently (Cancer Genome Atlas Research Network, 2015; Robinson et al., 2015). This pathway is involved in a broad range of cellular processes including proliferation and cell motility. Activated Phosphoinositide 3-Kinase family genes (encoded by *PIK3CA, B* and *D*) bind directly to, phosphorylate and subsequently activate Protein Kinase B (AKT) mediated by second messengers PIP2 and PIP3. (Figure 1-5). This allows AKT to localise to the plasma membrane and modify other key signalling molecules, including p21 (CDKN1B), FOXO1 and MTOR (Carnero and Paramio, 2014; Crumbaker et al., 2017) (Figure 1-5). A hyperactive PI3K-AKT pathway promotes cell growth, proliferation and cell cycle acceleration, and in addition has been shown to have substantial cross-talk with AR signaling (Carver

et al., 2011; Sarker et al., 2009). In addition, the oncogenic properties of PI3K genes (in particular *PIK3CA* and *PIK3CB*) have been well established, and hyperactivity of the pathway is a feature of numerous cancer types (Samuels and Waldman, 2010).

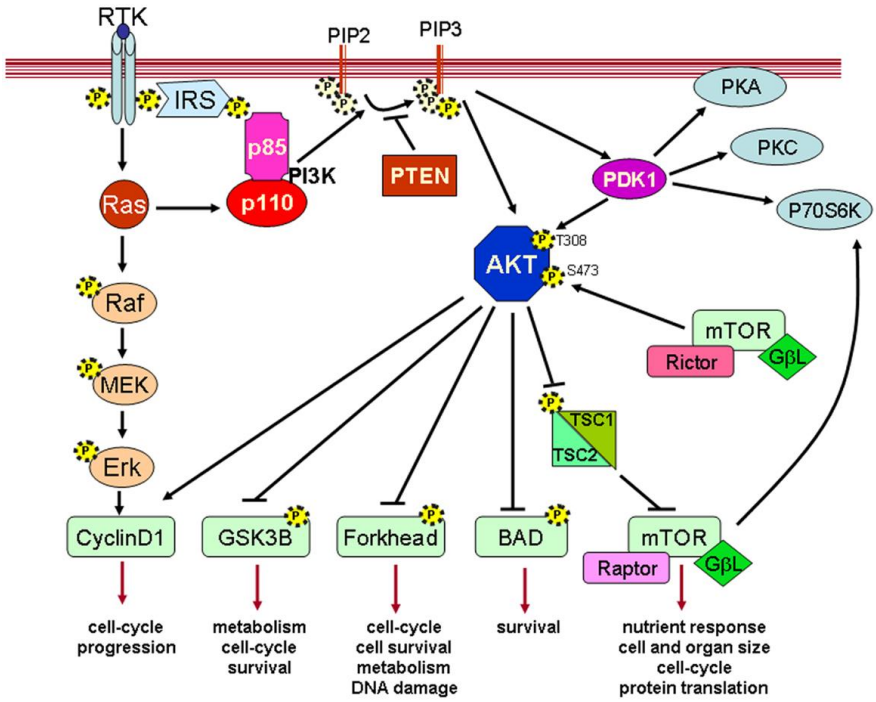


Figure 1-5 Summary of PI3K-AKT Pathway

Activated PI3K signals through PIP2/3, activating AKT and subsequently modulating a host of downstream processes. Substantial cross-talk with other molecular signalling pathways also occurs (Carnero and Paramio, 2014).

Aside from up-regulation of the PI3K/AKT axis, oncogenic activity of the pathway may also be mediated by functional loss of the tumour suppressor *PTEN* (phosphatase and tensin homolog). Under normal conditions, *PTEN* acts to reverse the conversion of PIP2 to PIP3, limiting AKT activation. Loss of *PTEN* - through a variety of mechanisms - occurs in ~40% of cases and leads to the development of more aggressive tumours and a worse prognosis (Grasso et al., 2012; Reid et al., 2010). In addition, losses of *PTEN* have been shown to increase in frequency as prostate tumours develop from primary to castration-resistant disease (Ferraldeschi et al., 2015; Vivanco and Sawyers, 2002).

Non-canonical signalling of the AR (ie. not transcription factor activity) is capable of activating the PI3K-AKT pathway by ligand-activated AR directly interacting

with PI3K, leading to subsequent accumulation of phosphorylated AKT. Further, downstream AKT substrates FOXO3 and FOXO1 can modify AR activity by binding to the AR promoter and interacting with histone deacetylases (HDACs) (Liu et al., 2008; Yang et al., 2005). AKT itself has been shown to bind directly to the AR protein through HER-2 activation, and can act as an androgen-independent AR activator in a feedback loop (Wen et al., 2000). Loss of *PTEN* modulates expression of AR-target genes and bypasses androgen-dependent signalling (Mulholland et al., 2011). These mechanisms are part of a reciprocal feedback loop: inhibition of one pathway activates the other (Carver et al., 2011). PI3K-AKT aberrations, aside from being oncogenic, are therefore key to prostate tumour response to AR inhibition, permitting growth in an androgen-deprived environment.

1.2.6 Chromatin Modification

Systems that control chromatin structure are also emerging as features of advanced prostate tumours, although the clinical relevance of this has not yet been studied in detail (Rodrigues et al., 2017). DNA strands are packaged into chromatin complexes, by first winding around histone proteins that are subsequently bundled into dense, ordered coils. Chromatin structures are dynamic: the degree to which a particular length of DNA is compressed has a subsequent effect on gene expression, as closed chromatin prevents molecular machinery accessing the DNA, and therefore limits transcription. However, 'open' chromatin is both more transcriptionally active and susceptible to damage (Längst and Manelyte, 2015).

Modifications of histone proteins – post-translational changes such as additions of methyl or acetyl groups (methylation and acetylation) - are common regulators of chromatin structure and have been shown to be aberrant in many tumours (Ruggero et al., 2018). Patterns of histone modification have been shown to segregate prostate cancers into high and low-risk groups (Seligson et al., 2005). Enzymes with the ability to add or remove post-translational modifications are thus capable of enabling or disabling transcription at specific sites. Remodelling of chromatin is common to prostate cancer, with the chromatin structure of androgen response elements (containing binding sites for androgen receptor transcription factor activity) known to be extensively relaxed during tumour initiation and progression (Pomerantz et al., 2015; Sharma et al., 2013). In

addition, castration-resistant prostate tumours exhibit markedly different levels of open chromatin at AR binding sites compared to primary cases, suggesting that this is another mediator of resistance to androgen deprivation.

Relaxing the chromatin structure may also be associated with structural events and genomic instability, a hallmark of prostate cancers, with these linked to sites of transcriptionally active chromatin (Gerhauser et al., 2018). Structural rearrangements in CRPC cluster in AR binding sites, and these regions present a higher rate of double-strand breaks (Gerhauser et al., 2018; Weischenfeldt et al., 2013). The chromatin remodeller CHD1 is commonly deleted in prostate cancer, and may represent a distinct subclass of prostate tumours. Loss of CHD1 leads to an altered proportion of double-strand break repair pathway utilisation, favouring NHEJ over HR and has been demonstrated to increase genome instability (Shenoy et al., 2017)).

Transcriptional repression by chromatin condensation can also be modified in tumours, with aberrations in the SWI-SNF pathway (and others) observed (Grasso et al., 2012). Chromatin condensation promotes winding of the DNA into tightly wound higher-order structures, inaccessible to transcription factor binding, and loss of SWI/SNF genes permits the activation of a suite of pro-proliferation genes such as the WNT/β-Catenin pathway (Kim and Roberts, 2014; Längst and Manelyte, 2015; Prensner et al., 2013).

Alterations to enzymes capable of adding or removing post-translational modifications to histone proteins such as histone lysine demethylases (KDM-family proteins) have also been linked to prostate tumours (Crea et al., 2012). Depending on their site-specific binding affinities, different KDMs may act with oncogenic or tumour-suppressive functions (Crea et al., 2012). For example, KDM1A activates the transcriptional functions of the MYC oncogene and suppresses p53 activity, and is also capable of forming a co-activation complex with AR, leading to transcription of AR target genes (Crea et al., 2012; Metzger et al., 2005; Stratmann and Haendler, 2012). Conversely, KDM2A maintains genome stability and represses transcription of pro-proliferative genes: accordingly it has been observed to be under-expressed in prostate tumours (Borgel et al., 2017; Frescas et al., 2008).

1.3 Actionable Subclasses of CRPC

1.3.1 Homologous Repair Deficiency and PARP Inhibition

As previously mentioned, ~30% of CRPCs feature deficiencies in DNA repair including homologous repair (HR) (Mateo et al., 2015; Robinson et al., 2015). HR deficiencies offer an advantage to tumours by leaving double strand breaks to be repaired by an alternate mechanism, non-homologous end-joining (NHEJ) – a more error-prone pathway which leads to a slow increase in chromosomal abnormalities, fuelling the evolutionary capacity of tumour cells. Prostate tumours with HR deficiencies feature an increased frequency of genomic alterations, and are effectively a subtype of CRPC (Mateo et al., 2017; Wu et al., 2018).

In these cases it is possible to exploit 'synthetic lethality', an approach that involves inhibiting single strand repair via the PARP genes. Impaired PARP allows for unrepaired single-strands to persist through replication, and potentially convert into double-strand breaks. In healthy cells the homologous repair pathway would repair these, but HR-deficient tumours rapidly accumulate double strand breaks, overload NHEJ and lead to severe cytotoxic consequences. Side effects of PARP therapy are relatively modest, particularly when compared to conventional chemotherapeutics (Fong et al., 2009; Lord and Ashworth, 2017).

Initially developed with a focus on breast and ovarian cancers, trials of PARP inhibition in advanced prostate tumours have shown promise: around a third of unselected individuals with CRPC respond to therapy, with respondents closely overlapping with those bearing HR pathway aberrations (Mateo et al., 2015; Ramakrishnan Geethakumari et al., 2017; Underhill et al., 2011). Identifying this patient subset in the overall CRPC population is therefore of particular importance in order to deliver this therapy with maximum impact.

1.3.2 Mismatch Repair Deficiency and Immunotherapy

A proportion of CRPCs (~3-10%) bear genomic deficiencies in mismatch repair (MMR) genes, which leads to elevated numbers of single nucleotide alterations (SNAs) and a micro-satellite instability (MSI) phenotype (abnormal length variability of short nucleotide repeats) (Nava Rodrigues et al., 2018; Pritchard et al., 2014). Higher quantities of SNAs in tumours, while providing a benefit to the tumour in terms of genetic variability, provoke the immune system through

generation of aberrant proteins (termed ‘neo-antigens’), and tumours develop methods to ‘escape’ immune system targeting by downregulating immune signalling pathways (Messerschmidt et al., 2016).

Deficiency in MMR genes may therefore sensitise tumours to immunotherapeutic agents, which attempt to reactivate the immune system and enhance the ability to detect tumour-derived aberrations. While initial trials of immunotherapy in CRPC were not successful, this could be due to trial designs not incorporating patient selection, as the number of MMR-deficient CRPC cases is relatively modest (Beer et al., 2017; Kwon et al., 2014). The drug Pembrolizumab was approved by the FDA for use in all solid malignancies with MMR defects in 2017, and inhibits the programmed cell-death pathway protein (PD-1). Upregulation of PD-1 is utilised by tumour cells to escape detection by the immune system. Emerging data suggests that susceptible (ie. MMR-deficient) tumours respond well (Hansen et al., 2018; Le et al., 2015). It is critical to identify and characterise this subclass of CRPC, particularly in the context of sequencing assays as these can be used to screen for the genomic signatures of MMR deficiency.

1.3.3 Others

The methyltransferase EZH2 and other KDM-family methylases are frequently deregulated in CRPCs (>10%), associating with a more aggressive phenotype and worse prognosis (Melling et al., 2015). These may present a treatment opportunity for CRPC patients, with small molecule inhibitors exhibiting promising preclinical results and phase-1 trials currently ongoing (Bai et al., 2019; Kim and Roberts, 2016; McAllister et al., 2016; Wee et al., 2014). A recent study has also shown that CHD1-deleted and SPOP-mutated prostate present a clinically relevant CRPC subtype, with increased AR-pathway activity and a heightened sensitivity to androgen blockade (Boysen et al., 2018).

Aberrations across the PI3K-AKT pathway have been shown to promote a castration resistant phenotype, and associate with a poor prognosis in various CRPC contexts (Carver et al., 2011; Ferraldeschi et al., 2015; Haffner et al., 2013; Mulholland et al., 2011; Reid et al., 2010). There is some evidence that targeting this hyperactive pathway in concert with other agents may have therapeutic value, although studies are currently on-going (Baselga et al., 2012;

de Bono et al., 2019). Similar to DNA-repair defective tumours, it is likely that targeting the PI3K-PTEN-AKT axis will only bear fruit in tumours which rely on it.

Despite widespread usage in advanced prostate cancer, taxane chemotherapies are given without patient genomic clinical stratification. Two licensed variants are commonly used in prostate cancer, docetaxel and cabazitaxel, which are similar in both structure and method of action (binding β-tubulin, microtubule stabilisation and cell-cycle arrest), although Cabazitaxel has been developed to overcome docetaxel-specific resistances. Several gene pathways have been linked to taxane resistance, which may be innate or acquired following therapy, including multidrug resistance proteins, modulation of tubulin expression, and mitotic spindle assembly (Bumbaca and Li, 2018; Fitzpatrick and de Wit, 2014; Galletti et al., 2017).

Identifying actionable subclasses of prostate cancer increasingly requires the integration of multiple assays, particularly involving genomic analysis.

1.4 Sequencing for Cancer Genomics

1.4.1 Biopsies

Classifying an individual based on their tumour genome requires tumour material (for this project the focus was primarily on DNA, rather than RNA). This can be acquired in a number of ways, and the methods offer various advantages and disadvantages. In the context of this work three main sources of genomic data will be featured: conventional tumour tissue biopsies, blood plasma, and circulating tumour cells, with the goal in each case being the acquisition of high-quality tumour DNA.

All diagnoses of prostate cancer require a prostate biopsy followed by histopathological examination, and individuals with more advanced tumours may undergo repeat biopsies of metastatic sites. Solid tissue biopsies are cheap to acquire, offer large amounts of genomic material, and are straightforward to perform. However, there are limitations to solid tissue biopsies, particularly: a) the intermingling of healthy tissue and tumour tissue (so-called ‘tumour purity’); and b) the fact that the biopsy procedure accesses only a small proportion of the tumour burden in the patient. “Liquid biopsies” offer ‘patient-friendlier’ alternatives

- serial using blood draws, commonly performed and less invasive, which can be readily repeated for easier longitudinal sampling.

Short fragments of tumour DNA have been observed circulating in the blood plasma of individuals with cancers, the origin of which has been reported to be necrotic or apoptotic cancer cells (Aucamp et al., 2018). This cell-free DNA (cfDNA) offers a potential insight into multiple tumour clones rather than the limited area of a single needle-core biopsy, but suffers from highly variable tumour purity and highly fragmented DNA. This presents a significant challenge in deconvoluting data from multiple tumour clones (Wan et al., 2017). Concentrations of cfDNA have also been linked to overall burden of disease, and offer some clinical utility in terms of monitoring treatment in CRPC patients (Goodall et al., 2017; Mehra et al., 2018).

Another application of liquid biopsies involves extracting circulating tumour cells (CTCs), a common feature of many cancers and a key driver of metastasis, isolating them and performing genomic analyses on each cell separately. They are found in the peripheral blood of local and metastatic cancers (including ~90% of CRPCs), and detection and quantification of CTCs has been shown to offer prognostic utility (de Bono et al., 2008; Punnoose et al., 2015). Extracting single cells, while time consuming, more costly and requiring sophisticated technology, allows for a high-resolution examination of tumour genomes present in the individual (Lambros et al., 2018). Low levels of input materials currently limit single-cell studies however, as acquiring sufficient cells of high enough quality for sequencing is challenging. In addition, analytical methods for making full use of single-cell data are not yet well established.

1.4.2 Sequencing

Following acquisition of genomic material for analysis (from biopsy, plasma, or circulating cells), DNA sequences need to be determined, a process most commonly performed by short-read sequencing, using workflows supplied by Illumina (Muzzey et al., 2015). Short-read sequencing is a ‘next-generation’ sequencing technology, a successor to the ‘first generation’ Sanger sequencing that drove the human genome project. Some ‘third-generation’ technologies are emerging, although for this work, due to ease of access and the prevalence of optimised approaches only short-read sequencing was used.

There are varying strategies for preparing samples for sequencing: essentially, input genomic material is fragmented (and size-selected) into short single-stranded sequences (often around 500 nucleotides in length) and molecularly tagged with nucleotide barcodes for downstream identification (Rizzo and Buck, 2012). This ‘library’ of tagged DNA fragments is then amplified using the polymerase chain reaction (PCR) to generate numerous copies of each fragment. Protocols for library generation range from those designed to access DNA from the entire human genome to those which provide data only on a set of pre-defined regions. While whole-genome sequencing (WGS) uses all available DNA templates for library preparation and amplification, targeted sequencing protocols selectively amplify DNA sequences that match certain criteria - ie. sequence similarity for genes or regions of interest in the genome, effectively ‘targeting’ those genomic regions.

The library is then loaded onto the DNA sequencer, where each strand of DNA is bound to a fixed position on a slide, and washes of fluorescently tagged artificial nucleotides are introduced; the complementary binding of specific nucleotides (ie. A, C, T, or G) identified by emission of differing fluorescent wavelengths (Muzzey et al., 2015). The fluorescent tags are then removed, and the process is repeated in order to extend this new complementary strand. After a set of sequencing cycles (commonly 100-150, although there are some varying approaches), the record of nucleotides at each strand position is reported as a single sequence - termed a ‘read’.

1.4.3 Genome Analysis

Sequenced reads are computationally ‘aligned’ to the human reference genome (a dataset derived originally from the one published by the human genome project), a process which assigns a chromosomal coordinate location for each read. Analysing sequencing data requires the application of specialised algorithms: the human genome is around 3 billion nucleotides in length, insurmountably huge for humans to remember, but relatively accessible for computers as a ~3 Gigabyte (Gb) text file. Software tools have been developed to aid this task; essentially consisting of a vast pattern-matching problem. The key to iterating rapidly through these files is the construction of indexes, for both the reference genome and the sample data in question.

The first genome alignment tools were derivatives of methods used for analysing protein sequence similarity: hash-table methods using the seed-extension approach. These rely on testing short stretches of each read (seeds) at numerous places across the genome, then extending the nucleotides either side and repeatedly testing the similarity of the read to the sequence. While accurate, these approaches suffer from long processing time, and for most short-read alignment tasks, are increasingly supplanted by algorithms based on the Burrows-Wheeler Transform (BWT) (Bao et al., 2011). The BWT is a data compression tool that is reversible and incurs no data loss during the process, and ideally suited to both storing NGS data (character sequences featuring many repeated characters) and performing many fast searches of the character string in question. Many current tools (such as BWA and BOWTIE) rely on the BWT to reduce the memory requirements and overall running time of short-read sequence alignment tasks: the first release of BWA was around ten times faster than pre-existing hash-table aligners (Bao et al., 2011; Canzar and Salzberg, 2017; Li and Durbin, 2009).

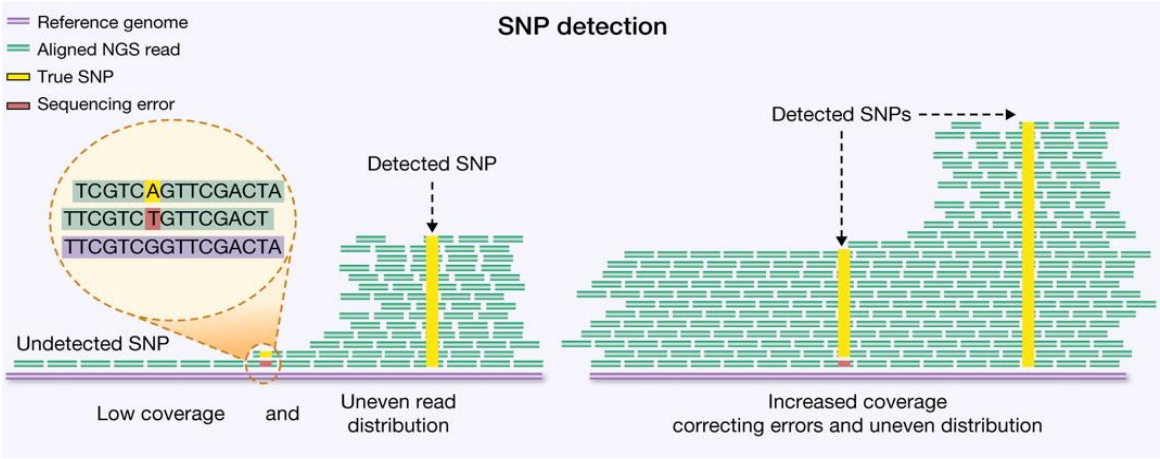
A limitation of next-generation sequencing is the presence of repetitive regions in the human genome - comprising approximately 50% of the genome (Treangen and Salzberg, 2011). Aligning short reads accurately to repetitive regions (comprised, for example, entirely of a short pattern <100 nucleotides in length repeated numerous times) is very challenging (or impossible) and therefore limits the downstream ability to detect mutations that may be present in these regions. Conversely, some reads may have multiple viable matches in the genome (multi-mapped reads) due to sequence similarity between genes. New technologies such as linked-read and long read sequencing may offer solutions to these problematic areas in the genome, although adoption of these methods is not yet widespread.

Another key concept in sequence alignment, and downstream analyses, is read coverage (or ‘read depth’). A reference base covered by one mapped read would have a read depth of one; a position covered by ten reads would have a depth of ten. Areas with higher coverage (ie. more successfully mapped reads) have a higher capacity to call mutations accurately (Figure 1-6). Sequenced samples may be described by the average coverage achieved for the whole genome – for example, a mean coverage of ten reads would be listed as 10X. Variation in

coverage can be due to technical biases (for example, introduced during library preparation), poorly mapped reads (due to repetitive regions) or due to the presence of somatic structural alterations, such as copy number changes. The term ‘breadth of coverage’ also is sometimes used to describe the proportion of the genome covered.

Figure 1-6 Varying depth of coverage in NGS.

Reads (green) are aligned to the reference genome (purple). Reads are not generated in an unbiased manner,



and some areas in the genome are challenging to align reads to: therefore the coverage of aligned reads varies throughout. The ability to detect variants changes depending on sequencing depth (Rizzo and Buck, 2012).

1.5 Genome Informatics

Before mutations (and other aberrations) can be identified in sequenced samples, data has to be pre-processed to remove biases and artefacts from the data. As PCR is commonly used to amplify DNA (and thus, the input signal for sequencing), marking of identical reads as is performed to reduce issues associated with quantifying unique events (Xu, 2018). For downstream detection of small insertions and deletions, local realignment of poorly-mapping reads is often performed prior to further processing, although modern software tools are increasingly incorporating this into downstream variant calling (Van der Auwera et al., 2013).

1.5.1 Single Nucleotide Variants

When considering mutations in tumour biology, it is critical to differentiate between inherited (so-called ‘germline’) variants, and those arising spontaneously in tumours (somatic variants).

A ‘mutation’ is typically considered a single nucleotide alteration (SNA) or a small insertion or deletion (indel) of several bases. The goals of somatic mutation-calling software tools are to: a) identify nucleotides in mapped reads that do not match the reference genome sequence, b) delineate the extent of the alteration, and c) in tumour samples, determine whether the variant is somatic or germline. In addition, resolving mutations in the context of noise is a challenging task: with sufficient read depth all mutations could in theory be observed, but mutations have to be disentangled from sequencing errors (Xu, 2018).

The ‘allele frequency’ (proportion of reads featuring the variant out of the total number of reads mapped to that site) is a key concept in examining mutations. Germline variants may be observed in heterozygous or homozygous states (a single-allele variation or one found in both alleles of a gene) - and therefore, theoretically, at frequencies of 50% or 100%. Somatic variants frequently occur at much lower frequencies, as biopsies often consist of mixed subpopulations of tumour cells (which may not all share the same alterations) and normal tissues. In theory, all somatic variants are heterozygous as the chance of acquiring mutations of the same type at the same position in both alleles is extremely unlikely. However, copy-number changes may impact the zygosity of alleles substantially. Allele frequencies are therefore both a key source of information about the tumour, and a limitation imposed upon the analysis, as they can be affected by changes in tumour ploidy and other chromosomal abnormalities – such as copy-number changes, which are discussed in section 1.5.2.

Identifying somatic events requires the ability to distinguish from germline mutations present in every individual. For maximum specificity, most protocols for tumour mutation calling require the presence of a same-patient matched germline sample, sequenced in the same manner. These patient-matched ‘normals’ are used by variant calling software as a baseline from which to classify events occurring in the tumour sample, ie. whether detected mutations are inherited germline variants or spontaneous somatic events.

Some variant calling software apply separate heuristics to the potential variants found in the tumour and normal, using statistical tests at each variant position to test the chance of being a real ‘non-reference’ site, before using the matched normal sample to filter out calls found in the germline data (Muzzey et al., 2015; Xu, 2018). The comparison of reads featuring non-reference variants found in the

germline sample to the non-reference variants of the somatic sample is key. Modern variant-calling tools often attempt to find a statistical model for the combination of unique genotypes observed at each nucleotide position, with a variety of approaches employed. For example, joint genotype analyses uses Bayes' rule to estimate the probabilities of the genotypes, with prior probabilities supplied by previously quantified values such as the genome-wide rate of germline variants. A key concept employed by several widely used tools (FreeBayes, Platypus, MuTect2 and HaplotypeCaller, among others) is that of a realigned haplotype: identifying 'active' regions (ie. regions containing variants), generating a candidate set of possible combinations of alleles in a de Bruijn-like graph, and then examining the read support that each allele can draw from the aligned reads in that region. This local realignment also improves detection of indels, which may be the true cause of a cluster of closely located SNAs.

Not all somatic alterations are equal. Only a few may be true driver events, while many will be so-called passenger alterations, with modest impacts on tumour fitness (Vogelstein et al., 2013). Identifying the relative importance of detected mutations - in the context of each unique tumour - remains a challenge in clinical genomics (Greenman et al., 2007).

Two main approaches are available for researchers to classify and explore mutations. Databases such as ClinVar and COSMIC catalogue mutations with clinical associations; the variants included are based on supporting evidence such as identification in clinical trials and experimental data (Forbes et al., 2017; Landrum et al., 2018). These are inevitably incomplete: not all genes are studied to the same degree, so popular (or more important, depending on perspective) oncogenes and tumour suppressor genes that are well characterised in the literature are proportionally well characterised in variant databases, and less well known (or less commonly altered) genes are less so. Annotation of mutations with associations from such databases is a core technique for judging the relevance; the function of a specific mutation may be unknown, but if we observe it at an elevated frequency in cancers, then it may be important.

The other primary method for annotating SNVs for functional impact relies on examining the sequence context in which the mutation has occurred and whether it lies within an area of the genome that codes for a protein (exonic regions contribute to under 5% of the human genome); although not all exonic

nucleotides are equally crucial to normal protein function. Exonic SNAs can change the subsequent protein sequence, either solely at that amino acid (‘missense’ mutations), or at all following amino acids until the next stop codon (‘frame-shift’ mutations). Others aberrantly convert the amino acid to be a stop codon, truncating any further transcription (‘nonsense’ mutations). Due to redundancy in the amino acid system, some SNAs in protein coding regions may result in no change to the protein structure (‘synonymous’ mutations). There is an implicit ordering in the severity of these mutational consequences – frame-shift and nonsense SNAs are likely to severely limit the function of the expressed protein, while synonymous mutations are generally considered to have mild (or non-existent) effects (Forbes et al., 2017; Li et al., 2017).

Extending this conceptual framework are tools that aim to predict the impact of non-synonymous SNAs (any variant which changes the amino acid sequence), by integrating rules about protein structure with sequence conservation data (Tang and Thomas, 2016). Changes to amino acids have downstream structural consequences for expressed proteins, and estimating the change in free energy between the folded and unfolded states of the mutant and wild-type proteins is key to understanding the consequences of the mutation. Numerous tools have been proposed to examine this, but generally require high-quality 3d protein structures to be available, simulate folding of nearby side chains, and incorporate statistical models that summarise the interactions of atomic forces (Compiani and Capriotti, 2013; Tang and Thomas, 2016). The extent to which specific amino acid sequences are ‘conserved’ between species is also highly informative for identifying mutations with unusual downstream effects, and compete favourably with structural modelling approaches (Ramensky et al., 2002). Emerging methods such as machine learning and meta-prediction add benefits over existing tools – for example, the PON-P software considers the output from several predictors spanning sequence conservation, structural stability prediction, and machine learning classification to make a final prediction (Olatubosun et al., 2012).

1.5.2 Copy Number Changes

In addition to smaller variants within genes (variants under 50 nucleotides in length), tumours also acquire broader abnormalities in chromosome structure. Humans have two copies of each autosome (non-sex chromosomes) - changes

from this are termed copy-number alterations (CNAs). Copy-number changes are frequently present as germline variants - some estimates place the proportion of the genome affected by an inherited CNA to be around 5%, with more events clustered around telomeric and centromeric regions (Zarrei et al., 2015). More drastic copy-number alterations occur frequently in prostate tumours, often alongside whole-genome duplications, and are enriched in tumours with DNA-repair deficiencies (Wu et al., 2018).

Next generation sequencing allows for the detection of copy number aberrations (CNAs), and several approaches have been established (Zare et al., 2017). The most common method relies on assessing the depth of aligned reads across the genome at a set of regions (bins). Typically a bias-correction step is applied; as discussed previously, read alignment is challenged by repetitive and low-complexity regions, and in addition the presence of GC-rich regions can skew the distribution of aligned reads (Benjamini and Speed, 2012).

The read depth is then normalised by the total number of reads in that sample, and compared to a reference data set - usually derived from a matched healthy tissue sample or a pool of unmatched healthy samples - to generate tumour-reference coverage ratios at each position. Theoretically, regions with a copy-number change will exhibit a shift in the resulting ratio, for example a tetraploid chromosome (4 copies) in a tumour will be overrepresented in terms of normalised read counts compared to a standard diploid chromosome (2 copies) in the germline reference.

Copy ratios are ‘segmented’ - a denoising step - by merging bins into regions of similar values. This step also effectively converts the data into a format that more closely mimics tumour biology, as CNAs in tumours are frequently large events spanning thousands (or tens of thousands) of nucleotides, rather than single locus events. If segments exhibit significant divergence from the reference this can indicate changes in the copy-number state at that region. Several algorithms are currently used for copy number segmentation, with different underlying mechanics but similar goals (Liu et al., 2013). Circular Binary Segmentation (CBS) functions by circularising the genome and performing an initial segmentation. It then iterates over the available data, adjusting segments to minimise intra-segment variability (Olshen et al., 2004). Hidden Markov Model

(HMM) based tools rely on the ability of HMMs to identify discrete states from data, and map these states to integer copy numbers.

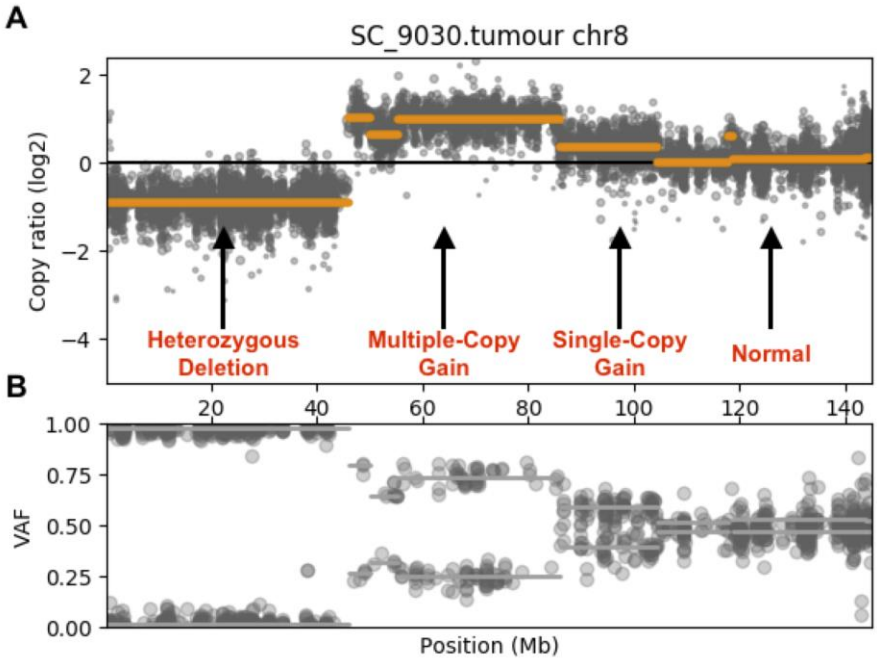


Figure 1-7 Segmentation of Coverage Ratios and Allele Frequencies

Copy number profiling of Chromosome 8 in a tumour sample analysed with whole-exome sequencing. Coverage data (A) and the allele frequencies of heterozygous germline alleles (B) illustrate changes in copy state throughout the chromosome.

Copy number information can also be extracted from the allele frequencies of germline mutations present in the data (Figure 1-7). Theoretically, the genotypes of a germline variant (in a diploid cell) can be considered: A/A (homozygous, for the ‘A’ allele), A/B (heterozygous, ‘A’ and ‘B’ alleles), B/A (also heterozygous), or B/B (homozygous, for the ‘B’ allele). Following variant calling the allele frequencies for these variants, assuming that ‘A’ is the wild-type allele and ‘B’ is the mutant, would be: 0, 0.5, 0.5 and 1, respectively. In a genomic region bearing a copy-number change, the frequencies of the heterozygous (A/B, B/A) variants may exhibit a shift away from 0.5: the direction and extent of this shift can be highly informative. Joint segmentation of allele frequencies with coverage ratios are critical to the calculation of allele-specific copy numbers, in addition to identification of copy-neutral copy changes which cannot be detected from read coverage alone.

Aside from focusing on the impact of single genes impacted by copy number changes, recent work has shown that the genome-wide burden of copy number alterations affecting a tumour is highly indicative of overall disease progression, and associates with prognosis (Hieronymus et al., 2018). The proportion of the genome affected by a CNA is likely a reflection of a) the age of the tumour, with additional alterations accumulating over time, or (and possibly in addition to) b) the presence of some pro-CNA phenotype, such as deficient homologous recombination DNA repair.

1.5.3 Systemic Genomic Data Sets

In addition to detection of specific pathogenic events (for example, a mutation or copy-number event affecting a tumour suppressor gene), an emerging trend is the study of genome-wide biomarkers. A high frequency of somatic mutations (ie. tumour mutational burden or TMB) is a pan-cancer indicator of poor prognosis, and is associated with increased rates of tumour-specific antigen production through aberrant protein expression (Owada-Ozaki et al., 2018; Samstein et al., 2019). These 'neoantigens' have been shown to provoke the immune system, and tumours that persist with a high neoantigen load are postulated to have a method to avoid immune surveillance – and are sensitive to immune checkpoint inhibition (Le et al., 2015; Messerschmidt et al., 2016). A recent study shows that a high number of indels correspond to a greater generation of tumour-specific neoantigens than single nucleotide variants (SNVs), and are likely to represent a highly immunogenic class of mutation (Turajlic et al., 2017).

Studying the aggregate of mutations in a tumour offers an additional approach: mutational signatures. Specific classes of passenger mutations occurring at high frequencies have been linked to certain sources of DNA damage, such as smoking or UV light (Alexandrov et al., 2013). Biological processes including MMR deficiency have also been linked to mutation signatures (Van Hoeck et al., 2019). Patterns of mutational signatures can therefore be used to characterise tumours.

Distinct from small events such as indels or SNVs, the quantity of copy-number aberrations (CNAs) in the genome has been demonstrated to associate with patient outcome in prostate tumours, independent from other prognostic factors such as PSA (Hieronymus et al., 2014, 2018). The distribution of CNAs has been

linked to various types of DNA-repair deficiency, and may indicate specific therapeutic pathways (Wu et al., 2018).

1.6 Biomarkers

When linking genomic (or experimental) data with clinical data, the identification and validation of potential markers for clinical outcome is of great interest. Studies may attempt to test the ability of a specific biomarker to inform on: changes in survival ('prognostic' biomarkers), effect a change in drug response ('predictive' biomarkers), other important endpoints ('surrogate' biomarkers), or changes in disease status following treatment ('response' biomarkers). To explore these putative biomarkers requires – in addition to the key assay in question – a detailed description of additional factors that characterise the study. This includes information about the patient cohort and selection critieria, the sample collection and assay methods, and downstream statistical methods. Detailed suggestions can be found in several articles by Altman and Sauerbrei (Altman et al., 2012; Sauerbrei et al., 2018). Exploratory studies – that aim to generate hypotheses to support further prospective work – also benefit from heeding these guidelines. While categorically proving the ability of certain genomic features to perform as clinical biomarkers is out of the scope of this thesis, I have attempted to work within these suggested frameworks.

2 Thesis Goals

2.1 Overall Hypotheses

- Genomic and molecular characterization of mCRPC can stratify individuals for well-established (anti-androgens, taxane chemotherapy) and emerging therapies (PARP- inhibitors, immunotherapies).
- Liquid biopsies can be used to dynamically monitor molecular features that predict resistance to therapy.
- Linking genotypes with novel phenotypic markers will characterize the development of CRPC and offer insights into the tumour.

2.2 Overall Aims

- Optimise and analytically validate methodology for robust copy-number calling from targeted and whole genome data. Test this in both biopsy data and plasma and show that key CRPC aberrations can be identified.
- Use clinical outcome data to identify putative mechanisms of resistance to chemotherapy in a large cohort of individuals treated with taxanes and with cell-free DNA targeted sequencing available.
- Apply whole-genome copy number profiling to investigate CRPC clonal structure and intra-patient heterogeneity by analysing single cells.

3 Classification of DNA-Repair Deficient CRPC Tumours from Targeted NGS

3.1 Introduction

3.1.1 Specific Hypotheses

- Targeted panels offer cost effective profiling of key genes, and are commonly used in clinical research (particularly for mutation calling) but have not been well established for precision medicine classification of prostate cancers.
- A large proportion (30-40%) of CRPCs bear a defect in DNA repair pathways such as HR and MMR and these are identifiable from genomic screening.
- Different techniques have to be leveraged to detected HR and MMR defects, including copy-number and micro-satellite profiling.

3.1.2 Specific Aims

- Optimise copy-number profiling from targeted-panel NGS data.
- Validate these data with orthogonal methods and external data sources.
- Optimise and apply MSI-scoring software to targeted-panel data.
- Show that key HR and MMR pathway defects can be identified using these approaches.

3.1.3 Research in Context

Recent studies have shown that 20% to 30% of prostate tumours bear defects in homologous DNA repair (HR) genes that render them sensitive to treatment with PARP inhibition, and that these are commonly mediated through gene deletions in these tumour suppressors (Goodall et al., 2017; Mateo et al., 2015). Individuals with HR defects have been shown to be highly sensitive to treatment with PARP gene inhibitors such as Olaparib (Mateo et al., 2015).

Despite the predominance of copy-number events in CRPC, emerging data have also highlighted the presence of defective mismatch repair (MMR) and

subsequent elevation of mutation rates in a small subset (5-10%) of mCRPC (Pritchard et al., 2014). These patients may benefit from immune-checkpoint blockade such as PD-1 inhibition, which aim to highlight tumours to the immune system (Graff et al., 2016). Taken together, HR and MMR deficient tumours may represent up to ~40% of all CRPC tumours, and therefore offer a subset of patients a precision medicine approach.

High-throughput, inexpensive, multiplex assays for molecular stratification are needed to guide therapeutic choices and support clinical trial design. Although high-coverage whole-exome or genome sequencing can reliably assess tumour genomics, cost and bioinformatic demands currently restrict their routine clinical implementation to stratify patients for treatment. Targeted next-generation sequencing (NGS) represents an opportunity for implementing genomics in clinical practice with the advantages of a rapid turnaround time, lower cost, and the ability to concurrently analyse multiple genes with a limited bioinformatic analysis burden. Existing approaches to call germline and somatic SNVs/SNAs and short indels from targeted sequencing data are well established and have led to a number of FDA biomarker-dependent drug approvals, including for the PARP inhibitor olaparib, in the past decade (Fong et al., 2009; Ledermann et al., 2014).

In the case of PARP inhibition for HR defective prostate cancers, screening for tumours with these DNA repair defects relies on detecting gene copy number changes and point mutations in the relevant genes (Mateo et al., 2015). Some data also suggests that applying a broader heuristic based on an overall burden of copy number changes (an HRD score) can also predict loss of these genes (Telli et al., 2016). Defining MMR deficient tumours in prostate cancer from NGS is similar, with deleterious variants in MMR pathway genes limiting function. In addition the number of somatic mutations (a ‘tumour mutational burden’ or TMB), and the presence of indels and multiple low-frequency variants at short nucleotide repeats (termed microsatellite instability, or MSI) may both be informative for detecting insufficient MMR (Pritchard et al., 2014; Salipante et al., 2014).

In this study I sought to develop and test computational approaches to classify HR and MMR deficient tumours by comparing copy number calls from targeted sequencing panel data with orthologous approaches. I initially focused on identifying gene copy number alterations (CNAs), based on targeted amplicon

sequencing of a focused gene panel dedicated to identifying DNA repair defects. I used a cohort of mCRPC patients to explore the utility of this approach. This targeted gene panel focused on the study of 113 genes important to DNA repair machinery, including HR genes such as: *BRCA2, BRCA1, ATM, PALB2*, and *CDK12* as well as MMR genes including MSH2/6 and MLH1. Other important potentially actionable prostate cancer genes were also studied, including *AR, SPOP, PIK3CA, PTEN, AKT1, AKT2*, and *MYC.*

I also tested a bioinformatic method for the detection of microsatellite instability (MSI), applied to samples sequenced using the same targeted amplicon panel and compare these data with a set of available orthogonal assays including immunohistochemistry and somatic mutation load. Currently, targeted panel NGS is not well established for the detection of MSI or MMR defects, so I also sought to validate these data with whole-exome sequencing. Accurately classifying mCRPC for MMR-defective and HR-defective states will expand the range of therapies available for advanced diseases, and enable precision medicine approaches in a disease for which these have traditionally been lacking.

3.2 Methods

3.2.1 Patient Selection and Tissue Sample Preparation

For the targeted-panel copy number analysis, patient tumour (biopsies) and germline (buccal swab and saliva) samples were collected as part of the Royal Marsden ethics committee–approved CCR2472 protocol. Samples were collected, annotated, stored, and reviewed as described previously (Mateo et al., 2015). This patient cohort is described in Supplemental Table 10-5, with subsets for the copy-number (homologous repair/HR) and mismatch repair (MMR) samples also shown. Briefly, biopsies were paraffin embedded, DNA from biopsies was extracted using the QIAamp DNA FFPE Tissue Kit (Qiagen), Quant-iT Picogreen High-Sensitivity double-stranded DNA (dsDNA) Assay Kits (Invitrogen, Thermo Fisher Scientific) and QC check by FFPE QC Kit (Illumina) as described previously(Nava Rodrigues et al., 2018; Ong et al., 2014; Seed et al., 2017). For whole exome analyses (as part of validating copy number events or MSI-scores), samples were collated from the SU2C/PCF Prostate Cancer Dream Team sample cohort.

3.2.2 Targeted Panel Design

A customized Generead v2 DNAseq Panel (Qiagen) panel was used for library construction; the exonic regions of 113 genes were included in the targeted panel, these being selected for being potentially actionable and/or involved in DNA damage repair processes and/or in prostate carcinogenesis (Supplemental Table 10-1). The panel covered 564Kb of the human genome, and each of the 113 genes was covered by an average of 18.69 targeted regions (SD, 13.64).

3.2.3 Targeted Panel Sequencing

Targeted amplicon sequencing was performed using the Illumina MiSeq platform following the manufacturers' protocol. FASTQ files were generated using the Illumina MiSeq Reporter (v2.5.1.3) (Ong et al., 2014). Reads were aligned to the human reference genome (hg19) using BWA. Whole-exome sequencing (WES) was performed at the University of Michigan (Ann Arbor, MI) as described previously (Robinson et al., 2015). Sequence alignments were performed using Burrows-Wheeler Aligner (BWA) tools and the Genome Analysis Toolkit (GATK)

variant annotator by the QIAGEN GeneRead Targeted Exon Enrichment Panel Data Analysis portal.

FastQC (v0.11.2) and Samtools (v0.1.19) were used to assess sequencing quality. Targeted sequencing samples were excluded from copy number analysis on the basis of: insufficient (<0.5 million) total read counts, low (<95%) percentage of properly paired reads, and low (<99.99%) percentage of on-target reads. Exome sequencing samples were rejected if the FastQC per-base quality score for 75% of the reads was less than Q20 over the first 80 bases, and alignment quality was monitored with Picard, as described previously (Robinson et al., 2015). I additionally processed aligned sequencing reads (.bam files) to: a) match the UCSC chromosome style (remove “chr” notation for chromosomes), b) sort and index the files, using Samtools.

3.2.4 Array Comparative Genomic Hybridization

Tumour DNA from prostate cancer patients and male reference DNA from Agilent were amplified using Sigma WGA2 Kit (Sigma-Aldrich) according to the manufacturer's recommendations, and the amplified DNA was quantified by the Qubit fluorometric quantitation method (Thermo Fisher Scientific). Amplified tumour DNA (500 ng) was then fluorescently labelled with Cy5, and male reference DNA labelled with Cy3, using the SureTag Complete DNA Labeling Kit (Agilent Technologies). Labeled DNA was subsequently hybridized utilizing the Agilent SurePrint G3 Human CGH Microarray Kit, 4 × 180 K according to the manufacturer's instructions. The slides were then scanned and analysed using the CytoGenomics Software v 4.0.3.12 (Agilent Technologies). To compare array comparative genomic hybridization (aCGH) and NGS results, we used the $\log_2$ ratio of the aCGH segments that overlap with genes present in the NGS panel. Genes not part of a copy-altered segment were included in the analysis by using the mean $\log_2$ ratio for probes, if the value was no greater than 0.25. Genes on the X-chromosome were excluded from comparison with results from the male reference–based panel data as a female reference was used for aCGH.

3.2.5 RB1 Gene Fluorescence In-Situ Hybridization

RB1 FISH was performed using a standard formalin-fixed, paraffin-embedded (FFPE) hybridization method on 3-µm FFPE tissue slices adjacent to hematoxylin and eosin sections that were confirmed to contain a minimum of 50 intact cells

(Ferraldeschi et al., 2015). Briefly, RB1 status was determined using Vysis LSI 13 RB1 (13q14) probe (catalog # 08L65-020; Abbott Laboratories) and a reference probe Vysis 13q34 (catalog # 05N34-020; Abbott Laboratories). Nineteen z-stacks were used to assess cell status.

3.2.6 MMR Protein IHC Staining and Scoring

FFPE samples were cut in 3-µm sections onto charged glass slides. Immunostaining was performed with antibodies against MSH2, MSH6, MLH1, PMS2 (M3639, clone FE11, 1:50; M3646, clone EP49, 1:500; M3640, clone ES05, 1:100; and M3647, clone EP51, 1:100; Dako, Agilent Technologies), and PD-L1 (catalog 13684, clone E1L3N, 1:200; Cell Signaling Technology). Heat-induced antigen retrieval was performed using Tris-EDTA buffer, pH 8.1, in a MenaPath Antigen Access Unit at 125°C for 2 minutes to detect MSH2, MSH6, MLH1, and PMS2 or microwave oven for 18 minutes for the detection of PD-L1. Endogenous peroxidase was inactivated using 3% H2O2, and nonspecific staining was blocked using protein block serum-free solution (X0909, Dako, Agilent Technologies). Detection by diaminobenzidine reaction was performed using a Dako REAL EnVision Detection System (K5007, Agilent Technologies). Scoring of MSH2, MSH6, MLH1, and PMS2 was achieved by segregating cases in a binary fashion between positives and negatives using College of American Pathologists criteria for biomarker reporting in colorectal carcinomas (Bartley et al., 2014).

3.2.7 PCR MSI Assay

DNA (1 ng) was amplified using the MSI Analysis System (Promega) according to the manufacturer's protocol. The MSI Analysis System is composed of 7 fluorescently labeled microsatellites including 5 mononucleotide repeat markers (BAT-25, BAT-26, NR-21, NR-24, and MONO-27) for detecting MSI and 2 pentanucleotide repeat markers (Penta C and Penta D) for sample identification. The PCR products were run in an ABI 3730 DNA Analyzer and subsequently analysed using GeneMapper 4.0 software (Thermo Fisher Scientific). Samples with microsatellite instability in 2 or more loci were defined as MSI-high, whereas samples with a single locus were defined as MSI-low; samples were microsatellite-stable (MSS) if no instability at any of the loci tested was detected. For the purpose of statistical analysis, cases were dichotomized between MSI-high and MSS/MSI-low.

3.2.8 NGS Copy Number Calling

The software CNVkit (v0.3.5) was used to analyse sequencing coverage and copy number in the aligned sequencing reads from targeted amplicon sequencing of tumour and germline samples (Talevich et al., 2016). Sequencing coverage of targeted regions in age-matched male germline samples (not same-patient) was assessed and used to create pooled reference data that incorporated the technical variability at each covered region. Regions that were poorly amplified or mapped were masked from further analysis. The analyses of germline samples were also performed with CNVkit to confirm reference sample quality. The read depths of tumour samples were accessed, normalized (corrected for GC content, target footprint size and spacing, and repetitive sequences), and individually compared with the reference, and the circular binary segmentation (CBS) algorithm was used to infer copy number segments (Olshen et al., 2004). Very small copy changes (≤3 bins) were treated as artefactual. Poor samples were also identified through high (>0.8) values of the segment interquartile range (IQR), a metric for the spread of bin-level copy ratios within each segment following application of the CBS algorithm. Copy number segments were aligned to genes, and regions bearing a log2 ratio of at least ±0.4 were identified as suggestive of shallow deletions or gains. Segments with $\log_2$ <−1.2 were classified as deep deletions, and those with log2 >2 were classified as amplifications. Experimental noise was identified as a $\log_2$ ratio SD of approximately 0.2.

WES CNA data were extracted from a previously reported Su2C-PCF dataset (Robinson et al., 2015). Briefly, log2 ratios were derived on a per-gene basis from circular-binary segmented, Lowess-normalized log2 transformed coverage ratio between each tumour and matched normal sample. The same reference genome (hg19) was used for both targeted and whole-exome copy-number datasets.

3.2.9 NGS MSI Assay

MSINGS software was used to score samples for an MSI-like phenotype by assessing targeted next-generation DNA-sequencing data (Salipante et al., 2014). I applied a customised approach to modify the system to the computing environment available, and the structure of input data. For input to the algorithm,

aligned sequencing reads (.bam files) were reformatted using the SAMtools (v1.3) multiple pileup (mpileup) utility with the following parameters: max depth (-d) 10000, count anomalous read pairs (-A), recalculate base alignment quality (-E), and zero-depth nucleotide positions then excluded (Li et al., 2009). The pileup file (.mpileup) was then used as an input to Varscan readcounts (v2.3.7) to perform a basic variant-calling step, calculating the reads that support different alleles at each position in the list of selected potential msi-loci. A minimum base quality of 10 was used as a filter. The resulting data was then used as an input to MSINGS (Salipante et al., 2014).

The MSINGS algorithm functions by (a) identifying possible DNA repeat regions; (b) examining the frequencies of these alleles bearing varying repeat lengths; and (c) comparing these values with a baseline reference from MMR-intact specimens. Loci with a significantly higher number of variable alleles are flagged. The output ("MSI-score") is the frequency of flagged loci that feature evidence of multiple low-frequency alleles, out of all possible loci in the targeted regions. The 113-gene targeted amplicon panel data contained 3,214 possible loci.

Prior to running the main algorithm on test samples, I prepared the reference dataset by using a combination of both tumour and normal samples, which were analysed concurrently to produce an enriched reference set of 698 loci. This is a small departure from the standard MSINGS methodology, which typically uses only germline (normal) samples for generation of a reference data set. The goal of this alteration was to focus on detection of tumour-derived microsatellites.

3.2.10 NGS Somatic Mutation Load

Targeted-panel and whole-exome sequencing mutation calling was performed with standard procedures as previously described (Ong et al., 2014; Robinson et al., 2015). Variants were first processed by removing low-quality mutations: haplotype score >200, read depth <60, low mapping quality <40, multi-allelic variants, indels, and previously identified false-positives. To identify somatic variants in the absence of germline data (in the targeted NGS cohort), variant calls were first annotated with Oncotator (v.1.8.0), and then excluded when: the allele frequency was greater than 5% across all analysed samples, or if the variant was present in 2 or more public mutation databases (dbSNP, ExAC or 1000Genomes) (1000 Genomes Project Consortium et al., 2015; Lek et al., 2016;

Sherry et al., 2001). Mutations present in the COSMIC database a minimum of 10 times were rescued, and included in the final output (Forbes et al., 2017).

3.2.11 NGS Mutation Signatures

Mutation signatures from whole-exome sequencing were collated from somatic mutations and the respective nearby nucleotide contexts. Two methods were applied to identify signatures from the R-packages SomaticSignatures (v.2.6.1.) and NMF (Gaujoux and Seoighe, 2010; Gehring et al., 2015). Overlapping signatures from both tools that overlapped previously identified MMR-deficient signatures were used to compare with other methods.

3.2.12 Study Approval

All SU2C/PCF study individuals provided written informed consent for collection of fresh tumour biopsies and for comprehensive molecular profiling of tumour and germline samples. All Royal Marsden Hospital patients gave written informed consent and were enrolled in institutional protocols approved by the Royal Marsden NHS Foundation Trust Hospital (London, UK) ethics review committee (reference no. 04/Q0801/60).

3.3 Results

3.3.1 Targeted Panel CNA Calling Identifies Key Events

Freshly collected metastatic CRPC (mCRPC) biopsies (n = 110) were acquired from patients with progressing disease and assessed for copy number changes by sequencing with our targeted panel. Germline DNA samples (buccal swab) were taken from 34 consenting patients to use as a baseline reference and for additional analyses. Tumour biopsies were from: bone (49.1%), lymph nodes (35.5%), liver (9.1%), soft tissue disease (3.6%), and transurethral resection of the prostate (3.6%).

High GC content and repetitive regions can bias accurately assessing the depth of coverage of genes, so I first evaluated CNA reproducibility on a per-gene basis between technical replicates (Olshen et al., 2004; Talevich et al., 2016). Thirteen samples were sequenced in duplicate to assess technical variation; 12 of these duplicate datasets included repeated library preparation from the same DNA extraction, while one involved re-sequencing of the same library twice.

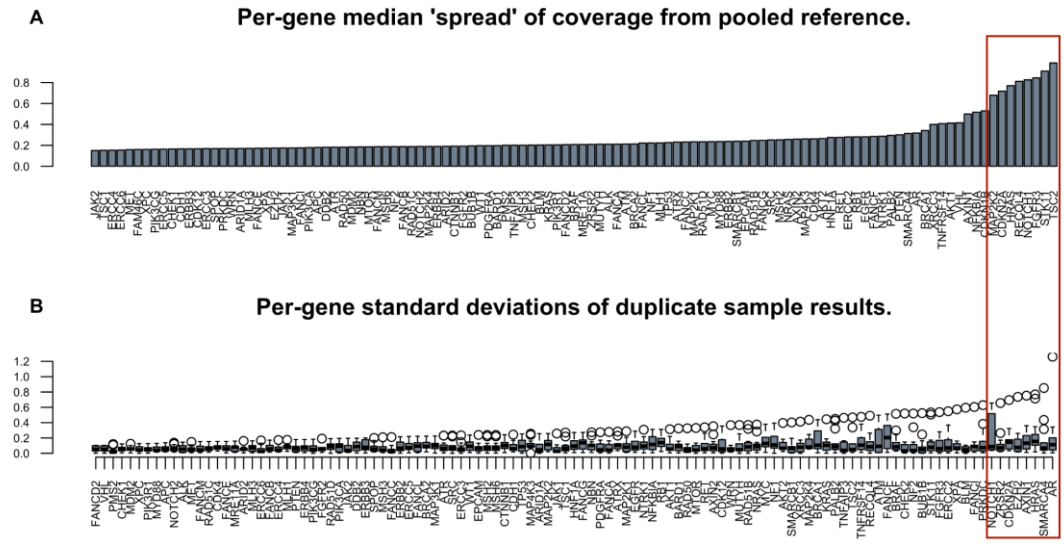


Figure 3-1 Variability of Per-Gene CNA Qualities

Not all genes perform equally well in copy-number analyses. The median spread of observed coverage values in each segmented gene shows that a subset of genes feature substantial variability across the cohort (A). Another subset of genes are common outliers in duplicate sample analysis (B).

While most genes offered good gene-level CNA data, I identified 15 genes that produced less concordant results between replicates, including five small genes

with insufficient number of targeted regions (ie. four or less probes) per gene (*NRAS, NFKBIA, FANCF, FAM46C, CDKN1B*), and 10 (*MAP2K2, CDKN2A, HRAS, RECQL4, NOTCH1, FGFR3, STK11, TSC2, SMARCA4, AXIN1*) that contained repetitive or polymorphic exonic regions that impeded accurate read alignment and possessed high coverage variability (Figure 3-1). These genes were excluded from downstream CNA analyses.

3.3.2 Panel CNA Data is Reproducible

Following removal of the aforementioned genes, CNA estimates between duplicates showed high correlation (Pearson *r* correlation coefficient = 0.92) (Figure 3-2). All genes with deep deletions ($\log_2$ ratio, <−1.2) or amplifications ($\log_2$ ratio > 2) had a similar (at least ±0.4) CNA estimation replicated in the duplicate sample. A total of 95% of samples with a $\log_2$ ratio change of at least ±0.4 had a similar result in the duplicate.

A matched germline sample is commonly used for identifying deviations in coverage ratios, so we sought to compare this with our pooled reference approach. For the pooled reference, we sequenced 34 unmatched germline samples (sequencing performed by J.M. and S.C.) and I assessed the depth of coverage using the CNVkit software. Regions with low variability were identified in the pooled reference and used to weight downstream CNA estimates. This approach produced a highly similar result (overall Pearson *r* = 0.93) when compared with the CNA estimation approach using a matched normal as reference, when matched normal samples were available (Figure 3-2); however, the CNA estimation confidence score (IQR value) was improved when using pooled germline samples (paired *t* test *P* = 0.05). This was in line with other reports that indicated that using pooled germline data as reference offered a robust method to identify gene CNAs (Grasso et al., 2015).

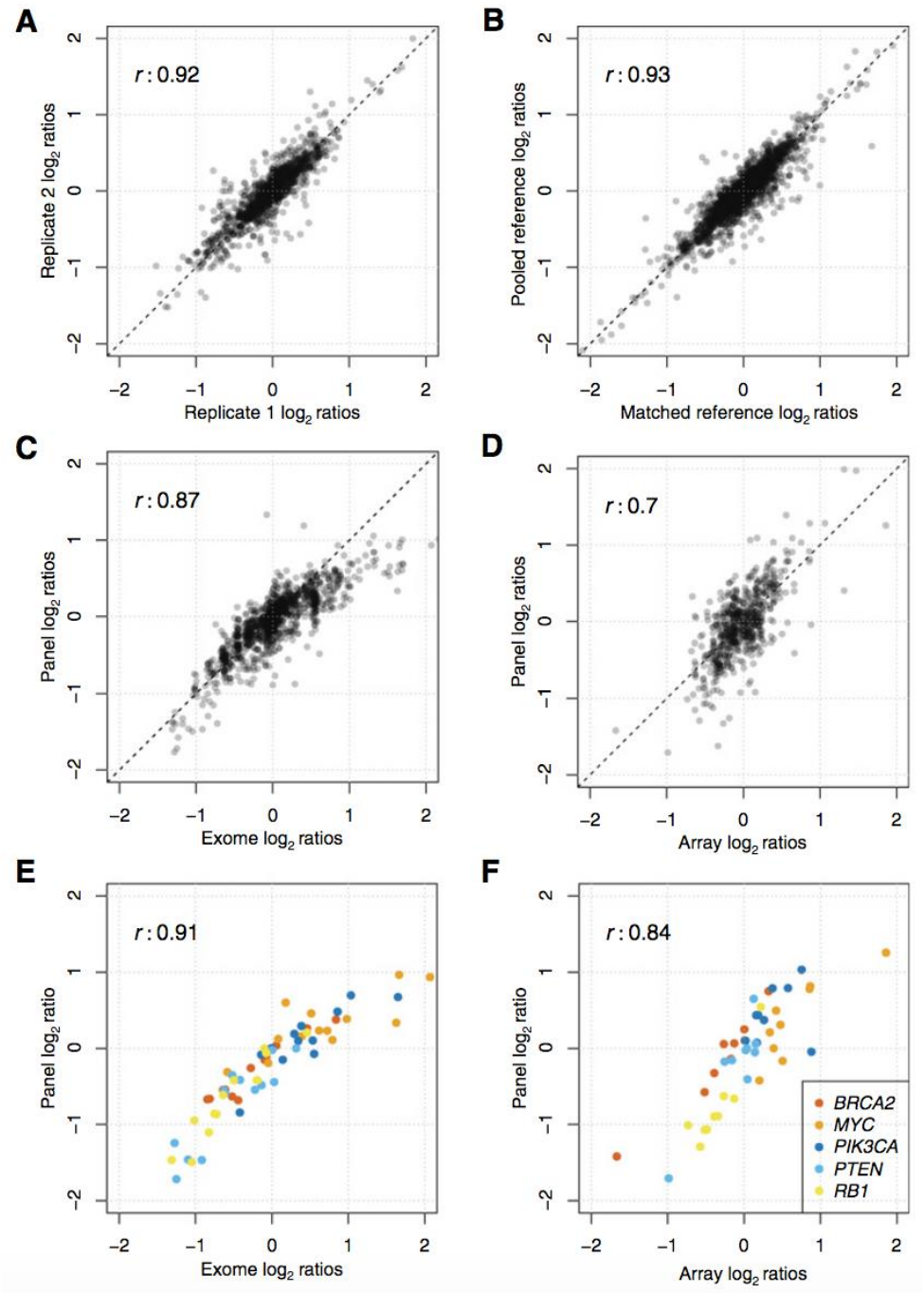


Figure 3-2 Targeted panel CNAs are highly reproducible

Correlation of log2 ratio data from the 113-gene panel between technical replicate samples (n=13) was highly correlated (A). A matched reference performed similarly to an unmatched pooled reference (n=25) (B). Orthogonal validation with (C) exome-sequencing data (n=13) and aCGH (n=9) (D) showed well-correlated results for the evaluable genes in the panel, and high correlations for five key CRPC genes of interest (E, F). Colours in (E) and (F) indicate gene. Pearson *r* values are shown.

Sample	Panel *BRCA2*	Exome *BRCA2*	aCGH *BRCA2*
p.4.1	-0.670	-0.845	
p.5.0	0.262	0.46	
p.7.0	-0.542	-0.611	
p.8.0	-0.152	-0.095	
p.9.0	-0.682	-0.442	
p.13.0	-0.109	-0.074	
p.14.1	-0.663	-0.815	
p.15.0	-0.546	-0.639	
p.18.0	0.036	0.057	
p.20.0	0.377	0.838	
p.22.0	-0.101	-0.101	
p.23.0	-0.259	-0.277	
p.25.0	-0.633	-0.52	
p.39.0	-0.324		-0.389
p.41.0	0.056		-0.264
p.45.0	-0.139		-0.17
p.110.0	-0.576		-0.516
p.69.0	-1.419		-1.665
p.98.0	0.102		0.008
p.102.0	0.065		-0.128
p.104.0	0.75		0.321
p.38.2	0.468		0.003

Figure 3-3 BRCA2 Gene Copy Numbers Are Highly Correlated

Raw gene-level copy number data of the *BRCA2* gene shows closely similar results with well-established methods of WES and aCGH (n=22). Cells are coloured for copy-losses (blues) and copy-gains (reds), with colour saturation reflecting spread of the value away from a neutral copy number (log2 ratio of 0).

I cross-compared copy number $\log_2$ ratios for all evaluable genes on the targeted panel with $\log_2$ ratios generated by exome sequencing (n = 13; Figure 3-2), resulting in a Pearson r correlation coefficient of 0.86. Moreover, all the genes with a $\log_2$ ratio <−1.2 by targeted sequencing, in keeping with putative “deep” deletions, exhibited a similar (at least ±0.4) result in the exome sequencing. For example, one sample bore *AR* amplification ($\log_2$ ratio 5.7 in targeted NGS and 4.91 in WES) alongside *WRN* deletion (−1.53 vs. −1.17) and *ATM* deletion (−0.68 vs. −0.56). We also compared $\log_2$ ratios from aCGH segments (n = 9) with gene copy number estimations by panel NGS among the genes covered by both platforms and found $\log_2$ ratios were also concordant (Figure 3-2), with a Pearson r value of 0.7. I then sought to further validate copy number calls for specific genes of clinical relevance, such as *BRCA2, MYC, PIK3CA, PTEN*, and *RB1* (Figure 3-2). For these genes, I found that the targeted panel produced results highly concordant with exome sequencing and aCGH data, with Pearson r values of 0.91 and 0.84, respectively. All clinically relevant deletions of *BRCA2* in samples with exome or aCGH data were identified in our panel with deletions of varying intensity clearly evident and highly correlated (Figure 3-3).

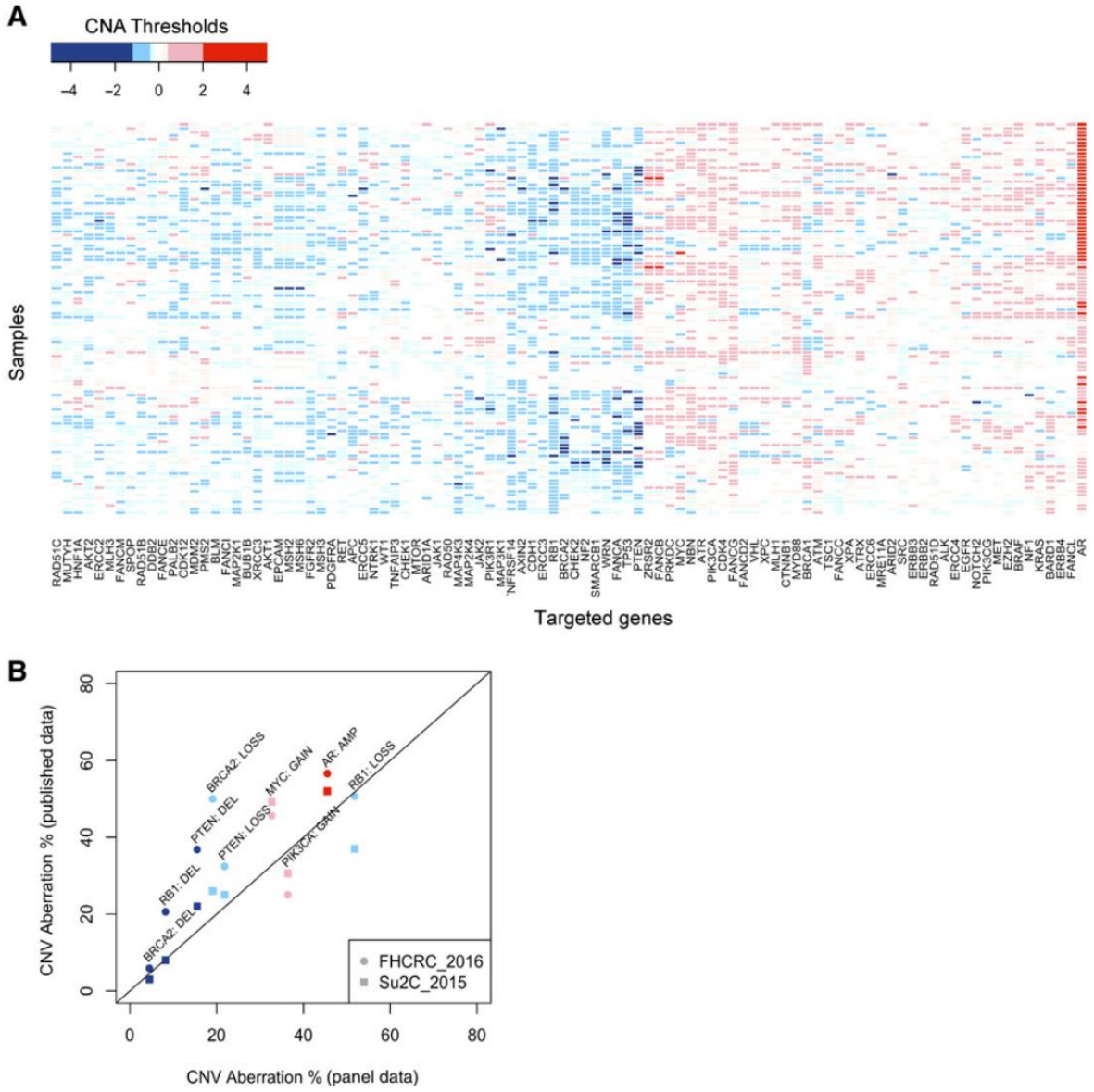


Figure 3-4 Targeted Panel CNAs Match Other Studies

Binned heatmap (A) of per-gene log2 ratio values shows that gene deletions (blues) and amplifications (reds) are detectable in this cohort at expected frequencies. Rows and columns clustered by Ward's method and Euclidean distance for visualisation purposes. The proportion of samples bearing copy changes in these samples is similar to those reported in two recently published datasets, from the Fred Hutchinson Cancer Research Center (circles) and Stand Up to Cancer Dream-Team (squares) studies (B) (Kumar et al., 2016; Robinson et al., 2015).

3.3.3 HR Gene CNA Frequencies Match Other Studies

In addition to validating the reproducibility of the targeted panel data, I sought to show that the prevalence of copy-number alterations in this cohort of CRPC tumours matched that of other datasets.

CNAs of *AR* (amplification; $\log_2 > 2$) were found in 45.5% (50/110) of tumour samples. Broad deletions of chromosome 13 involving the loss of *BRCA2* and *RB1* were detected as previously described with 23 of 26 (88%) samples with any *BRCA2* loss also having some loss of *RB1*, while overall 23 of 66 (34.8%) samples with *RB1* loss had loss of *BRCA2*

Overall, the evaluation of the deep deletions of commonly aberrant CRPC genes revealed results in keeping with previous reports from two recent studies of CRPC genomics; *PTEN* genomic deep deletions were detected in 17 of 110 (15.5%), *BRCA2* deep deletion in 5 of 110 (4.5%), and *RB1* deep deletion in 9 of 110 (8.2%) mCRPC biopsy samples (Kumar et al., 2016; Robinson et al., 2015). In addition, gain of the *MYC* locus was found in 36 of 110 samples (32.7%). The copy number alteration frequencies in this cohort were in line with expected CRPC frequencies (Figure 3-4).

3.3.4 CNA Calling is Limited by Purity and Heterogeneity

To explore how tumour purity altered the capacity of this targeted assay to detect CNAs, I pursued the serial dilution of tumour DNA acquired from a sample estimated by pathology review as having 80% tumour content, with same-patient germline DNA (sequencing performed by J.M. and S.C). This sample harboured a somatic, *BRCA2* homozygous deletion (independently confirmed with WES). The resulting dilutions were sequenced with targeted panel NGS and re-analysed with CNVkit. I found that homozygous deletions were easily distinguishable with tumour content purities >60% and that clonal CNAs were detectable with purity as low as 30%.

To examine the ability of DNA sampling methodology to improve tumour material acquisition, I analysed a sample for which microdissected DNA was available (dissection performed by D.N.R). This sample was originally scored by targeted-sequencing analysis as bearing a borderline *BRCA2* deletion ($\log_2$ ratio of −1.15) and had a pathologist-estimated tumour purity of ∼40%). Re-sequencing the

resulting tumour-enriched DNA sample (purity estimate, ~80%) and re-analysing the CNAs with CNVkit shifted the CNA estimates away from normal, with the $\log_2$ ratio for *BRCA2* dropping to −2.82, an unambiguous homozygous deletion (Figure 3-5).

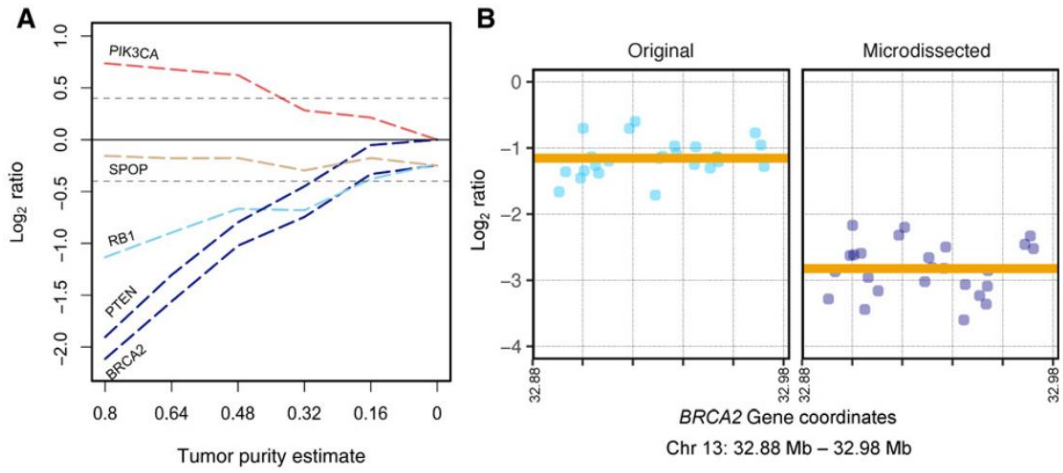


Figure 3-5 Tumour Purity Limits CNA Detection

Serial dilution of a mCRPC tumour sample with matched-patient germline DNA and resequencing (A) highlights the functional limits of copy-number detection. Homozygous deletions (*BRCA2, PTEN* – dark blue), hemizygously deleted (*RB1* – light blue), copy-neutral (*SPOP* – grey), and copy-gained (*PIK3CA* – pink) genes are shown. Enriching tumour DNA by microdissection (B) shows a radical increase in detectable signal, with a shift in exon-level log2 ratios (blue points) and the corresponding assigned segment (orange line) away from 0.

The $\log_2$ copy ratios estimated from sequencing a bulk of cells, as in this protocol, can be biased by tumour sample heterogeneity. To assess this relationship, I compared FISH assays of the *RB1* gene with the targeted panel data (FISH performed by S.M. and scored by D.N.R.), counting the number of cells where each copy number (0 through 3+) was observed (Figure 3-6). I found a significant association between the proportion of cells with 1 or 0 copies of *RB1* detected by FISH (ie. hemizygous or homozygous deletions) and putative copy loss inferred from our targeted sequencing panel (unpaired *t* test $P = 0.02$, $n = 18$, df = 15.884). However, the full complexity of aneuploidy in individual tumor cells is masked by bulk tissue sequencing, which effectively averages the estimated DNA content across all cells in a sample. FISH analyses also revealed that many tumour samples contain a mixture of cells with different copy number changes at the *RB1* locus, in highly heterogeneous cell populations (Figure 3-7).

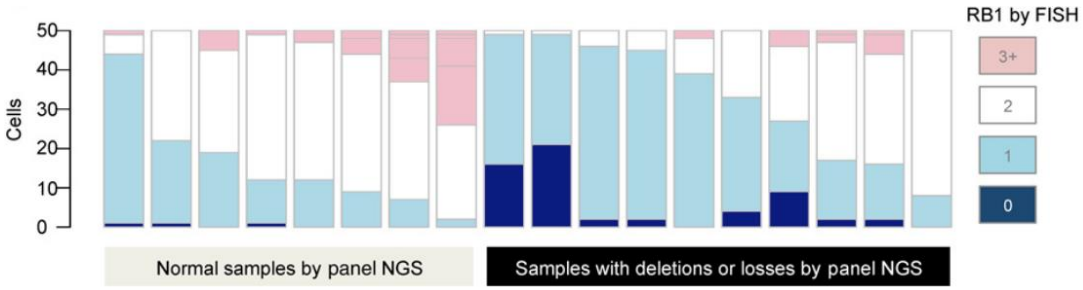


Figure 3-6 Distinguishing copy-number states is limited by tumour heterogeneity

Stacked bar plot illustrating *RB1*-FISH statuses in 18 mCRPC samples, with 50 cells scored per sample. Colours indicate numbers of *RB1* gene detected; dark blue: 0 copies, light blue: 1 copy, white: 2 copies, pink: 3 or more copies. Samples are split by panel-sequenced *RB1* gene deletion status.

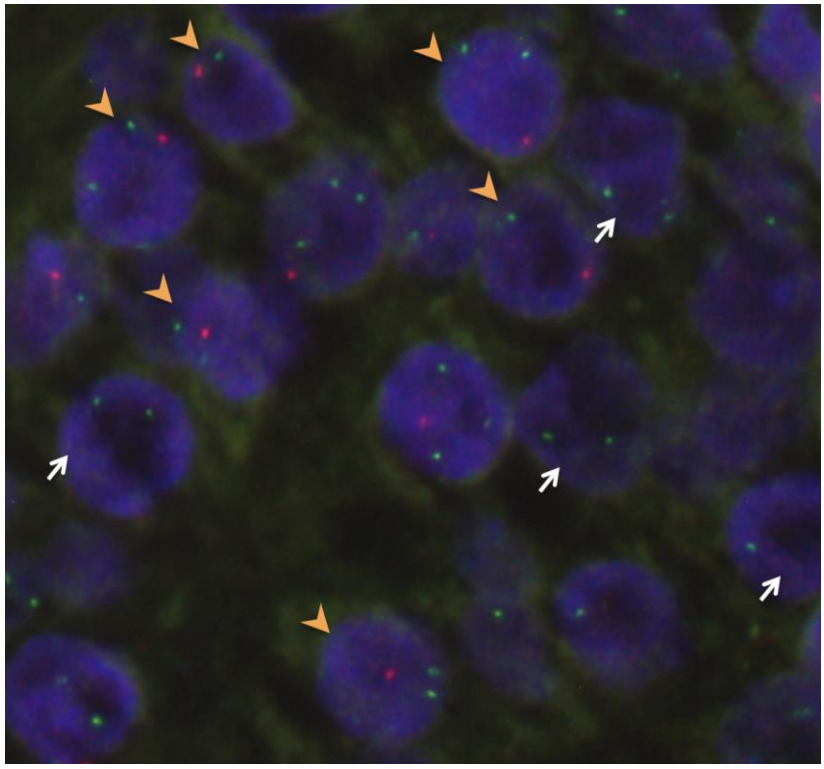


Figure 3-7 RB1-FISH shows adjacent cells with variable copy-states

FISH from sample p.16.0 showing *RB1* (red) and reference (green) showing a mixture of: hemizygously deleted cells (white arrows) and homozygously deleted cells (orange arrows).

3.3.5 MSINGS Identifies MSI-loci in CRPCs

I sought to test the ability of the available targeted panel data to identify MMR-defective tumours by detection of the microsatellite instability phenotype. I applied the MSINGS algorithm to a cohort of 127 mCRPC samples (sequenced with the available targeted panel) that also had orthogonal assays available. MSINGS works by: constructing a position-based summary of all reads mapping

to a set of pre-defined loci, checking for non-reference alleles at these positions, and identifying loci that bear numerous alternative alleles, indicating variable indel lengths. Samples with high numbers of these loci are termed ‘microsatellite unstable’, with a MSINGS score derived from the frequency of unstable to stable loci. The targeted panel featured 698 potentially unstable loci. I modified the MSINGS algorithm to use a reference dataset assembled from both germline and tumour samples, instead of germline alone, to access the maximum number of tumour-specific microsatellites.

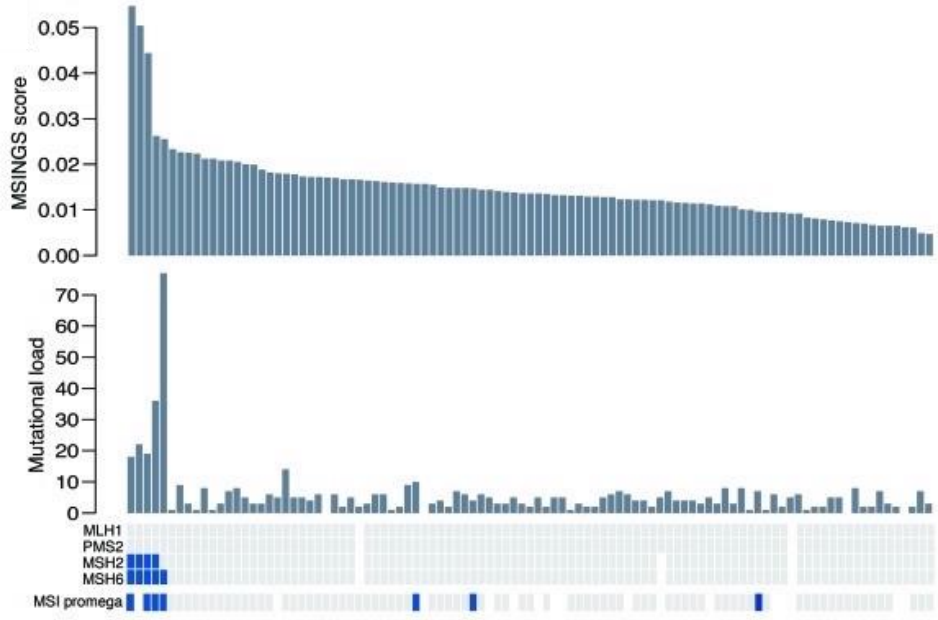


Figure 3-8 Targeted panel NGS MSI-scoring associates with MMR markers

Orthogonal assay validation of NGS-assessed microsatellite instability (MSI) shows that it is well correlated with somatic mutations, MMR IHC and a commercial MSI assay.

Numerous unstable loci were detected in this targeted panel cohort, however no overall pattern of events emerged (Supplemental Figure 10-1). In this cohort, high MSINGS scores were closely aligned with: high somatic mutational load (generated by M.C.), IHC for MMR proteins (prepared by I.F. and D.N.R.), and results from an orthogonal PCR-based assay (run by M.B.L.) (Figure 3-8). However, the overall frequency of possibly MSI-high tumours in this mCRPC cohort was low (~5-8%). A cut-off of 0.0244 with this targeted MSINGS panel had an AUC of 0.79, a sensitivity of 60%, and a specificity of 98% to predict MMR cases defined positive by IHC and/or pcr-MSI (Figure 3-9).

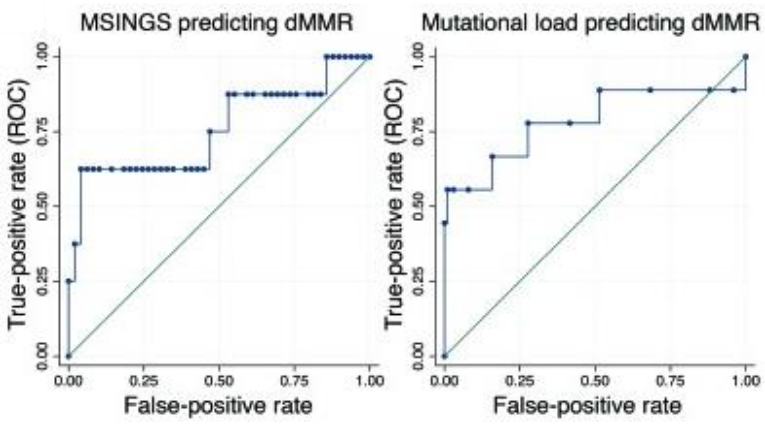


Figure 3-9 Targeted Panel Data is Predictive of MMR Defects

ROC curves of MSINGS and somatic mutational load scores predicting MMR protein loss or pcr-MSI assay positivity. Sensitivity was 60% and 78% respectively, while specificity was 98% and 72%.

3.3.6 Further MSI Validation Confirms Panel NGS Data

To confirm the viability of MSI testing from NGS data in CRPCs, I applied the MSINGS algorithm to a set of whole-exome sequencing samples generated as part of a Su2C/PCF funded effort. These samples (n=254) had targeted sequencing MSI data available, and the scores were correlated ($r = 0.73$, $P < 0.0001$) (Figure 3-10). Mutation signature analyses (performed by W.Y. and an external collaborator D.L.) revealed that samples with high MSINGS scores had a strong correlation with high numbers of mutations linked to MMR-derived signatures (Figure 3-10). COSMIC signatures 6 and 26 were strongly associated with MMR deficient tumours in this cohort (Forbes et al., 2017). Interestingly, several cases in this cohort exhibited high quantities of MMR-like mutations but had a modest MSINGS score and expressed normal quantities of MMR proteins.

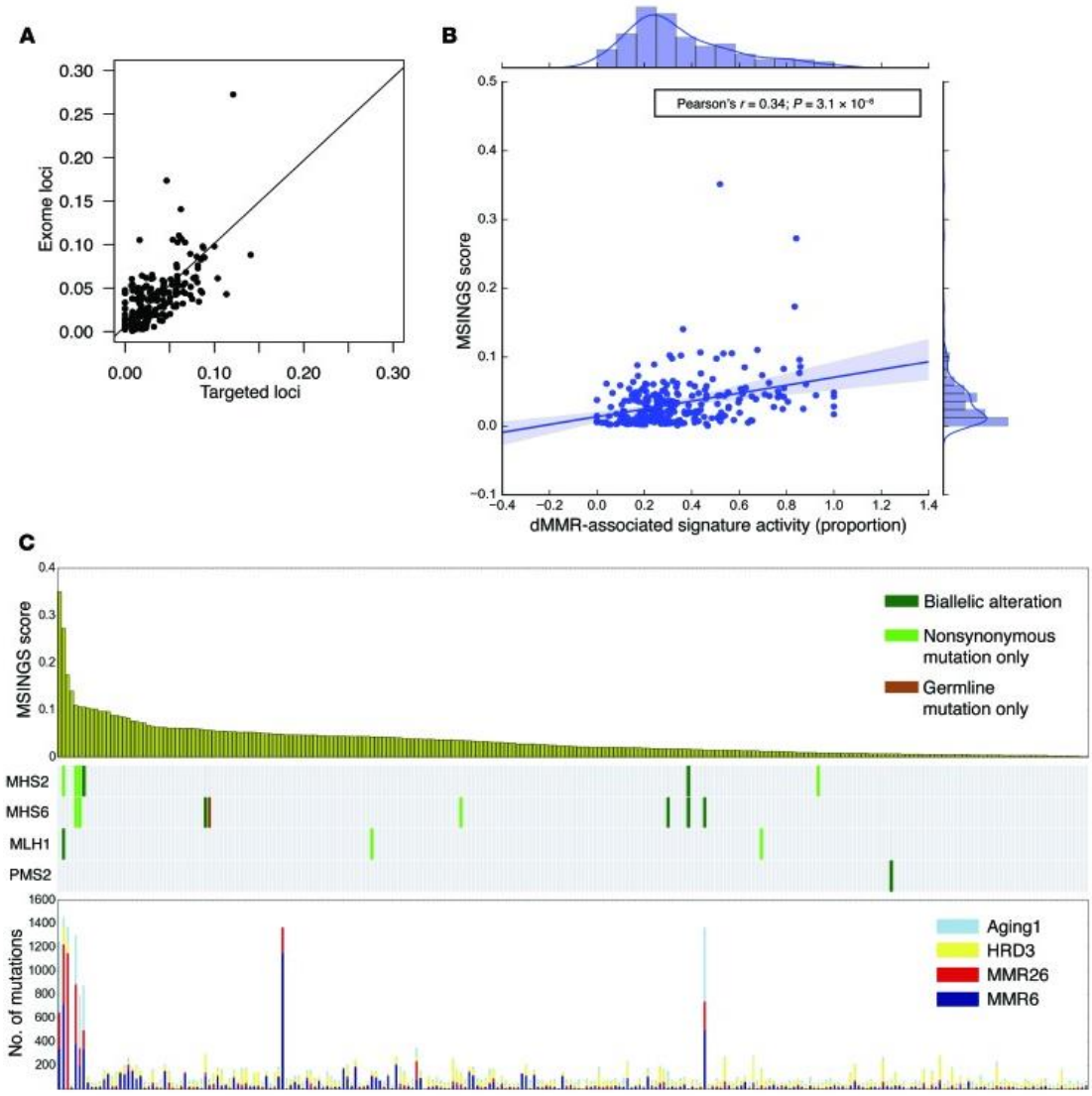


Figure 3-10 Whole-exome sequencing validation of MSI-scoring

Targeted panel and whole exome NGS MSI-scores are well correlated (A) (Pearson's r = 0.73). MSINGS results showed limited correlation with an MMR-deficient gene activity (RNA) signature (B). Whole-exome MSI-scores were associated with both a higher overall count of somatic mutations and MMR-deficient mutational signatures (C).

3.4 Discussion

This chapter details approaches to robustly identify clinically viable copy-number changes in mCRPC biopsies through the application of targeted sequencing. This can support classification of mCRPC tumours as HRD-defective (~30 of cases), primarily through accurately calling losses of key tumour suppressor genes such as *BRCA2*. The results generated by this method are also comparable with widely used assays, including exome sequencing, aCGH, digital droplet PCR, and FISH. I also confirmed that using a pooled reference provides copy-number calls equally accurate to those generated by using matched normal sample, raising the possibility of stratifying patients even in situations where only tumour DNA is available, and also potentially reducing the sequencing cost of large studies.

The gene panel I utilised also covered several key genomic aberrations that are of emerging interest as possible clinically predictive biomarkers that may be actionable in prostate cancer, including loss of *PTEN*, gain of *PIK3CA*, and amplification of *MYC* (de Bono et al., 2019; Crumbaker et al., 2017; Ferraldeschi et al., 2015; Reid et al., 2010; Sarker et al., 2009; Wee et al., 2014). The frequencies of the genomic events that I observed were in line with other recently reported studies of mCRPC copy-number changes (Robinson et al., 2015).

However not all genes performed equally well in this analysis (including *FGFR3* and *SMARCA4*). There are several sources of possible noise: genes are distributed unevenly throughout in the genome, are of variable length, and may feature repetitive or high-GC regions that limit accurate read alignment. In addition, amplicon sequencing (as opposed to capture-based techniques) generates very little off-target reads (reads mapped to regions not targeted in the panel), which could otherwise be used to copy-number calls by providing coverage estimates throughout the genome (Talevich et al., 2016; Zare et al., 2017). Due to varying primer sets used for amplification, different targeted panels may perform differently for CNA estimation, and are likely to require separate validation.

As I show here, the relative purity of tumour cells to normal cells in the biopsied material applies hard limits to detection of CNAs in this technological setting, particularly in the context of tumour heterogeneity. These represent challenges

for all bulk sequencing applications, and I estimate that interpreting CNA data from targeted NGS sequencing of samples with less than 30% tumour purity may be challenging. My data also show that it is possible to mitigate these limitations by modifying sample collection procedures (i.e. microdissection) to enrich for tumour cells prior to sequencing. The manner in which tumour biopsy DNA is collected and enriched is a vital consideration for clinical applications of tumour genomics.

I went on to apply the MSINGS algorithm for identification of the MSI phenotype in a cohort of samples sequenced using the same targeted-amplicon panel. In concert with somatic mutations and copy-number changes in MMR genes, these data show that targeted panel NGS can classify MMR-defective CRPC tumours, which are an emerging subtype of clinical interest in cancer (Nava Rodrigues et al., 2018; Pritchard et al., 2014).

I validated the MSINGS results against orthogonal assays for MMR pathway deficiency, including IHC and PCR-based assays. The targeted-panel data were also highly correlated with whole-exome derived MSINGS results, with an entirely different set of loci considered for analysis. Collating the MSINGS results and somatic mutation load, in addition to specific alterations in MMR genes (such as deletions or mutations in *MLH1* and *MSH2/6*, which were observed in these analyses) can therefore be used to identify a discrete subset of CRPCs (around ~5-10% of the population). While the varying approaches I show here were overall well correlated, there is observable noise, which is related to the challenge of using systemic continuous variables rather than binary decisions. Classifying an individual based on the presence of a single event, such as the presence or absence of a specific mutation, is a simpler task than classifying based on a collection of biomarkers – such as numeric values of mutation load or MSINGS.

Taken together, these results show the validation and clinical application of a accessible next-generation sequencing-based assay capable of categorising a substantial proportion (~40%) of mCRPC tumours into two categories of DNA-repair deficiencies.

4 Prognostic and Predictive Studies of Low-Pass Plasma NGS

4.1 Introduction

4.1.1 Specific Hypotheses

- Taxanes are used to treat mCRPC and other cancers, and tumour genomics studies can identify those that benefit from these drugs.
- Plasma cell-free DNA offers key insights into tumour genomics but is limited by low tumour purity.
- Copy number events are characteristic of CRPC genomes, and studying CNAs in a cohort of individuals treated with taxanes may identify associations with treatment resistance and sensitivity.

4.1.2 Specific Aims

- Use low-pass whole genome sequencing to generate copy-number profiles from cfDNA samples acquired from men with mCRPC treated with taxanes on two prospective phase 3 trials (FIRSTANA and PROSELICA).
- Validate methodology and demonstrate reproducibility and sensitivity of these assays.
- Address tumour purity in these serial cfDNA samples, acquired pre- and post-treatment and assess their utility as markers of response.

4.1.3 Research in Context

- Taxane are commonly used to treat multiple cancer types, with docetaxel and cabazitaxel approved for treating mCRPC.
- No stratification exists for treating prostate tumours with taxanes.
- Taxanes interfere with microtubule function by stabilising tubulin, thereby impacting mitosis and tubulin-dependent interphase interactions.
- There are limited options for monitoring taxane response, and no genomic stratification exists for this class of chemotherapy in prostate tumours.

- I analysed lp-WGS data from plasma cell-free DNA, collected from individuals treated on two trials of taxane therapies, for genome-wide copy number changes.
- Changes in tumour purity and other genomic signals offer an insight into disease progression.

In advanced cancers – including prostate cancer – taxane-based chemotherapies are frequently used. Taxanes exhibit chemotherapeutic activity by binding to microtubules and preventing their disassembly, leading to cell-cycle arrest and resultant apoptosis (Jordan and Wilson, 2004). Taxanes are also thought to have some anti-androgenic properties; having been shown to potentially block nuclear translocation of the microtubule-dependent androgen receptor (AR) (Fitzpatrick and de Wit, 2014). In prostate tumours, Docetaxel was the first to lead to an improvement in overall survival (OS), prostate-specific antigen (PSA) falls and quality of life in mCRPC in two phase III trials (Berthold et al., 2008; Tannock et al., 2004). Cabazitaxel, a semi-synthetic taxane, was devised to overcome the emergence of taxane-resistance and has been shown to be efficacious in the post-docetaxel setting (de Bono et al., 2010). Taxanes are not currently applied in a 'precision medicine' context, as they are a broad-spectrum chemotherapy. Potential sources of stratification for these drugs are therefore of keen interest.

Whilst both of these licensed chemotherapeutic agents result in demonstrable improvements in OS, PSA and quality of life, responses are variable and resistance inevitable. Monitoring response to treatment using available biomarkers, such as changes in prostate-specific antigen (PSA) and radiographic changes in bone disease can poorly reflect taxane anti-tumour activity, and superior clinical variables are much needed. Recent work has shown that mCRPC treatment response can also be monitored using plasma cell-free DNA (cfDNA); this can be both quantitative and qualitative (Annala et al., 2018; Mehra et al., 2018). Using these liquid biopsies to monitor disease has many advantages, and when analysed with next-generation genomics, cfDNA offers key insights into the elements driving an individual's disease (Carreira et al., 2014; Goodall et al., 2017).

Low-pass whole genome sequencing is an effective approach to assess genome-wide copy number events, with advances in technology making this more cost-efficient and allowing for a high-throughput analyses of samples (Ulz et al.,

2016). Aside from gene-level copy number events, other signals from sequencing cfDNA, such as whole genome copy-number burden and average copy-number fragment size have also been shown to be informative (Hieronymus et al., 2014; Wu et al., 2018). Several studies have shown that tumour purity estimated from cfDNA has value as a surrogate marker for tumour burden, although the precise origins and precise clinical use has not been fully established (Aucamp et al., 2018; Nygaard et al., 2014; Valpione et al., 2018).

In this study I sought to apply lp-WGS based copy number profiling to cfDNA samples acquired from prospectively collected clinical trial samples and therefore linked with detailed, high quality clinical data, with the goal of identifying prognostic as well as predictive biomarkers of taxane sensitivity. I accessed survival and response data from individuals treated on two trials of taxane chemotherapy, one taxane naïve and the other taxane pre-treated. In addition to identifying markers of poor prognosis, I pursued the study of tumour purity – ie, the proportion of tumour derived DNA from the overall quantitiy of circulating DNA - in serial samples as a measure of response to treatment.

4.2 Methods

4.2.1 Sample Collection and Processing

Blood samples were collected from patients treated on the FIRSTANA and PROSELICA clinical trials; study designs for both trials have been reported previously (Eisenberger et al., 2017; Mehra et al., 2018). The primary endpoint for both studies was OS, with secondary endpoints including radiological progression free survival (rPFS), PSA response and pharmacokinetics. Sanofi Aventis (S.A.) provided clinical data as a database. Clinical data included baseline measurements for: prostate specific antigen (PSA), lactate dehydrogenase (LDH), haemoglobin (HB), serum albumin (ALB), alkaline phosphatase (ALP), and Eastern Cooperative Oncology Group performance status (ECOG PS). Serial blood samples for cfDNA analyses were collected prospectively in both trials at two baseline timepoints screening (SCR) and Cycle 1 Day 1 (C1), and at Cycle 4 Day 1 (C4) and where possible when a patient came off treatment (End of Study; EOS). Blood was collected in lithium heparin tubes, as when these studies were initiated STRECK^{TM} tubes were not available (BD Vacutainer; BD Biosciences, San Jose, CA, USA). Healthy volunteer plasma (n=10) was collected in STRECK^{TM} tubes and pooled prior to library preparation.

4.2.2 Cell-Free DNA Extraction and Quantification

Cell-free DNA (cfDNA) was isolated from 1-4mls of plasma using QIAamp Circulating Nucleic Acid kit (Qiagen, Hilden, Germany) as per manufacturer's protocol. Of the 60ul eluate, 3ul were used for quantification using the Quant-IT Picogreen HS DNA kit (ThermoFisher, Massachusetts, USA) as per manufacturer's protocol, and concentrations read on a BioTek microplate spectrophotometer at 480ex/520em.

4.2.3 Library Preparation and Sequencing

Following extraction and prior to library preparation the samples were treated with heparinase I (Sigma-Aldrich, Missouri, USA) as per the manufacturer's guidelines. Heparinase concentration was titrated per sample (1U of heparinase to clean 1ug of DNA), vortexed briefly, spun down and incubated at 37C for 2 hours (Taylor, 1997). Following this, low-pass whole genome library preparation was carried out using the QIAGEN QiaSeq FX DNA library kit (96) as per

manufacturer's guidelines. Samples were then sequenced on the Illumina NovaSeq 6000™ (Illumina, San Diego, USA). Technical replicate samples were prepared as above and then sequenced on the Illumina MiSeq (Illumina, San Diego, USA) as previously described (Mateo et al., 2015).

4.2.4 NGS Data Processing

Low-pass WGS sequencing data was converted to Paired-end reads using Illumina bcl2fastq2 software (v.2.17.1.14, Illumina) with the default filters to select sequence reads for downstream analyses. All sequencing reads were aligned to the human genome reference sequence (GRCh37) using the BWA-MEM (v.0.7.12) algorithm, with indels being realigned using the Stampy (v.1.0.28) package to produce aligned read files (.bam format). Quality control checks were performed using Picard CollectWgsMetrics (v.2.8.1) and FASTQC (v.0.11.8.). Samples were excluded if coverage of less than 0.05X was achieved or if overall read quality by FASTQC was classed as a failure.

Aligned reads (.bam) were converted into interval-formatted counts of reads (.wig) using HMMcopy readCounter (v. ABC) with quality filter set to 20 and interval width set to 500kb. Copy-number (CN) modelling of read depth data was performed using ichorCNA (v0.1.0). Transition strength parameters (--txnE and --txnStrength) were set at 0.99999 and 100000 respectively, and the maximum copy number was set to 20 to account for high-level amplifications. Normal contamination (initial values 40% to 90%), ploidy (initial values 2 and 3), and sub-clonal events were modelled. The previously generated 500kb reference coverage and GC-content data supplied with the ichorCNA software were used.

4.2.5 Data Handling and Statistical Analyses

The ichorCNA software provides segmented copy data with both integer copy values and broad classifications, ranging from homozygous deletion to high-level amplification (Adalsteinsson et al., 2017). I further processed the results in R (v3.6.1). Segment copy-states were normalised to the inferred sample ploidy by deducting the ploidy value from the segment copy number, to remove spurious copy-gains due to overall triploid or tetraploid whole-genome ploidy states (common in mCRPC). In addition, homozygous deletions were re-classified as hemizygous if the log2-ratio of the segment was greater than -0.4. In this study, amplifications were defined as having 6 or more copies than the sample ploidy.

Survival analyses were performed using R (v.3.6.1) with the survival (v.2.44-1.1) and survminer (v.0.4.4) packages. Genome and chromosome-level visualisations were produced using karyoploteR (v.1.10.4) and GenomicRanges (v.1.36). Forest plots were produced using the forestmodel (v.0.5.0) package. Cox proportional-hazards models were used for multivariable survival analysis. Kaplan-meier curves and log-rank tests were used for univariable survival comparisons. For logistic regressions, a binomial (logistic) generalised linear model framework was used with the stats package (v.3.6.1). Unpaired tumour purity comparisons were made using the Wilcoxon rank-sum test as implemented in the ggpubr package (v.0.2.1)

4.3 Results

4.3.1 Low-pass WGS Provides High-Quality Copy Number Data

I accessed lp-WGS data that was generated from 528 samples acquired at three time-points pre- and post-taxane treatment (Baseline [SCR and C1], C4, and EOS). The samples were split between the FIRSTANA (FIRS) and PROSELICA (PROS) cohorts as shown below (Table 4-1). FIRSTANA represents individuals who had no prior taxane exposure, while PROSELICA patients had at least one course of prior docetaxel (Supplemental Figure 10-2). This supplied 188 unique patients. Some patients were excluded from analyses due to low coverage metrics (a median coverage less than 0.05X) or missing baseline samples (Figure 4-1). Due to financial limitations, this represents a small proportion of the total number of patients treated on the two trials, and was

	Baseline			
	SCR	C1	C4	EOS
FIRSTANA	101	55	77	57
PROSELICA	84	33	58	63

Table 4-1 Breakdown of cfDNA-lpWGS sample data available for downstream analysis across timepoints.

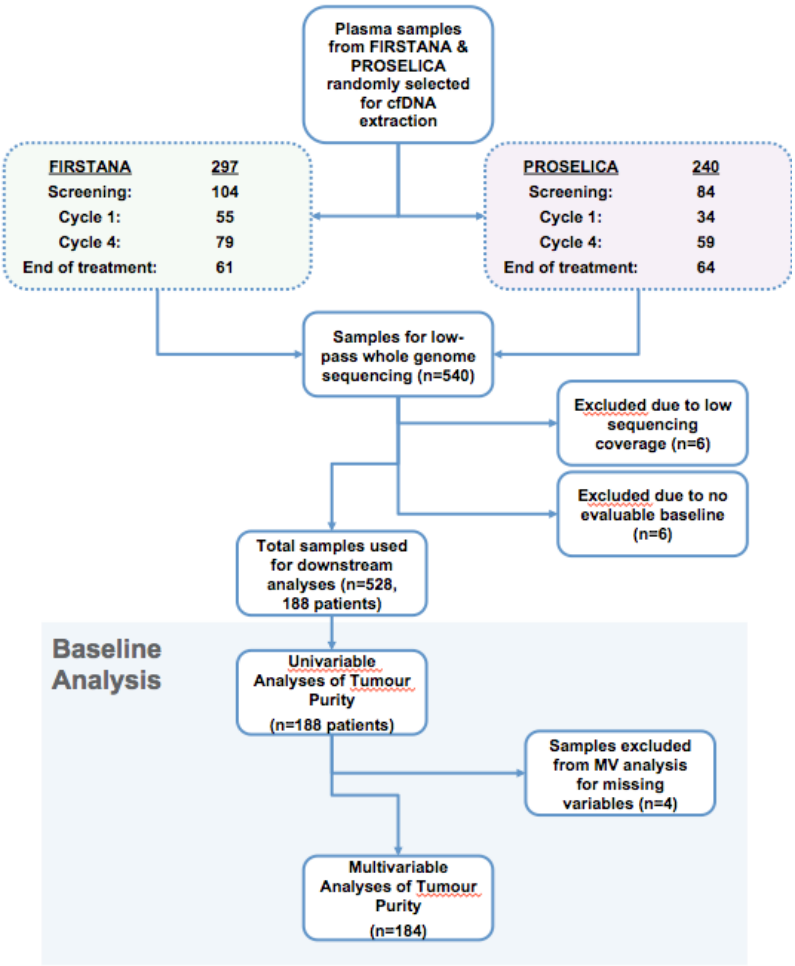


Figure 4-1 Available samples for baseline studies.

Flow chart showing substudy design following collation of samples that passed quality control sequencing. Some samples were excluded from multivariable analyses due to missing clinical variables.

For analysis of baseline genomic data, I preferentially selected screening samples (n=185), but included several (n=3) samples collected at Cycle 1 for individuals who had no screening sequencing data available. Of these baseline samples, 5 were excluded from multivariable analysis due to missing clinical variables (Figure 4-1).

For sequencing (performed by S.S.), average cfDNA input was 10ng, and the average sequencing depth achieved for the sample cohort was ~1.7X (n=528, median = 1.72, SD = 1.23). Copy number frequencies generated by ichorCNA were plotted as log2 ratios along with the assigned segments, showing that profiles consistent with emergent CRPC subtypes are clearly visible including: sparse alterations with highly focal amplifications (including AR), frequent large-scale copy number changes previously described in homologous repair deficient

tumours, and a ‘sawtooth’ profile with numerous small-scale events previously linked to CDK12 deficient tumours (
Figure 4-2, below) (Wu et al., 2018).

4.3.2 The characteristics of this cohort match other studies

I collated the genome-wide copy number profiles of the baseline sample cohort (n=188), and calculated the frequency of aberrations (shallow/hemizygous deletion, deep/homozygous deletion, copy-gain, and amplification) at each 500kb genomic bin (Figure 4-3, below). These results show a similar profile to the publicly available International Stand Up To Cancer/Prostate Cancer Foundation (SU2C/PCF) Prostate Cancer Dream Team dataset of whole exome sequencing performed from 150 mCRPC biopsies (Robinson et al., 2015). The lp-WGS dataset featured events characteristic of CRPC genomes, with frequent amplifications of *AR* (~30%) and *MYC* (~13%), copy-gain of PI3KA (~45%) and copy-loss of *NKX3-1* (~75%), *RB1* (~70%) and *PTEN* (~50%).

Comparing the baseline clinical characteristics of this subset of patients studied herein to the wider trial population (data collated by S.S.) shows some minor differences, despite random selection of patients for analysis. Significant ($p=<0.05$, unpaired t-test) differences were observed in ALB, HB and pain in the FIRS substudy patients compared to the overall FIRS population; while a difference in PROS substudy patients was observed for age, ALP, HB and PSA. No difference in response data, however, was found in either substudy population. Comparing the characteristics of FIRS and PROS patients showed, as previously reported, significant differences in ALP, LDH, HB, and PSA. Response rates were also lower in PROS (38%) compared to FIRS (68%), in line with patients receiving second line chemotherapy (Mehra et al., 2018; Oudard et al., 2017). Similarly, median overall survival for FIRS was longer than PROS (~22 months compared to ~13 months).

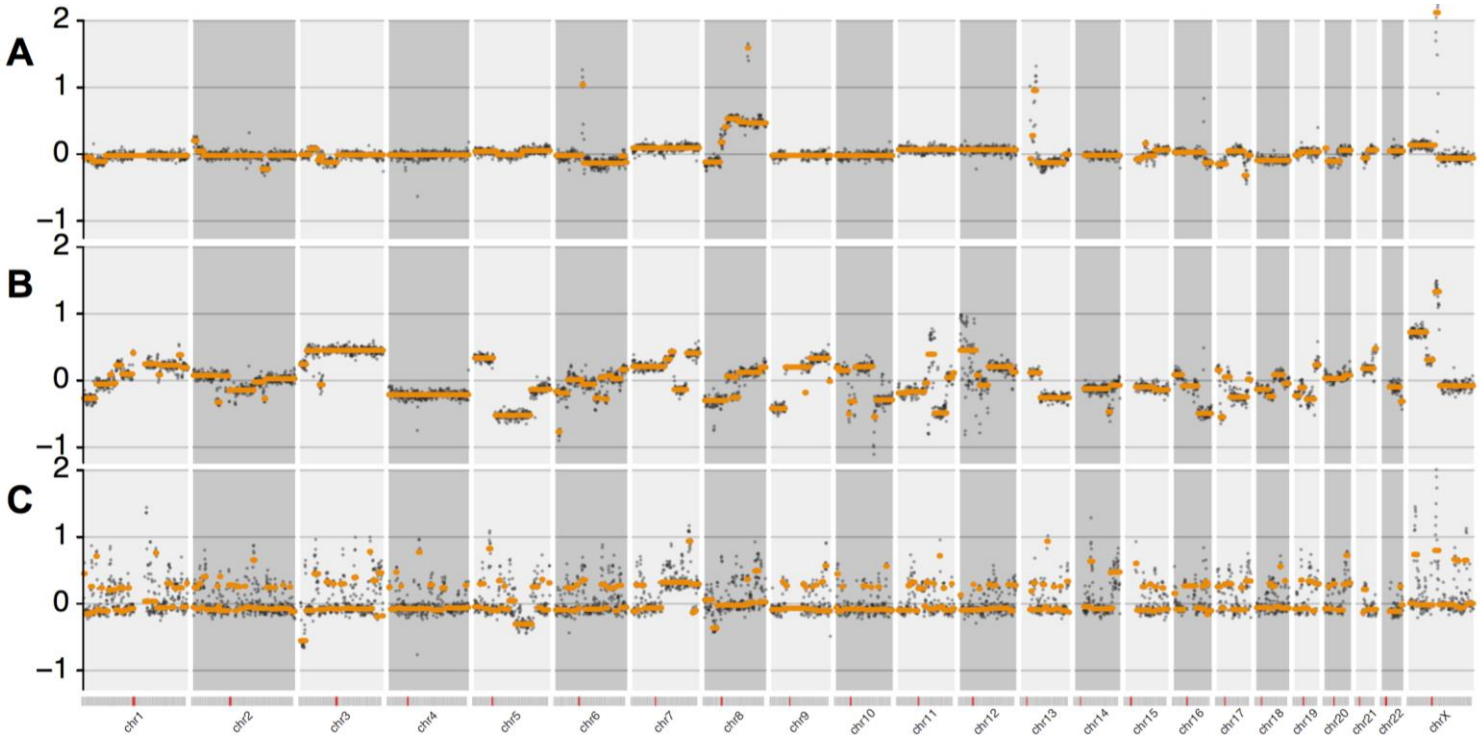


Figure 4-2 Example CNA profiles found in this cohort

Three genome-wide copy-number aberration plots from samples analysed during this study, showing per-bin log2 ratios (grey points) and assigned segments (orange lines). Characteristic prostate cancer events are observable in each case including AR amplification and chromosome 8 losses/gains. The samples include a case with modest alterations and several highly focal amplifications (A), a case with large chromosome-spanning changes (B) and a case exhibiting frequent small alterations throughout the genome (C).

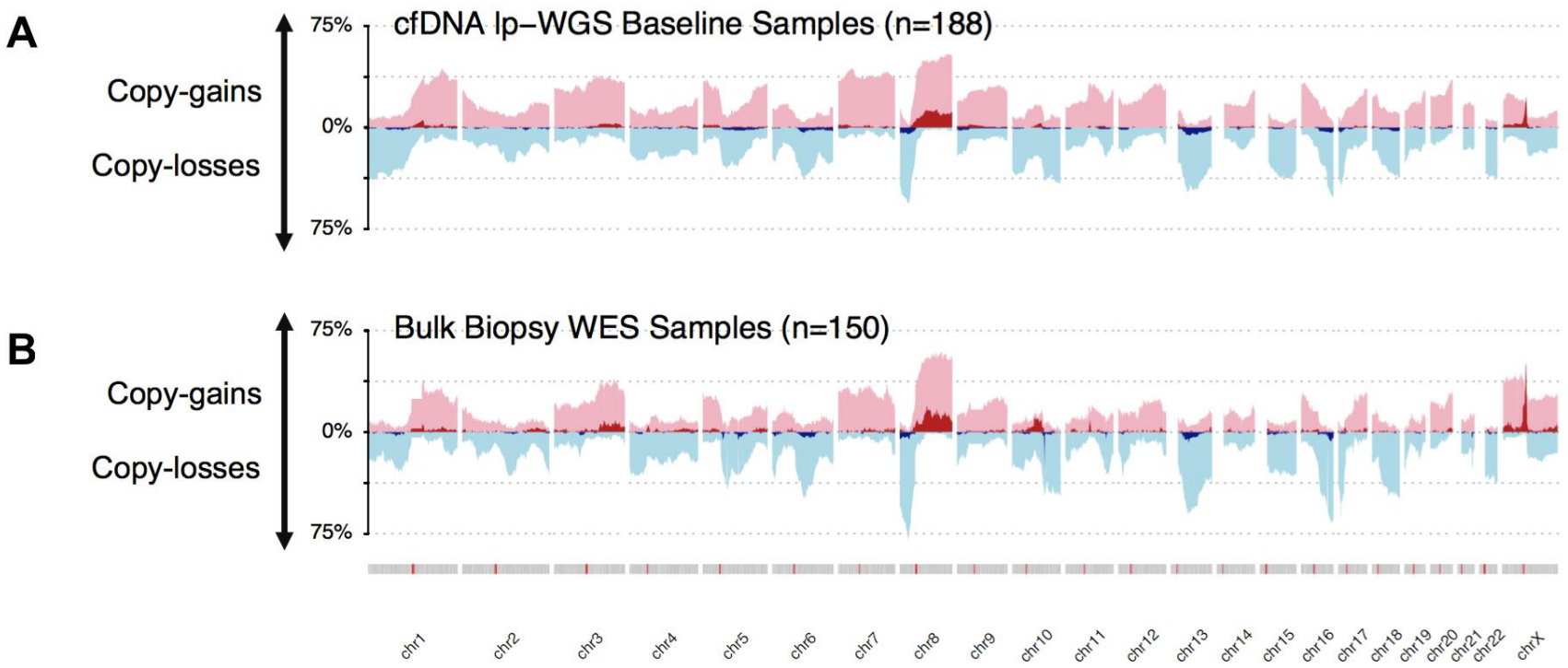


Figure 4-3 Comparing CNA frequencies between studies

The genome-wide low-pass WGS copy number profiles of the baseline cfDNA samples analysed in this study (A) closely match the data from the Stand Up to Cancer mCRPC dataset (B) [ref]. Characteristic alterations are seen (chr8p deletions, chr8q amplifications, chr13 deletions, chrX amplifications).

4.3.3 Biological and Technical Replicates are Closely Correlated

A baseline sample is shown in (Figure 4-4, below) where common CRPC aberrations are clearly visible, including broad arm-level copy events of chromosome 8 with *NKX3-1* (loss) and *MYC* (gain) as well as *RB1* loss on chromosome 13 in the context of an unaltered *BRCA2*. This sample, which had an estimated lp-WGS generated tumour purity of 50% (ie. the proportion of tumour-derived DNA in the total cfDNA concentration) was then diluted using a pool of healthy volunteer cfDNA, fresh sequencing libraries prepared, and resequenced (performed by S.S.) to assess how the same sample performed at decreasing purity levels. The copy number profiles of these technical replicates (diluted to 40%, 20%, 10% and 5% purity) were well correlated, with a high concordance of called copy-number events. Overall, the lp-WGS estimated tumour purity values for these dilutions were strongly correlated (Figure 4-5, below) with the target dilution value (Pearson's r = 0.994). As the concentration of tumour-sample derived DNA drops, the log2-ratios of copy-number segments move closer to 0, the neutral value. This is illustrated by linear modelling of the dilution samples (Figure 4-5), showing the regression line trend to horizontal as the purity decreases.

To assess how consistently the lpWGS assay would perform at different timepoints and study biological variability over time, I sought to compare same-patient biological replicates. Screening and cycle 1 samples, taken at 2 separate blood draws but within 14 days of each other and without any interim treatment, for 85 pairs were analysed (Figure 4-6, below). Despite being collected at different pre-treatment baseline time-points (approximately one week apart), the per-bin log2 ratio profiles correlate well between these biological replicate samples (Pearson's r=0.778). Estimated lp-WGS-derived tumour purities were also correlated but exhibited greater variation (Pearson's r = 0.656), particularly at higher purity values. For subsequent analyses, I elected to use predominantly screening samples so as to have a single baseline sample for each patient.

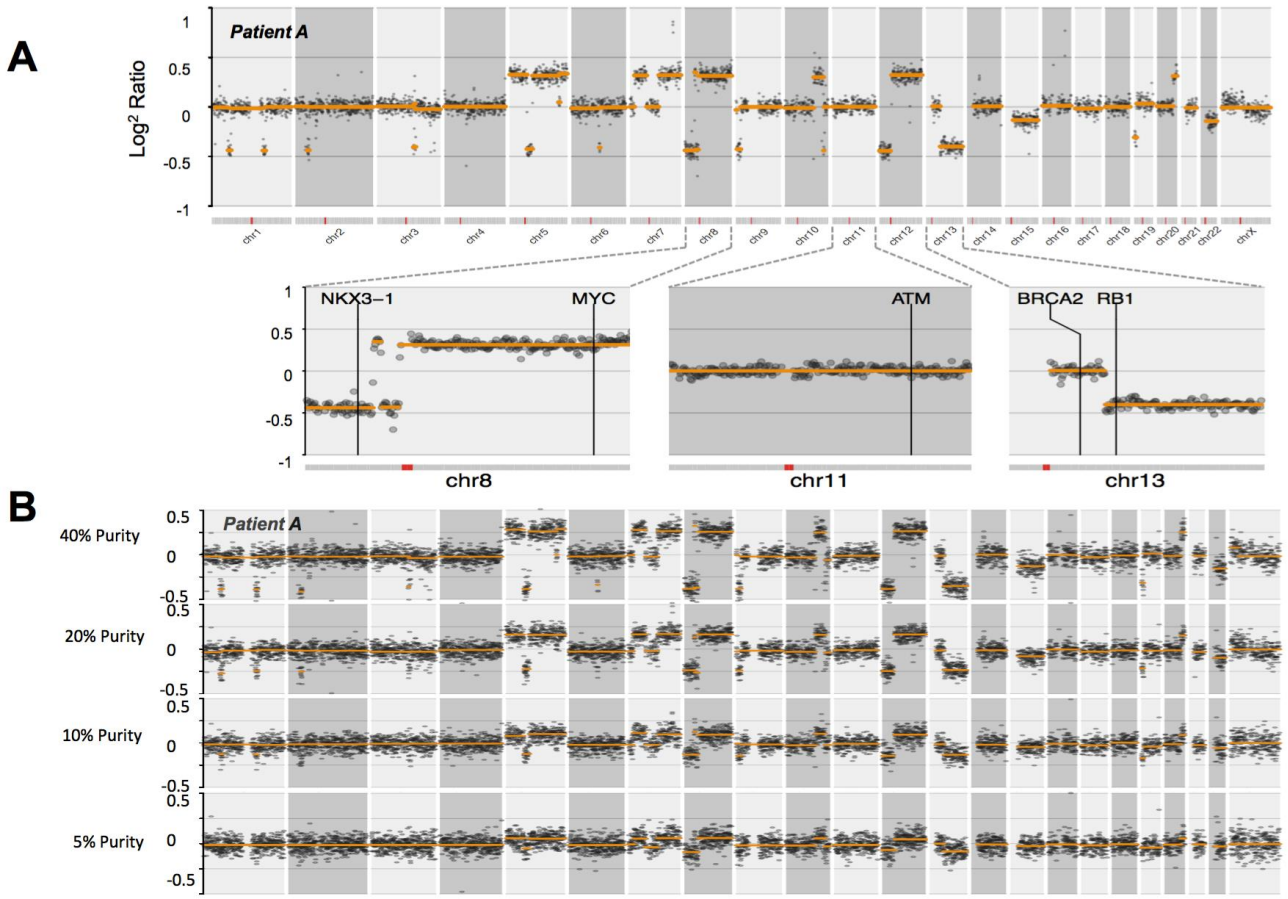


Figure 4-4 Dilution and resequencing of technical replicates

A sample with high (ichorCNA-predicted 50%) tumour purity (A) and clearly observable CNAs was diluted and resequenced (B), with similar genome-wide results observed at each dilution value. As before, grey points are genomic bins and orange lines are assigned segments.

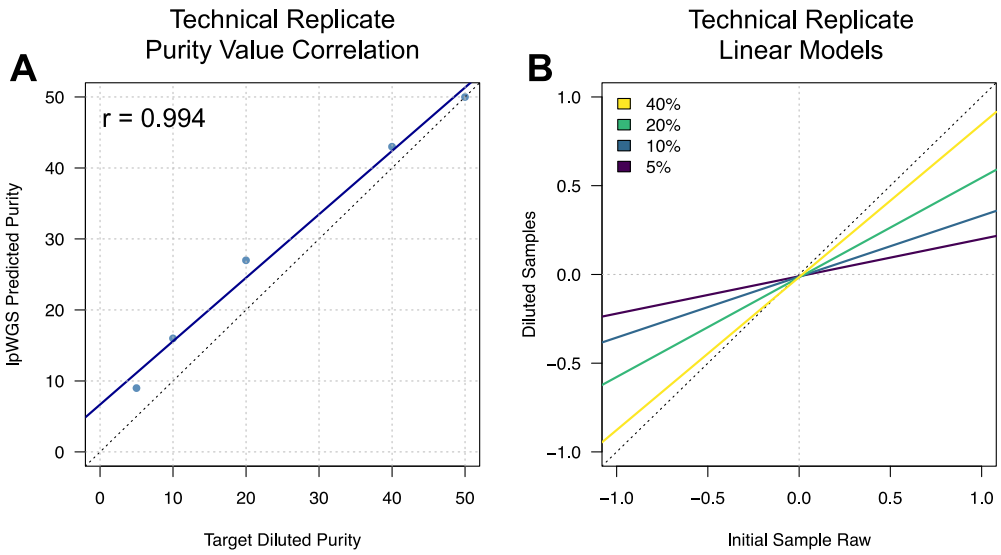


Figure 4-5 Technical replicate validation

The lpWGS copy-number calling was well correlated across technical replicates. Purity values predicted by ichorCNA (y-axis) correlate well with the estimated dilution (x-axis) concentration (A). Linear modelling of the diluted sample CNAs correlated with the original sample demonstrate the effect of a falling DNA concentration (B).

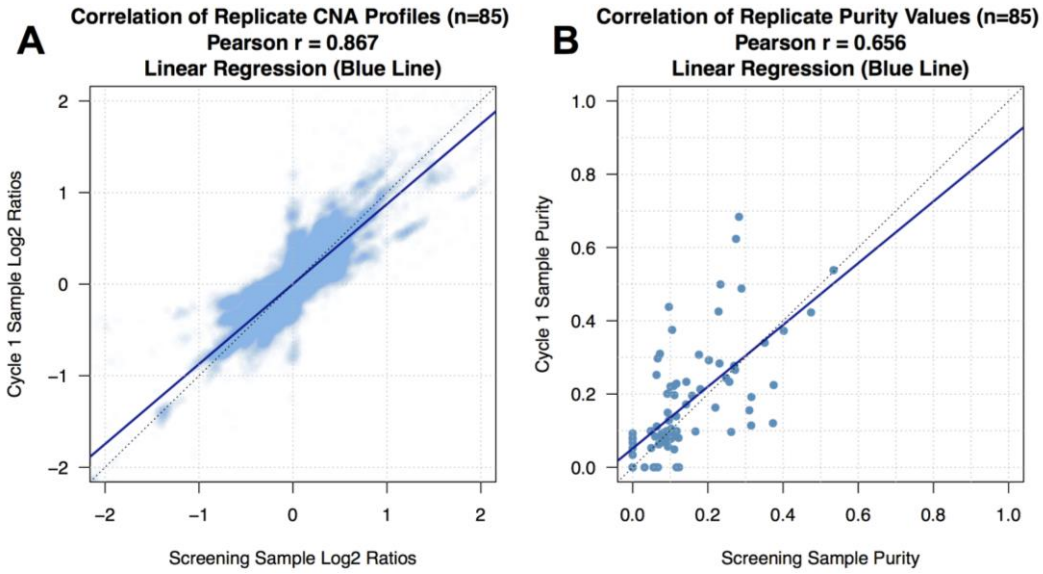


Figure 4-6 Biological replicate correlation

Same-patient biological replicates (taken ~1 week apart) show well correlated log2 ratios (B) and WGS-inferred tumour purity values (B). Screening samples (x-axis), Cycle 1 samples (y-axis)

4.3.4 Baseline Tumour Purity Values are Highly Prognostic But Not Predictive of Drug Response

I assessed baseline lp-WGS data for all 188 individuals, and observed no significant differences in tumour purity or ploidy (Figure 4-7) between the two clinical trials (Wilcoxon rank-sum test *p*-values of 0.29 and 0.75 respectively). Interestingly, there was however a significant increase in the number of copy-number segments called in baseline PROSELICA compared to baseline FIRSTANA (Wilcoxon *p*-value of 0.045), perhaps indicating a change in genomic composition following therapy. Patients were then classified by median baseline cfDNA tumour purity values for univariable survival analysis (ie. high or low purity). This stratification was highly prognostic of overall survival (OS) in both FIRSTANA (log-rank test *p*-value 0.0056) and PROSELICA samples (log-rank test *p*-value <0.0001), with the high purity group (i.e. greater than the median purity value 9.56%) having a ~10 months shorter median overall survival than the low-purity group (Figure 4-8, below).

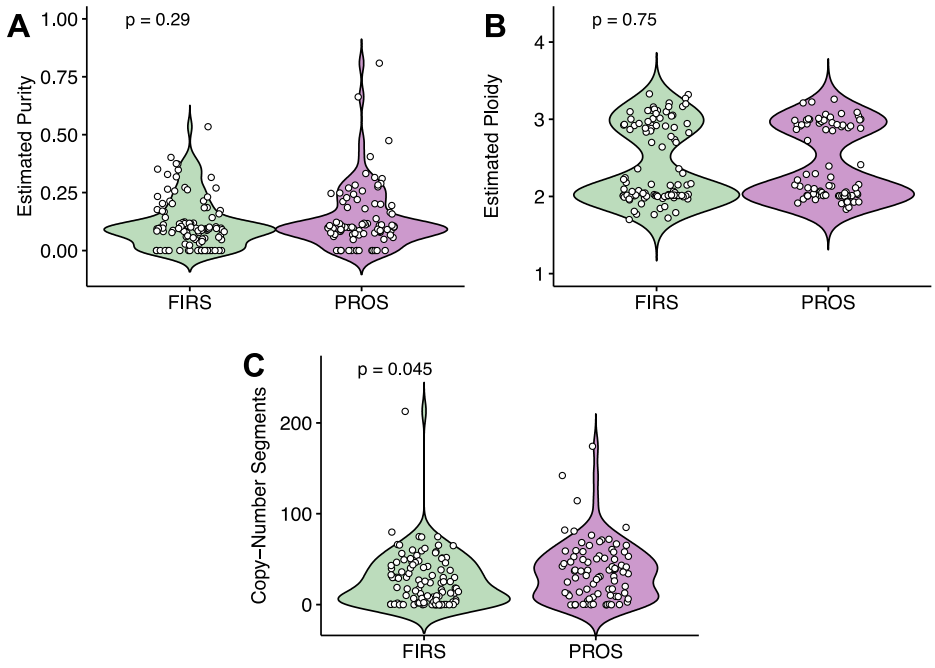


Figure 4-7 Baseline genomic characteristics of lpWGS cohort

Violin plots showing distribution of observed genomic characteristics for baseline FIRSTANA (green) and PROSELICA (violet) cohorts. Neither lp-WGS estimated purity (A) and ploidy (B) values were significantly different between studies. More copy-number segments were observed in PROSELICA when compared to FIRSTANA (C), however.

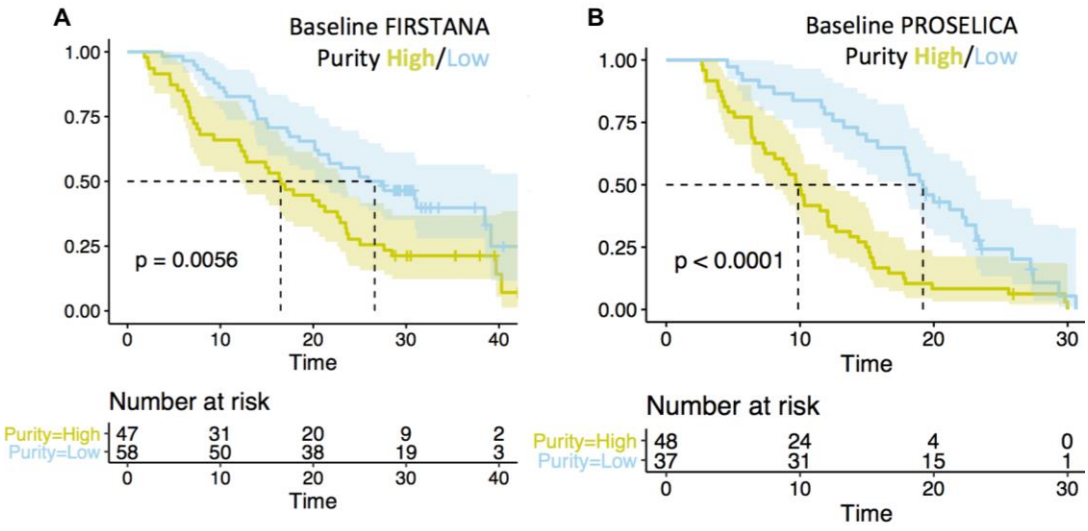


Figure 4-8 Baseline tumour purity values are highly prognostic of overall survival

Segregation of lp-WGS derived tumour purity into high (yellow) and low (blue) groups shows that it is highly prognostic in both FIRSTANA (A) and PROSELICA (B) with ~10 months difference in median survival in both patient cohorts.

For 183/188 patients who had complete clinical data sets, I applied a multivariable Cox proportional hazard model to assess the impact of tumour purity on overall survival (OS) (Figure 4-9, below) and a multivariable logistic regression model to examine the likelihood of response to treatment (Figure 4-10). While baseline tumour purity (treated as continuous variable) from lp-WGS was prognostic for overall survival (HR = 4.42, CI: 1.09-17.82, p=0.037), it was not predictive of response to taxanes (OR = 3.75, CI: 0.22-83.01, p=0.377), indicating that tumour burden does not dictate drug responsiveness. Taxane response was classed as having either radiographic tumour shrinkage or a 50% PSA fall. In addition, other clinical variables (ECOG status, the presence of visceral metastases, baseline LDH and HB concentrations) were also significantly associated with shorter survival. Some clinical variables were associated with tumour purity (Supplemental Figure 10-4), although interestingly the presence of visceral metastasis was not. None were predictive of response to taxane chemotherapy, however.

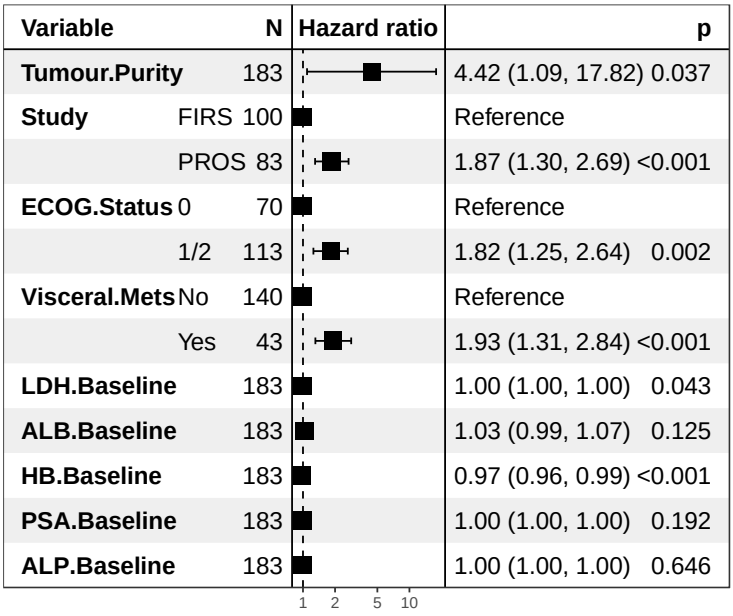


Figure 4-9 Multivariable prognostic model for tumour purity

Tumour purity estimated by lp-WGS is prognostic for tumour overall survival in a multivariable (cox proportional hazards model) context, alongside other established prognostic variables.

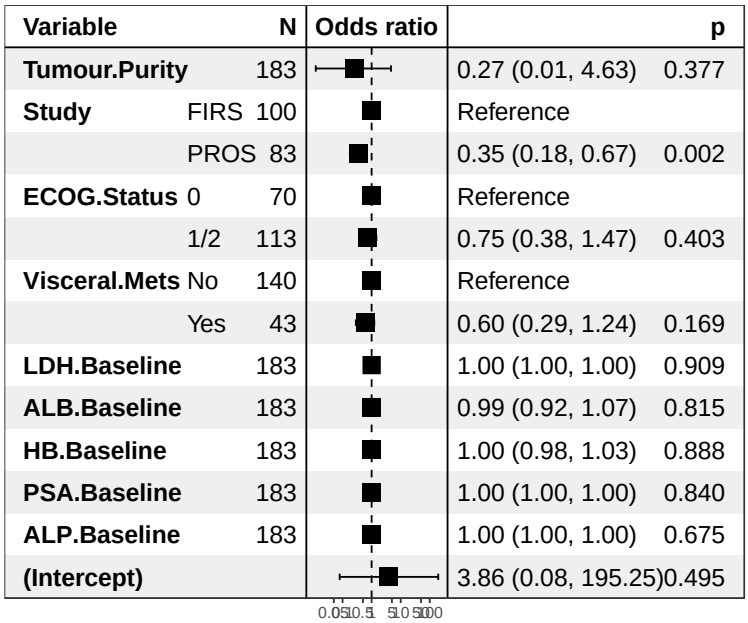


Figure 4-10 Multivariable predictive model for tumour purity

Tumour purity is not predictive of drug response status in a multivariable logistic regression model.

The inclusion of these variables was based on a published set of prognostic variables for mCRPC studies (Halabi et al., 2014). Study ID (ie. being enrolled on PROSELICA rather than FIRSTANA) was also associated with both shorter overall survival and a lower response rate, which is expected due to PROSELICA

being overall a more advanced and heavily treated cohort. These variables were broadly in line with expected results for metastatic CRPC cancers, and were broadly representative of the overall clinical trial population from which these samples were drawn as shown in Supplemental Table 10-7. Some minor differences were observed:

- PROSELICA patients in the cfDNA sub-study were younger at diagnosis (p-value 0.01).
- Serum alkaline phosphatase concentration was greater in the PROSELICA sub-study compared to the wider PROSELICA population (p=0.02).
- Serum haemoglobin concentration was significantly less in both FIRSTANA and PROSELICA sub-studies (p=0.01,0.02 respectively).
- Prostate-specific antigen (PSA) concentration was greater in the PROSELICA sub-study compared to the wider PROSELICA population (p=0.001).

4.3.5 Serial cfDNA Tumour Purity Measures Associate with Treatment Response

I examined individual patient datasets, and observed that longitudinal changes in tumour purity corresponded with response status. Copy number profiles from the same patient are shown (Figure 4-11) at three time points (baseline/C4/EOS), and an initial clinical response to therapy matched with a fall in tumor purity at C4. By EOS the patient had evidence of disease progression coinciding with a rise in lp-WGS tumour purity and detectable copy number events in the EOS sample.

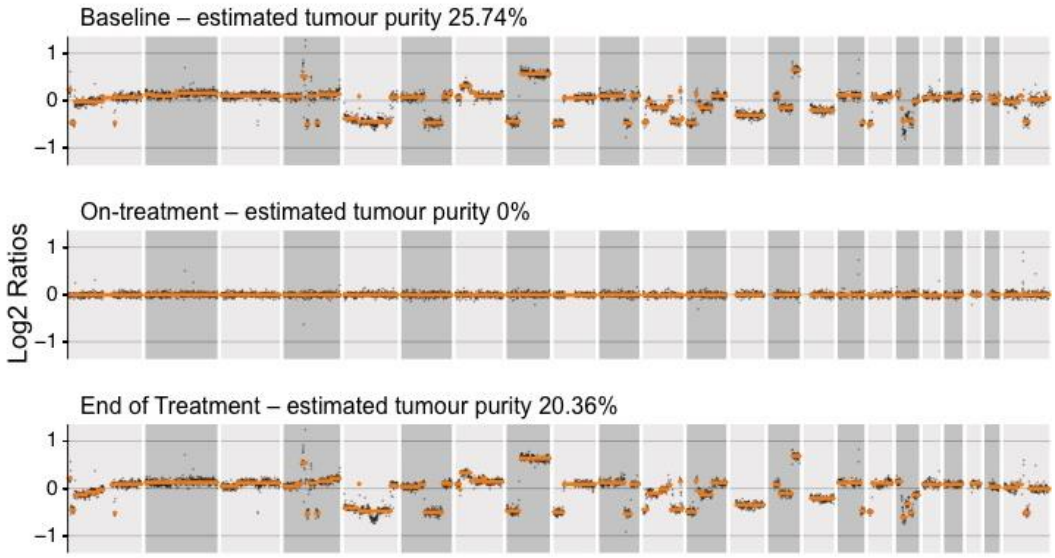


Figure 4-11 Example purity changes in a taxane-responsive patient

Copy-number plots showing genomic regions (grey points) and assigned segments (orange lines) for three same-patient samples at baseline, on-treatment (cycle 4) and end of treatment timepoints. Tumour-derived copy number changes disappear in the on-treatment sample, but return by the end of treatment. This individual was classed as a responder.

These changes were also evident across the entire cohort (unmatched; all evaluable samples analysed), with the proportion of low tumour purity (<5% purity) cases being highest at C4 (48% of cases) increasing from 23% at baseline. Responding individuals (either exhibiting a radiological or PSA-defined response) exhibited significantly decreased (Wilcoxon rank-sum test p value < 0.05) tumour purity values at both on treatment (C4) and end of treatment (EOS) compared to non-responding cases across both trials (Figure 4-12).

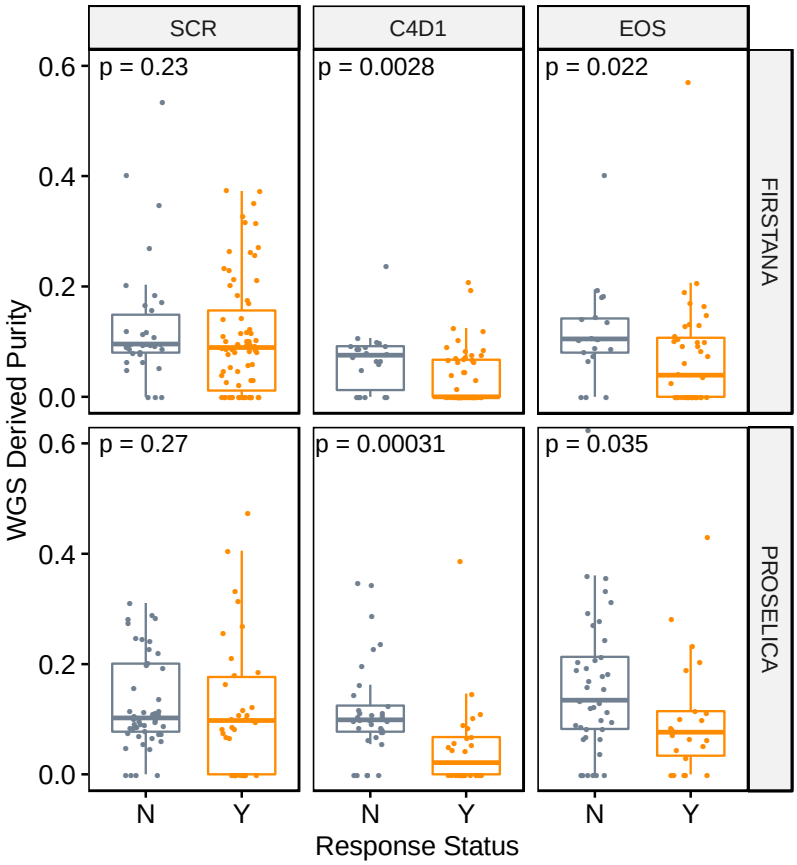


Figure 4-12 Decreases in tumour purity on treatment indicate therapy response

Box-and-whisker plots showing lp-WGS tumour purity values split between taxane responders (orange) and non-responders (grey), across screening (SCR), cycle 4 (C4D1) and end of study (EOS). Samples are segregated between FIRSTANA and PROSELICA trials. Significant differences in tumour purity were observed at C4 and EOS across both trials. Wilcoxon Rank Sum test *p*-values are shown.

Patients where matched baseline (SCR samples preferred except in 3 cases where no SCR was available and C1 was used instead) and on treatment (C4) timepoints with complete clinical data were available (n=132) were grouped into 4 categories based on their baseline and on treatment (C4) purity states. These were defined as high or low purity (split by the median baseline value in this subset of ~9.4%) and classified as: 1) Baseline High; C4 High, 2) Baseline High; C4 Low, 3) Baseline Low; C4 Low; 4) Baseline Low; C4 High. Logistic regression was used for multivariable analysis of the impact of these categories on response rates (Figure 4-13, below). Both Baseline high; C4 low and baseline low; C4 low were significantly associated with a higher likelihood of response to taxanes. As before, study (PROSELICA) was associated with a lower response rate.

Variable		N	Odds ratio	p
Purity	BL_High; C4_High	19	Reference	
	BL_High; C4_Low	47	4.86 (1.29, 21.41)	0.03
	BL_Low; C4_Low	55	4.60 (1.15, 21.48)	0.04
	BL_Low; C4_High	11	1.32 (0.20, 8.69)	0.77
Study	FIRS	75	Reference	
	PROS	57	0.42 (0.18, 0.96)	0.04
ECOG.Status	0	57	Reference	
	1/2	75	0.77 (0.34, 1.75)	0.54
Visceral.Mets	No	106	Reference	
	Yes	26	0.78 (0.29, 2.12)	0.62
LDH.Baseline		132	1.00 (1.00, 1.00)	0.73
ALB.Baseline		132	1.05 (0.96, 1.15)	0.29
HB.Baseline		132	1.00 (0.97, 1.04)	0.91
PSA.Baseline		132	1.00 (1.00, 1.00)	0.21
ALP.Baseline		132	1.00 (1.00, 1.00)	0.46
(Intercept)			0.09 (0.00, 13.81)	0.36

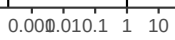


Figure 4-13 Stratification by tumour purity enables use as a response factor

Forest plot showing multivariable cox-proportional hazards model. Patients are stratified based on high or low tumour purity at baseline (BL) and on-treatment (C4) timepoints.

4.4 Discussion

This chapter details the analyses of cell-free DNA collected from two prospective Phase 3 clinical trials of mCRPC with taxane-based chemotherapy: FIRSTANA, taxane naïve patients treated with docetaxel or cabazitaxel and PROSELICA, patients with prior docetaxel exposure treated with two different doses of 20mg/ml^2 or 25mg/ml^2 of cabazitaxel (Oudard et al., 2017). There was no significant differences in survival outcomes between the three arms in FIRSTANA or the two arms in PROSELICA, which demonstrated non-inferiority. These data show that low-pass (~1.5X) whole genome sequencing is capable of producing sensitive, high-quality and reproducible copy-number profiles. In addition, this methodology (using ichorCNA) allows for the estimation of tumour purity (the proportion of tumour DNA in the total cfDNA concentration), a signal distinct from specific genomic alterations. In cases with very low purity values, detecting precise CNAs can be challenging: but low tumour purity is also a prognostic factor in mCRPC itself. I also show that that thanks to the broad coverage and high quantity of data points available for each potential copy-number segment, plausible CNAs can be detected at tumour purities as low as 5%. These data agree with previously published analyses of cfDNA CNAs, and the suggested functional limits of lpWGS (Adalsteinsson et al., 2017; Johansson et al., 2019). While the biological replicate data were linearly correlated, some variation was observed, particularly in purity values, which is likely due to biological variation, as cfDNA production is a dynamic, variable process (Bromberg et al., 2017).

I have shown that baseline values of tumour purity are strongly prognostic in both univariable and multivariable analyses, with a substantial increase of hazard even in the presence of other known clinical variables. In addition, I have found that tumour purity was independently associated with other clinical variables. These data are similar to previously published studies of total cell free DNA concentration; specifically, Mehra *et al* showed that baseline cfDNA concentrations correlated overall survival, with a hazard ratio of 1.53, a significant result but a smaller effect than this study, reflecting the difference between assays (Mehra et al., 2018). However, using the concentration of cell-free DNA as a clinical vairable alone may to be less sensitive than tumour purity estimation from lpWGS, as a significant proportion of cfDNA has been shown to be from non-tumour sources (Aucamp et al., 2018; Bronkhorst et al., 2019).

Studying samples collected at different time-points shows that longitudinal changes in tumour purity values reflect response to therapy, as tumour DNA concentration in the plasma in responding patients falls due to reduced tumour burden. I found significant changes in both studies at both on-treatment and end of study timepoints. This supports the concept that tumour purity is representative of overall tumour burden; therefore these assays may provide a response biomarker capable of tracking treatment response – in addition to providing potentially clinically actionable CNAs (Nygaard et al., 2014; Valpione et al., 2018).

I did not observe any association of baseline tumour purity values with drug response, which confirms data from Mehra *et al*, and suggests that tumour purity is unlikely to serve as a predictive biomarker for response to taxanes (Mehra et al., 2018). The trends for prognostic and response analyses were consistent across both FIRSTANA and PROSELICA cohorts, despite different clinical backgrounds, indicating that these approaches have a broad applicability.

This study may be limited by some inadvertent sample selection: samples were sequenced preferentially on the basis of detectable DNA in the plasma, so this may represent a biased group of cases. This is an intrinsic limitation of the assay, however: blood samples without available DNA cannot be sequenced. Advanced prostate cancer overall does have, however, some of the highest quantities of cfDNA compared to other adult solid tumour types, making these studies especially feasible in mCRPC (Perkins et al., 2012). The availability of STRECK™ tubes to better collect the plasma may improve cfDNA acquisition when concentrations are low and improve the future feasibility of these studies.

To summarise, here I show that patient tumour burden can be accurately and reproducibly estimated by application of low-pass WGS to cfDNA, and have clinical use as markers of survival and drug response. Tumour purity measurements compete with other well-characterised clinical biomarkers, and have a strong association with survival on treatment with taxanes. It was not, however, able to predict drug response.

5 Genomic Copy-Number Signatures of Taxane Chemotherapy

5.1 Introduction

5.1.1 Specific Hypotheses

- Certain copy-number alterations present in CRPC genomes may provide sensitivity or induce resistance to taxane chemotherapy.
- Chemotherapy levies a pressure on tumours that may select for or against these aberrations.
- These aberrations may change in frequency between treated and untreated cohorts.

5.1.2 Specific Aims

- Examine the impact of copy-number burden on overall survival, radiographic progression-free survival, and response to taxane chemotherapy.
- Examine differences between the FIRSTANA and PROSELICA cohorts in terms of CNA frequencies, and longitudinally within these cohorts.
- Expand this approach by screening all available genomic regions for survival and response associations.

5.1.3 Research in Context

- I have shown in Chapter 4 that low-pass WGS is capable of providing accurate copy number data, with alterations corrected for estimated tumour purity and ploidy measurements.
- Aside from tumour purity, specific CNAs are likely to be prognostic of survival in the FIRSTANA and PROSELICA cohorts.
- Mechanisms of resistance and sensitivity may, however, involve multiple diverse processes rather than one specific alteration, and tumour CNAs are frequently large in size, spanning many genomic bins.

I have shown in chapter 3 that targeted NGS analyses can stratify around 40% of mCRPCs into clinically actionable subtypes based on deficiencies in DNA repair. In chapter 4, I found that lpWGS analysis of cfDNA was capable of providing robust whole-genome copy-number profiles, and that in addition the proportion of cfDNA comprised of tumour-derived DNA was highly prognostic of overall survival. As this cfDNA dataset was derived from individuals treated with taxane chemotherapy (commonly used in multiple tumour types), I attempt in this chapter to, in an unsupervised manner, identify genomic predictive biomarkers in CRPCs that may provide useful stratification for taxane treatments, similar to the classifications for PARP inhibitors and immune checkpoint inhibitors. Currently, no such classification exists, and taxane therapies are applied without molecular stratification, despite over half the patients not demonstrating a response, and many having significant side effects.

Data has emerged highlighting the role of tumour copy-number burden in multiple tumour types, including prostate cancer, with high values associated with poor survival and higher risks of relapse (Hieronymus et al., 2014, 2018). CNA burden is associated with aneuploidy (and genomic instability), similar to the results of taxane treatment. Varying results have been shown for a possible relationship between genome instability and taxane resistance – with both pro- and anti-survival effects demonstrated (Jamal-Hanjani et al., 2015; Swanton et al., 2009).

In this chapter, I explore the possibility that copy number burden could associate with taxane response status. I go on to explore the changes in CNA frequencies between cohorts, which may be induced by taxane therapy, and identify significantly altered regions. In addition, as the available dataset is not limited to single genes, I also depict the chromosomal regions identified by a genome-wide screen for association with response and survival.

5.2 Methods

5.2.1 Sample Selection and Data Collection

Samples were collated and initial data processed as described in Chapter 4 from the FIRSTANA and PROSELICA clinical trials. FIRSTANA included patients who had had no prior taxane exposure, although most had had hormone therapy, and the population was randomised to docetaxel or cabazitaxel (20 and 25 mg/ml^2). Patients enrolled on PROSELICA had all been previously treated with docetaxel, and were randomised to two different doses of cabazitaxel (20 and 25 mg/ml^2). The details of both trials have been previously published (Eisenberger et al., 2017; Oudard et al., 2017).

5.2.2 Data Analysis

As in Chapter 4, I used ichorCNA to generate whole genome CNA data from cfDNA for each sample, in addition to purity and ploidy estimates. A bin size of 500kb was used.

To access bin-level copy number ratios, I intersected each sample's ichorCNA segment data with the initial 500kb intervals used for collation of coverage data (n=5226). Intersection and management of genomic coordinate data was performed using the GenomicRanges (v.1.36.0) package in R.

When analysing genomic data, I excluded samples that had an estimated tumour purity of less than 5%. Whole-genome and chromosome plots were generated using the karyoploteR package (v.1.10.4).

5.2.3 Statistical Modelling

For univariable and multivariable models predicting response to therapy, I used generalised linear models (as logistic regressions), provided by the glm() function from the stats package (v3.6.1) in R. Survival models were built using the coxph() function from the survival package (v2.44-1.1). 'Study' (ie. FIRSTANA or PROSELICA) was included as a co-variable in both logistic and survival models. Where stepwise selection was used to modify multivariable models, the Akaike information criterion (AIC) was used in a bi-directional manner using the stepAIC() function from the MASS (v7.3-51.4) package. Elastic net regressions

were performed using the glmnet (v2.0-18) package, with optimal lambda values estimated by: bootstrapping over the cv.glmnet() function, using 10-fold cross validation each time, and returning the lambda with the minimum mean lambda value across all bootstraps. This lambda value was used as an input parameter for the final elastic net, and the alpha parameter was set to 0.5, as a compromise between lasso (fewer final variables) and ridge (more final variables).

5.3 Results

5.3.1 Copy-Number Burden and Taxanes

Following exclusion of samples with less than 5% purity, 144 baseline samples were available, split between FIRSTANA (n=75) and PROSELICA (n=69) cohorts. Both sample sets featured frequent CNAs with a large proportion (median 47%) of the genome impacted.

To investigate the potential impact of overall genomic copy number burden, I performed multivariable analyses of baseline CNA data, with variables representing: the estimated ploidy value, the number of assigned CNA segments, and the overall proportion of the genome impacted by a CNA (Figure 5-1).

A

Drug Response

Variable		N	Odds ratio	p
Ploidy_Round	2	78		Reference
	3	66		0.84 (0.42, 1.70) 0.635
Number_Segments		144		0.99 (0.98, 1.01) 0.278
Percent_Altered		144		1.02 (1.00, 1.03) 0.031
Study	FIRS	75		Reference
	PROS	69		0.33 (0.16, 0.65) 0.002
(Intercept)				1.06 (0.45, 2.53) 0.898

B

Overall Survival

Variable		N	Hazard ratio	p
Ploidy_Round	2	78		Reference
	3	66		1.00 (0.69, 1.44) 0.986
Number_Segments		144		1.01 (1.00, 1.02) 0.020
Percent_Altered		144		1.00 (0.99, 1.01) 0.864
Study	FIRS	75		Reference
	PROS	69		1.77 (1.22, 2.56) 0.003
Purity		144		1.66 (0.28, 9.77) 0.577

C

Progression-Free Survival

Variable		N	Hazard ratio	p
Ploidy_Round	2	78		Reference
	3	66		1.23 (0.75, 1.99) 0.4
Number_Segments		144		1.02 (1.01, 1.03) <0.001
Percent_Altered		144		1.00 (0.99, 1.01) 0.6
Study	FIRS	75		Reference
	PROS	69		1.44 (0.91, 2.28) 0.1
Purity		144		0.59 (0.07, 5.00) 0.6

0.1 0.2 0.5 1 2 5

Figure 5-1 Multivariable models of CNA burden

Forest plots showing models of (A) drug response (logistic regression), (B) overall survival (Cox proportional hazards) and (C) radiographic progression-free survival (Cox proportional hazards) models for copy number burden data.

I found that the percentage proportion of the genome impacted by a CNA modestly associated with odds of responding to therapy; a higher percentage associated with higher odds (p=0.03). This was not reflected in the OS or RPFS, however. The estimated ploidy value did not associate with either drug response

or overall survival. Interestingly, there was an association between the number of copy-number segments and a slightly poorer overall survival (p=0.02), although the change in hazard was modest (HR 1.01). This was supported by the RPFS (p=<0.001), although again the change in hazard was again slight (HR 1.02).

5.3.2 Baseline CNA Frequency Differences Between Cohorts

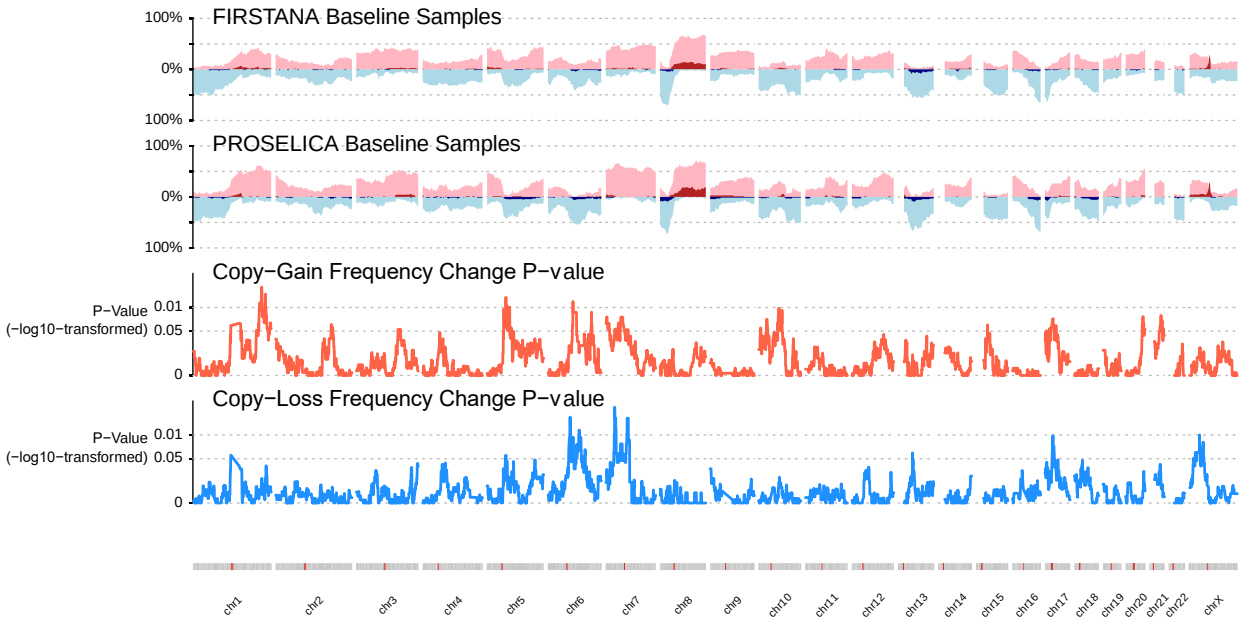


Figure 5-2 Genome-wide copy number frequencies observed in the FIRS and PROS baseline cohorts

Stacked plots indicating CNA frequencies and p-values of fisher's exact tests comparing gains and losses between FIRS and PROS baseline cohorts. Frequencies displayed for copy-gains (pink), amplifications (red), copy-losses (light blue) and deletions (navy). P-values presented as –log10 transformed.

To assess changes in CNA frequency between cohorts, I compared the frequency of copy-gains and copy-losses in a genome-wide manner (Figure 5-2). This analysis offers an unbiased approach. Additionally, while many statistical tests are performed as part of this analysis, individual CNA data points are not unrelated, with many adjacent bins bearing similar results. For this reason, and as this is a hypothesis-generating analysis, I did not perform multiple testing correction.

When comparing FIRS (taxane naïve) baseline samples against PROS (docetaxel treated) baseline samples, although both cohorts had broadly similar mCRPC-like profiles, I found that several regions showed highly significant ($p <= 0.01$) shifts in either copy-gain or copy-loss CNA frequencies: chr1 (211-230.5Mb), chr5 (59.5-63Mb and 71-73Mb), chr6 (70-72Mb, 79.5-8Mb, 99-100.5Mb), chr7 (27-33Mb and 67-72Mb) and X (34-34.5Mb), covering 60 bins and 30Mb in total. Each bin is 500kb in size – the input bin size selected for the copy number analysis.

Several other regions displayed some significance ($p <= 0.05$), including areas on chromosomes 2, 3, 10, 13, 15, 17, 20 and 21 (along with wider areas on 1, 5, 6, 7 and X) (Figure 5-2). In total, 491 bins (245.5Mb) exhibited some significant changes for either copy-gain or -loss frequencies.

5.3.3 Longitudinal CNA Frequency Shifts in FIRSTANA and PROSELICA

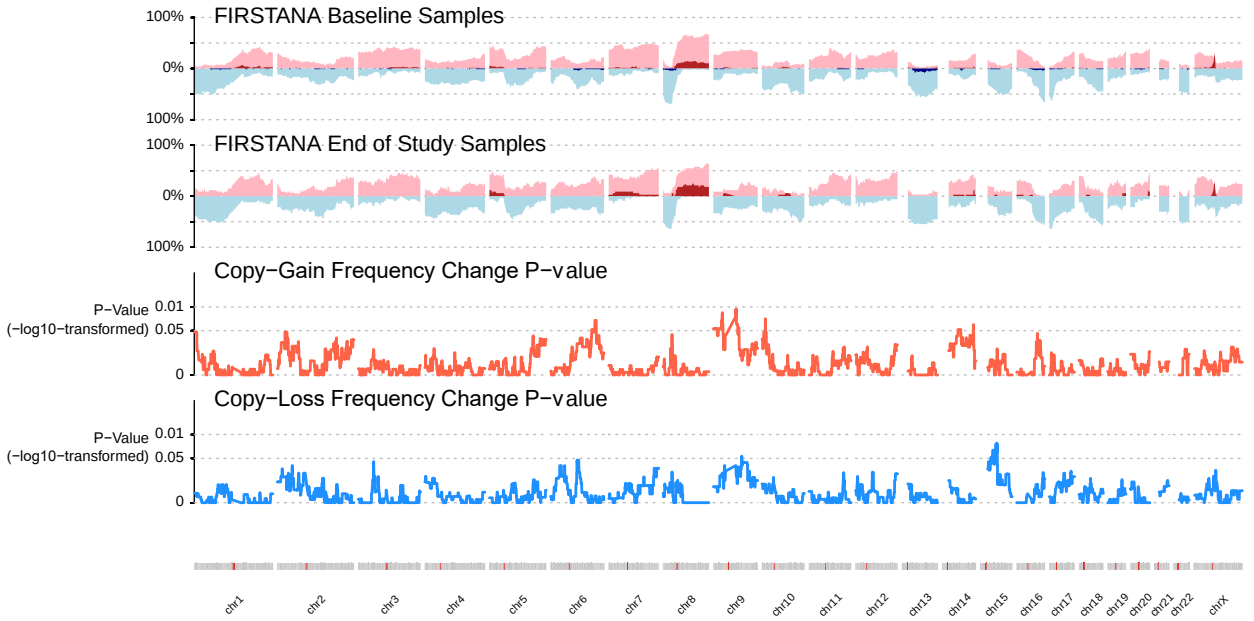


Figure 5-3 Genome-wide copy number frequencies observed in the FIRS baseline and end of study cohorts

Stacked plots indicating CNA frequencies and p-values of fisher's exact tests comparing gains and losses FIRSTANA baseline and end of study sample sets. Frequencies displayed for copy-gains (pink), amplifications (red), copy-losses (light blue) and deletions (navy). P-values presented as –log10 transformed.

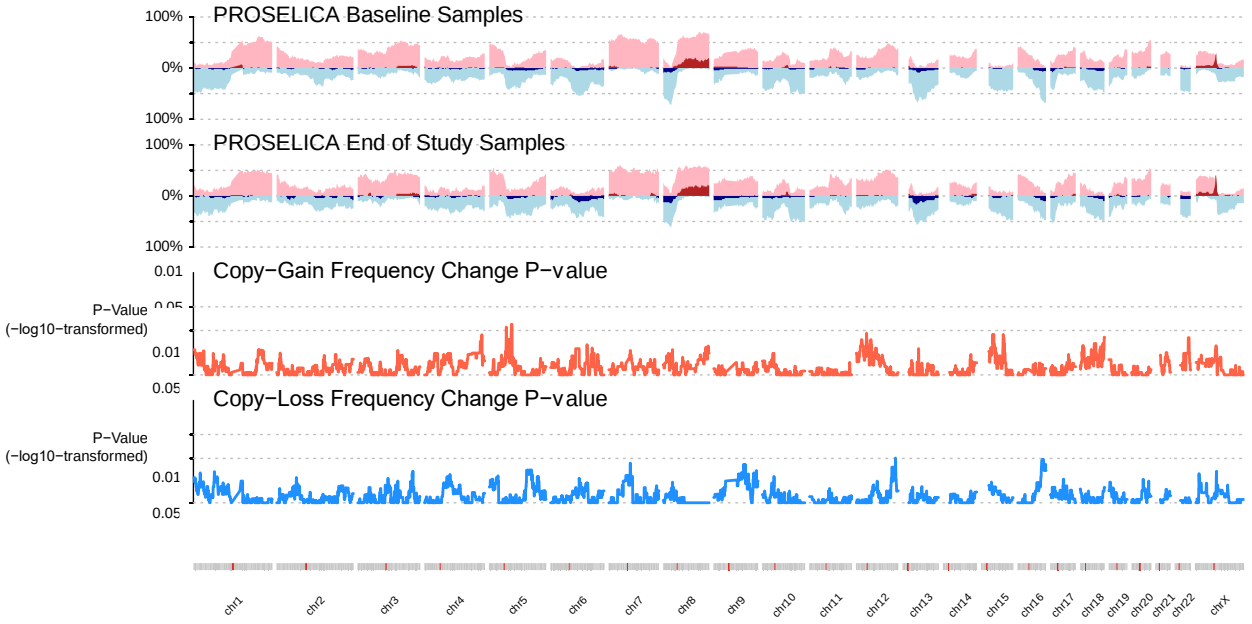


Figure 5-4 Genome-wide copy number frequencies observed in the PROS baseline and end of study cohorts

Stacked plots indicating CNA frequencies and p-values of fisher's exact tests comparing gains and losses PROSELICA baseline and end of study sample sets. Frequencies displayed for copy-gains (pink), amplifications (red), copy-losses (light blue) and deletions (navy). P-values presented as –log10 transformed.

I performed similar analyses comparing the longitudinal (baseline to end of study) FIRSTANA and PROSELICA sample cohorts. When comparing baseline (n=75) and EOS (n=33) samples from FIRSTANA, several regions (on chromosomes 6, 9 10, 14 and 15, 146 bins and 73Mb) were highlighted as having significant changes ($p<0.05$), surprisingly almost all distinct from those in the prior comparison of baseline FIRS/PROS samples (Figure 5-3).

There were, however several regions that were significantly (<0.05) altered in both this comparison and the BL FIRS/PROS analysis, including chr6 139Mb-140Mb (q23) and chr15 36-37Mb (q14). CNAs in these regions may be related to changes induced by chemotherapy.

I performed similar analysis using the PROSELICA cohort, comparing baseline (n=69) and EOS (49) samples, and found that far fewer regions showed a significant frequency shift: only six bins, of which two overlapped with those from the BL FIRS/PROS: chr5 53-56Mb (q11) and chr5 71-73Mb (q13) (Figure 5-4). This is perhaps unsurprising, since the PROSELICA cohort were already docetaxel exposed.

5.3.4 Whole-Genome Screening of Response Associations

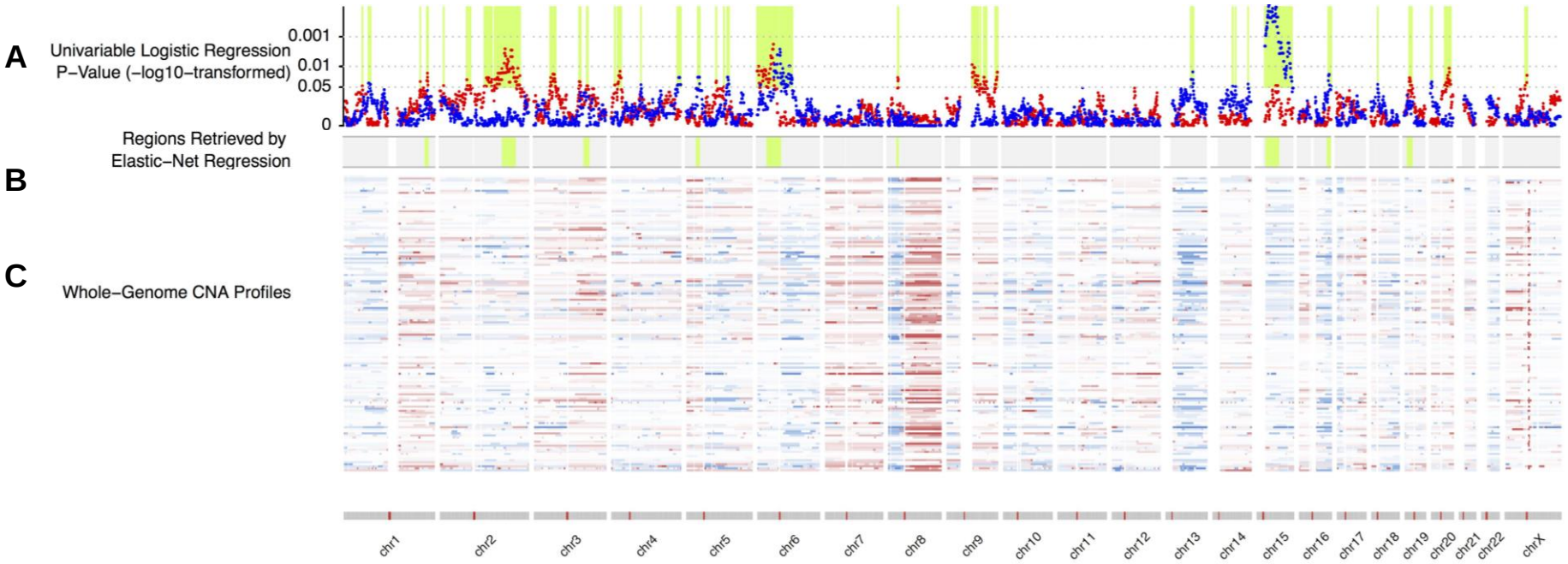


Figure 5-5 Genome-wide screen for response-associated CNAs

Stacked plots showing: univariable logistic regression results for CNAs in each binned genomic region (A), -log10 p-values shown (copy-gain red, copy-loss blue), with significant regions highlighted in green; the regions (green) retained after variable selection with elastic net regression (B); the input CNA data for this analysis with copy-gains in red and copy-losses in blue, sorted by response status.

To identify genomic regions that were significantly predictive of a change in drug response in an unsupervised manner, I performed a multi-step analysis on the baseline data set to identify areas in the binned genomic CNA that can be candidates for predictive biomarkers of drug response, and convert the bins into merged genomic regions (Figure 5-5). The approach was as follows:

1. Sequentially apply a univariable logistic regression model, as before, to each genomic bin, and identify regions with any significant (<0.05) copy-gain or copy-loss associations with taxane response. At the start of this analysis, 762 bins were included.
2. Use these regions as input variables for a penalised elastic-net regression model, which performs a multivariable analysis and shrinks coefficients that do not contribute to the model. This results in a small set of variables suitable for further analysis, n=56.
3. As the data features many variables that are highly collinear with each other (ie. regions of CNA cover many bins and have similar results in each bin), and elastic-net regression tends to exclude these variables, expand the list of variables with any other adjacent bins that have highly (r>0.9) correlated CNA data, n=246.
4. Collapse the set of bins (each one 500Kb) into broader genomic regions (of varying length), based on adjacency (a gap of 10Mb was permitted between regions). This resulted in 9 broad regions, covering 152Mb.
5. Calculate the modal value of CNA of bins within that region for use in downstream regression models.
6. Perform follow-up multivariable logistic and survival analysis on the final variables.

I applied multivariable modelling on the 9 genomic regions highlighted using this methodology, for both drug response and survival (Figure 5-6, below). I performed stepwise selection to refine the set of variables for the final models. Six regions were retained in the multivariable analysis, with all these final regions showing some association with drug response. In addition, I did not distinguish between varying intensities of CNA in the subsequent analyses, ie. between hemi/homozygous deletions, or copy-gain/amplification. I made two small modifications to the models: including a variable ‘Study’ in all analyses (including nominally univariable models), as both studies had different response rates; and for CNAs with extremely low event frequencies (ie. copy-losses or gains

occurring in less than 10% of cases) these were merged with the ‘Neutral’ category, to avoid over-interpreting results with very sparse events.

Some regions had significant associations with **response** ratio and changes in **survival** time.

- Copy-losses in the chr15 (30-63.5Mb) region strongly associated with response with a 64.5% response rate (p=0.002, OR 4.55), and an improved OS (p=0.01, HR 0.58), although RPFS did not appear associated (p=0.87, HR 1.05).
- A region on chr19 (12.5-24.5Mb) also displayed a clear result, with copy-gains having a significantly worse response rate of 33.3% (p=0.005, OR 0.2), and shorter OS and RPFS (p=0.03 and 0.01, HR 1.67 and 2.12).
- Aberrations on chromosome 6 (29-65Mb) in this cohort presented counterintuitive results: both copy gains and losses were associated with improved response rates (71.4% and 65.2%), which were statistically significant (p=0.008 and 0.030, ORs 5.27 and 4.93), however there was no multivariable association with OS or RPFS.

These regions are interesting candidates for predictive biomarkers. In particular, chromosome 15 has been highlighted in section 5.3.2 and 5.3.3, and contains several genes of interest – including the BUB1B and PAK6 locus. Both genes are serine/threonine kinases with links to cell cycle progression and mitosis. Additionally, following false discovery rate correction (using the Benjamin-Hochberg procedure, with the initial univariable statistical tests), this region was also the only one remaining with a significant (<0.05) p-value.

A Taxane Drug Response

Variable		N	Odds ratio	p
chr15_30000001_63500000	Neutral	82	Reference	
	Loss	62	4.55 (1.82, 12.21)	0.002
chr16_82000001_89000000	Neutral	41	Reference	
	Loss	103	3.07 (1.08, 9.44)	0.040
chr19_12500001_24500000	Neutral	85	Reference	
	Gain	33	0.20 (0.06, 0.59)	0.005
	Loss	26	0.71 (0.20, 2.43)	0.583
chr2_172000001_206000000	Neutral	86	Reference	
	Gain	29	3.42 (1.08, 11.74)	0.041
	Loss	29	0.32 (0.08, 1.09)	0.080
chr6_29000001_65000000	Neutral	93	Reference	
	Gain	28	5.27 (1.62, 19.46)	0.008
	Loss	23	4.93 (1.26, 23.26)	0.030
chr8_27500001_30000000	Neutral	43	Reference	
	Loss	101	0.30 (0.09, 0.88)	0.033
purity		144	1.67 (0.04, 61.19)	0.783
studyID	FIRS	75	Reference	
	PROS	69	0.17 (0.06, 0.40)	<0.001
(Intercept)			0.91 (0.31, 2.69)	0.864

B Overall Survival

Variable		N	Hazard ratio	p
chr15_30000001_63500000	Neutral	82	Reference	
	Loss	62	0.58 (0.38, 0.90)	0.01
chr16_82000001_89000000	Neutral	41	Reference	
	Loss	103	1.14 (0.74, 1.76)	0.54
chr19_12500001_24500000	Neutral	85	Reference	
	Gain	33	1.67 (1.06, 2.66)	0.03
	Loss	26	0.81 (0.49, 1.36)	0.43
chr2_172000001_206000000	Neutral	86	Reference	
	Gain	29	1.03 (0.61, 1.72)	0.91
	Loss	29	1.50 (0.83, 2.71)	0.18
chr6_29000001_65000000	Neutral	93	Reference	
	Gain	28	0.89 (0.50, 1.56)	0.68
	Loss	23	1.28 (0.68, 2.42)	0.45
chr8_27500001_30000000	Neutral	43	Reference	
	Loss	101	0.87 (0.56, 1.37)	0.56
purity		144	4.57 (1.12, 18.60)	0.03
studyID	FIRS	75	Reference	
	PROS	69	1.98 (1.36, 2.88)	<0.001

C Radiographic Progression-Free Survival

Hazard ratio	p
Reference	
1.05 (0.61, 1.81)	0.87
Reference	
1.00 (0.55, 1.81)	0.99
Reference	
2.12 (1.19, 3.76)	0.01
0.73 (0.34, 1.55)	0.41
Reference	
0.72 (0.38, 1.38)	0.32
1.75 (0.91, 3.36)	0.10
Reference	
1.18 (0.58, 2.39)	0.66
1.10 (0.52, 2.34)	0.81
Reference	
0.77 (0.43, 1.36)	0.36
0.98 (0.15, 6.55)	0.98
Reference	
1.58 (0.98, 2.53)	0.06

Figure 5-6 Final multivariable models of unsupervised genomic regions

Forest plots showing models of (A) drug response, (B) overall survival and (C) radiographic progression-free survival for CNA data derived from genomic regions highlighted in unsupervised analyses. Variables for tumour purity and study are included to model the impact of tumour burden and varying hazard between cohorts, respectively.

5.4 Discussion

Currently, no genomic stratification for taxane treatment exists. I illustrated in Chapter 4 how lpWGS from cfDNA offers high-quality genome-wide CNA data, but is limited by tumour purity. I also showed that tumour purity is itself highly prognostic, and may be considered a surrogate marker for tumour burden. Critically, I found that changes in cfDNA tumour purity associate with drug response status, but baseline concentrations did not. In this chapter, my work focused on first examining the impact of copy-number burden on clinical outcome, before studying on the genomic differences observed between patients with varying amounts of prior treatments (taxane naïve vs docetaxel pre-treated), and between taxane non-responders and responders.

The CNAs studied in these analyses are typically large in scale, frequently covering tens of thousands of nucleotides and hundreds of genes. Therefore, while studying individual genes is of interest in certain contexts, here I performed genome-wide studies to try and identify clinically relevant loci. Dissecting these loci with further functional studies will be key to identifying mechanisms impacting taxane sensitivity. This cohort of patients represents a set of advanced diseases, progressing on prior hormone therapies, and all bear significant genomic aberrations that likely drive their disease. This may explain why CNA burden was not strongly associated with survival or response in these data, compared to other studies.

There are no published tools or ‘best practices’ for the application of clinical statistical models to genome wide copy-number data, so for this study I designed an approach that was both conceptually simple and statistically robust. I did not include false-discovery rate (FDR) correction during the initial univariable analysis step, as the number of variables analysed would reduce the number of potential hits to a negligible number, and this step was designed to provide hits for further validation, rather than fully explain the variables. Additionally, FDR correction may not be appropriate to CNA data, due to the presence of large regions of near identical p-values. However, the subsequent use Elastic-net regression avoids some of the pitfalls related to multiple testing, and I carefully merged adjacent genomic loci to avoid having multiple highly collinear variables in the final models. Interestingly, several genes of interest are located in the

highly significant chr15 (30-63.5Mb) region, including *BUB1B* and *PAK6*, both of which have been implicated in taxane resistance and sensitivity (Bargiela-Iparraguirre et al., 2014; Wen et al., 2009). Defining and exploring the genes of interest in these regions will require future studies.

In summary, in this chapter I have sought to examine changes in CNA profiles due to taxane therapy, and identify potential predictive factors that may explain drug resistance or sensitivity. I found no association with CNA burden and taxanes, although this may be related to the cohort analysed in this study. However, several genomic regions were highlighted in this dataset, which may provide a valuable starting point for further investigations. These results are to be considered hypothesis-generating, and therefore validation of these regions with independent samples and orthogonal technologies will be key, with identification of the genes responsible crucial to offering a clinical stratification for taxanes.

6 Supervised Analysis of Taxane Associated Molecular Pathways

6.1 Introduction

6.1.1 Specific Hypotheses

- The mechanism of action of taxanes relies on inhibition of mitosis and cell cycle interruption, and molecular pathways that control these processes may play a role in taxane resistance and sensitivity.
- Taxanes bind directly to beta-tubulin subunits present in microtubules, so CNAs of this gene family may impact drug efficacy.
- Key copy-number events found in mCRPCs may play roles in taxane chemoresistance, in addition to serving as prognostic markers.

6.1.2 Specific Aims

- Analyse a targeted set of tubulin family genes for both changes in CNA frequency and also clinical impact.
- Similarly, test previously highlighted gene copy-number events in other important molecular pathways for prognostic and predictive impact in the FIRSTANA and PROSELICA cohorts.
- Distinguish between predictive and prognostic CNAs in these cohorts.

6.1.3 Research in Context

- I have shown in Chapter 5 that several genomic regions are associated with changes in response and survival.
- Additionally, I detected significant alterations in the frequency of CNAs between untreated and treated sample sets.
- Studies have reported several possible mechanisms for taxane resistance, but there are currently no predictive biomarkers in place for stratifying patients for taxane therapy.
- Tubulin genes are an axis of potential taxane resistance due to these being the main drug target, with these being commonly altered in many tumour types.

In chapter 4 I showed that lpWGS analyses can provide robust CNA data, and in chapter 5 I found that several regions, such as on chromosomes 2, 6, 15 and 19 may contain key genes that confer taxane resistance or sensitivity. Various studies have also implicated several functional pathways in taxane resistance, so here I seek to test molecular pathways for significant clinical associations.

The taxanes are a class of agents that bind to the beta-tubulin protein, which normally forms heterodimers with alpha-tubulin and subsequently assembles into microtubules (a key part of the mitotic spindle) (Fitzpatrick and de Wit, 2014; Ganguly et al., 2012; de Leeuw et al., 2015; Parker et al., 2014). Microtubule function relies on being able to rapidly polymerise and depolymerise, growing and shortening in a dynamic process. Taxane binding, promotes over-stabilisation of the complex, preventing depolymerisation, which traps the filament in place, thus limiting microtubule function, preventing cell division and interphase functions. Identified in the 1970's, paclitaxel was the first of this drug class, followed by subsequent developments of docetaxel and cabazitaxel, sharing the same chemical backbone but with modified functional groups (de Bono et al., 2010; Bumbaca and Li, 2018; Tannock et al., 2004).

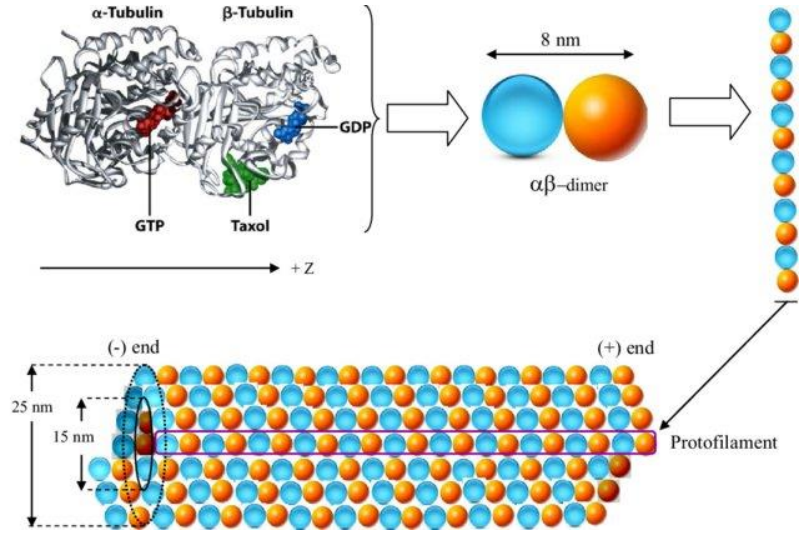


Figure 6-1 Tubulin dimers and microtubule assembly

Tubulin α and β isoforms heterodimerise before aligning head-to-tail into protofilaments, and then polymerising together (13 filaments together) into hollow microtubules (Saeidi et al., 2014).

The numerous possible combinations of alpha and beta tubulin isoforms, along with post-translational modifications, allow for regulation of heterodimer binding

affinities and thus downstream microtubule dynamics (Miller et al., 2010; Vemu et al., 2017). Differential tubulin isoform expression is found between different healthy tissues, and represents a method of control for core cellular processes, such as cell division and motility.

The acquisition of mutations in βI-tubulin (*TUBB*), particularly at residues 270 and 364, has been associated with taxane resistance in cell line models (Bumbaca and Li, 2018; Giannakakou et al., 1997). Another site between residues 215 and 228 has been implicated as well; as with other variants the precise mechanism of resistance remains unknown, although the position of variants around the taxane binding site is suggestive (Gonzalez-Garay et al., 1999). Heightened expression of the β3 isoform (encoded by the *TUBB3* gene) has also been associated with reduced sensitivity to taxanes and poorer survival in multiple tumour types, although it has also been linked to other therapies including hormone treatment (Hwang et al., 2013; Loeser et al., 2017; Ploussard et al., 2010; Sobue et al., 2016). One study reported frequent copy number events and expression changes in both tubulin family members and tubulin-associated genes, including both *TUBB* and *TUBB3*, in paclitaxel resistant breast cancer cell lines (Nami and Wang, 2018). These tubulin-based adaptive mechanisms have been observed across the various members of the taxane family of drugs (to varying degrees), including cabazitaxel, which was engineered to bypass docetaxel resistance (Duran et al., 2015; de Leeuw et al., 2015). Aside from providing mechanisms of resistance, modulation of tubulin family proteins in cancer cells is likely to provide a broad suite of pro- and anti-oncogenic processes, thanks to diverse roles in the cytoskeleton, cell motility and intracellular transport (Cirillo et al., 2017; Ganguly et al., 2012; Parker et al., 2014).

Multidrug Resistance (MDR) family proteins are also well established as causes of chemotherapy resistance across cancer types by transporting chemicals out of the cell. Members of the ABC protein superfamily (featuring ATP-binding cassettes), such as ABCB1 and ABCC4 have been demonstrated to confer taxane resistance (Duran et al., 2015; Oprea-Lager et al., 2013; Sissung et al., 2008). Some cell lines that resist docetaxel are sensitive to cabazitaxel, which has reduced affinity for MDR-mediated efflux (Fumoleau et al., 2013; Oprea-Lager et al., 2013).

Additionally, several key CRPC oncogenic pathways, including the androgen receptor pathway, the PI3K-AKT axis, and cell cycle regulation have been reported to impact on taxane resistance or sensitivity (Duran et al., 2015; Wang et al., 2013). Several studies have shown that the tumour microenvironment can play a role in tumour cell survival in the presence of taxanes. MDSCs may induce senescence, protecting tumours from spindle poisons by slowing cell division, and interestingly tumour cells may shelter within the perivascular niche, protected by interactions with neighbour cells and vascular endothelium (Carlson et al., 2019; Di Mitri et al., 2014; Kampan et al., 2015). The development of cancer stem cells has also been shown to provide some protection from taxane-induced cell death, with high expression of stem cell markers (CD133, CD44) linked to this phenomenon (El-Khattouti et al., 2014; Mittal et al., 2017).

It is clear that a plethora of approaches are available for tumour cells to escape microtubule-induced arrest, ranging from specifically removing drug molecules from the cytosol to induction of senescence and alternative mechanisms of cell division. In this chapter, I plan to interrogate a select group of genes previously reported in taxane chemoresistance, to explore the possibility that copy number changes common to CRPCs could provide predictive biomarkers. I also present here comparisons in CNA frequencies between taxane-naïve and taxane-treated cohorts, and data showing treatment-induced changes in CNA frequency over time.

6.2 Methods

6.2.1 Sample Selection and Data Collection

Samples were collated and initial data processed as described in Chapter 4 and 5 from the FIRSTANA and PROSELICA clinical trials. FIRSTANA included patients who had had no prior taxane exposure, although most had had hormone therapy, and the population was randomised to docetaxel or cabazitaxel (20 and 25 mg/ml^2). Patients enrolled on PROSELICA had all been previously treated with docetaxel, and were randomised to two different doses of cabazitaxel (20 and 25 mg/ml^2). The details of both trials have been previously published (Eisenberger et al., 2017; Oudard et al., 2017).

6.2.2 Data Analysis

As in Chapters 4 and 5, I used ichorCNA to generate whole genome CNA data.

As before, to access Per-gene copy number data, I intersected each sample's ichorCNA segment data with the genomic coordinates. Gene coordinates for hg19 retrieved using biomaRt (v2.36.1). When analysing genomic variable data, I excluded samples that had an estimated tumour purity of less than 5%.

6.2.3 Statistical Modelling

For univariable and multivariable models predicting response to therapy, I used generalised linear models (as logistic regressions), provided by the glm() function from the stats package (v3.6.1) in R. Survival models were built using the coxph() function from the survival package (v2.44-1.1). 'Study' (ie. FIRSTANA or PROSELICA) was included as a co-variable in both logistic and survival models. Where stepwise selection was used to modify multivariable models, the Akaike information criterion (AIC) was used in a bi-directional manner using the stepAIC() function from the MASS (v7.3-51.4) package.

6.3 Results

6.3.1 Frequency Changes in Tubulin Genes

Beta-Tubulins are the target of taxane cytotoxics. Due to the reports implicating regulation of tubulins in: a) tumour development and b) taxane drug sensitivity and resistance, I pursued an investigation of tubulin genes and tubulin-beta specific chaperone proteins (Bumbaca and Li, 2018; Giannakakou et al., 1997; Hwang et al., 2013; Nami and Wang, 2018; Ploussard et al., 2010). Following the generation of whole-genome copy number profiles, I intersected those data with the genomic coordinates of this gene set. Genes analysed are shown in Table 6-1. All tubulin group genes and chromosomal coordinates were accessed from the Hugo Gene Nomenclature Committee database, retrieved 2nd of July 2019. Adjacent genes were merged, as shown below in Table 6-1.

I excluded any samples with <5% tumour purity from all subsequent analysis of genomic data, as these would erroneously appear to be false-negatives. This left 144 Baseline (BL) and 82 End-of-Study (EOS) samples available for study, split between the FIRSTANA (FIRS) and PROSELICA (PROS) cohorts.

Gene	Group	Chrom	Band
TBCE	Chaperone	1	q42.3
TUBA3D/TUBA3E	Tubulin A	2	q21.1
TUBA4A	Tubulin A	2	q35
TBCA	Chaperone	5	q14.1
TUBB2A/TUBB2B	Tubulin B	6	p25.2
TUBB	Tubulin B	6	p21.33
TBCC	Chaperone	6	p21.1
TUBE1	Tubulin E	6	q21
TUBB4B	Tubulin B	9	q34.3
TUBB8	Tubulin B	10	p15.3
TUBAL3	Tubulin A	10	p15.1
TUBA1A/TUBA1B/TUBA1C	Tubulin A	12	q13.12
TUBA3C	Tubulin A	13	q12.11
TUBB3	Tubulin B	16	q24.3
TUBG1/TUBG2	Tubulin G	17	q21.2
TUBD1	Tubulin D	17	q23.1
TBCD	Chaperone	17	q25.3
TUBB6	Tubulin B	18	p11.21
TUBB4A	Tubulin B	19	p13.3
TBCB	Chaperone	19	q13.12
TUBB1	Tubulin B	20	q13.32
TUBA8	Tubulin A	22	q11.21

Table 6-1 Tubulin genes and beta-tubulin chaperones selected for analysis

I first assessed frequency changes in the overall unmatched cohort between baseline FIRSTANA and PROSELICA samples. The results are shown in (Supplemental Table 10-3). I found that:

- *TBCE* (tubulin beta chaperone E, chr1 q42.3) and *TUBB1* (tubulin beta 1 class VI, chr 20 q13.32) were copy-gained at significantly higher frequencies in PROSELICA (taxane pretreated) baseline samples compared to FIRSTANA (taxane naïve) baseline samples. The *TBCE* copy-gain frequency changed from 35% to 53%, (p=0.04), while *TUBB1* copy-gain frequency changed from 39% to 58%, (p=0.03).
- No significant changes in copy-loss frequency were observed, although *TUBE1* (tubulin epsilon 1, chr6 q21) losses increased from 33% in FIRS BL to 48% in PROS BL (but this was not statistically significantly, p=0.08).

I went on to examine the longitudinal changes in the FIRSTANA trial (between BL and EOS timepoints). No genes exhibited statistically significant changes, although I observed that:

- Copy-losses of *TUBB* (tubulin beta class 1, chr6 p21.33) increased from 11% to 24%, although this was a not a statistically significant change (p=0.08).
- Copy-gains of *TUBB8* increased from 7% to 18% (with a reduction in copy-losses of 39% to 27%), although this again was not statistically significant (p=0.088) (Supplemental Table 10-3).

No significant changes were observed in the PROS BL/EOS analysis, which is relatively unsurprising, as this cohort had already had prior taxane exposure.

Using McNemar's test for paired data sets to examine shifts in proportion, I analysed same-patient longitudinal matched samples: FIRSTANA had 29 pairs available, while PROSELICA had 43.

- In FIRSTANA, I found that: *TUBAL3* (tubulin alpha like 3, chr10 p15.1) had a significant shift in the frequency of copy losses, with 6 aberrations in FIRS found in baseline samples that were all absent by EOS (p=0.041).
- In the PROS cohort (n=43 matched sets) the *TUBA1A/B/C* locus (tubulin alpha isoforms a/b/c, chr12 q13.12) showed a marked increase in copy-gains between BL and EOS (p=0.027), with 9 additional events from baseline to EOS (Supplemental Table 10-4).

The limited number of paired samples may limit interpretation of these data.

6.3.2 Univariable and Multivariable Associations of Tubulin Genes with Taxanes

I performed univariable analyses of the gene set for both response and survival associations. As before, I included the 'Study' variable to account for biases due to varying response rates between cohorts. I did not distinguish between varying intensities of CNA in the subsequent analyses, ie. between hetero/homozygous deletions, or copy-gain/amplification. I made two modifications to the univariable models: including a variable 'Study' in all analyses, as both studies had different response rates; and for CNAs with extremely low event frequencies (ie. copy-losses or gains occurring in less than 10% of cases) these were merged with the 'Neutral' category, to avoid over-interpreting models with very sparse events.

	Raw Counts (Percentages)				Logistic Regression Results for Response Prediction					
	Losses		Gains		Losses			Gains		
Gene	Responders	Non-Responders	Responders	Non-Responders	OR	CI	p-value	OR	CI	p-value
TBCE	7 (38.9)	11 (61.1)	32 (51.6)	30 (48.4)	0.59	0.19-1.76	p=0.349	1.41	0.67-3.00	p=0.368
TUBA3D	22 (47.8)	24 (52.2)	12 (75.0)	4 (25.0)	1.19	0.56-2.54	p=0.650	4.38	1.33-17.43	p=0.021
TUBA4A	12 (46.2)	14 (53.8)	21 (61.8)	13 (38.2)	1.13	0.45-2.83	p=0.799	2.16	0.93-5.14	p=0.076
TBCA	24 (45.3)	29 (54.7)	13 (72.2)	5 (27.8)	1.04	0.50-2.18	p=0.913	2.67	0.87-9.29	p=0.097
TUBB2A	23 (52.3)	21 (47.7)	17 (70.8)	7 (29.2)	1.94	0.89-4.36	p=0.100	4.49	1.64-13.59	p=0.005
TUBB	12 (63.2)	7 (36.8)	20 (64.5)	11 (35.5)	3.15	1.10-9.71	p=0.036	2.87	1.21-7.15	p=0.019
TBCC	13 (61.9)	8 (38.1)	22 (73.3)	8 (26.7)	3.26	1.19-9.52	p=0.024	4.66	1.87-12.66	p=0.001
TUBE1	28 (48.3)	30 (51.7)	0 (NaN)	0 (NaN)	1.16	0.58-2.36	p=0.672	NA	NA	NA
TUBB4B	8 (44.4)	10 (55.6)	30 (62.5)	18 (37.5)	1.3	0.43-3.86	p=0.630	2.63	1.23-5.82	p=0.014
TUBB8	25 (43.9)	32 (56.1)	7 (46.7)	8 (53.3)	0.73	0.35-1.50	p=0.393	1.01	0.31-3.26	p=0.990
TUBAL3	26 (43.3)	34 (56.7)	7 (46.7)	8 (53.3)	0.68	0.33-1.39	p=0.291	0.97	0.29-3.14	p=0.955
TUBA1A	6 (40.0)	9 (60.0)	16 (48.5)	17 (51.5)	0.81	0.25-2.54	p=0.722	0.89	0.39-2.03	p=0.783
TUBA3C	27 (56.2)	21 (43.8)	13 (48.1)	14 (51.9)	1.8	0.84-3.97	p=0.136	1.39	0.55-3.57	p=0.489
TUBB3	54 (53.5)	47 (46.5)	0 (NaN)	0 (NaN)	2.18	1.03-4.80	p=0.046	NA	NA	NA
TUBG1	17 (60.7)	11 (39.3)	16 (55.2)	13 (44.8)	1.83	0.75-4.59	p=0.191	2.04	0.84-5.07	p=0.118
TUBD1	17 (65.4)	9 (34.6)	21 (50.0)	21 (50.0)	2.55	1.00-6.92	p=0.056	1.6	0.73-3.59	p=0.244
TBCD	10 (55.6)	8 (44.4)	29 (52.7)	26 (47.3)	1.66	0.57-5.05	p=0.357	1.62	0.78-3.44	p=0.200
TUBB6	16 (57.1)	12 (42.9)	22 (45.8)	26 (54.2)	1.85	0.73-4.81	p=0.198	1.04	0.48-2.25	p=0.928
TUBB4A	22 (53.7)	19 (46.3)	10 (43.5)	13 (56.5)	1.39	0.63-3.08	p=0.416	0.71	0.26-1.87	p=0.492
TBCB	12 (48.0)	13 (52.0)	16 (47.1)	18 (52.9)	1.04	0.41-2.66	p=0.929	0.87	0.38-2.00	p=0.748
TUBB1	0 (NaN)	0 (NaN)	35 (50.7)	34 (49.3)	NA	NA	NA	1.52	0.76-3.11	p=0.245
TUBA8	36 (55.4)	29 (44.6)	0 (NaN)	0 (NaN)	1.7	0.86-3.41	p=0.131	NA	NA	NA

Table 6-2 Univariable analyses of taxane-response associations for copy-gains and copy-losses in tubulin genes

I found that several tubulin family genes were univariably associated with drug **response** (Table 6-2).

- Copy-gains of the *TUBA3D/TUBA3E* locus (tubulin alpha 3D and 3E, chr2 q21.1, ~11% of cases) were significantly associated with better response with a response rate of 75% (p=0.021, OR 4.38).

- Similarly, copy gains of the *TUBB2A/TUBB2AB* locus (tubulin beta 2A and 2B, chr6 p25.2; ~16% of cases) were also associated with a very high response rate of 70.8% (p=0.005, OR 4.49).
- Copy-gains for both *TUBB* (chr6 p21.33, ~22% of cases) and *TBCC* (tubulin beta chaperone C, chr6 p21.1, ~20% of cases) also correlated with higher than expected response rates, at 64.5% and 73.3% respectively (p=0.019 and 0.001, ORs 2.87 and 4.66). Strangely, copy-losses for these two genes also exhibited a significant trend of improved response rates: 63.2% and 61.9% (p=0.036 and 0.024, ORs 3.15 and 3.26).
- Copy-gains of *TUBB4B* (tubulin beta 4B, chr9 q34.3, ~33% of cases) were also associated with significantly more responses with a rate of 62.5% (p=0.014, OR 2.63). Interestingly, *TUBB4B* copy-losses (~12% of cases) exhibited a worse OS (p=0.043 HR 1.73) and a similar but non-significant trend for RPFS (p=0.101 HR 1.76) – aligning with the response association.
- Conversely lower copy-number of *TUBB3* (tubulin beta 3, chr16 q24.3, ~70% of cases) was associated with improved responses, with a rate of 53.5% in the copy-lost group (p=0.046, OR 2.18).

	Cox Proportional Hazard Results for Overall Survival Prognosis						Cox Proportional Hazard Results for Radiographic Progression-Free Survival					
	Losses			Gains			Losses			Gains		
Gene	HR	CI	p-value	HR	CI	p-value	HR	CI	p-value	HR	CI	p-value
TBCE	1.3	0.74-2.28	p=0.360	1.25	0.85-1.83	p=0.262	1.02	0.47-2.22	p=0.954	1.06	0.64-1.74	p=0.829
TUBA3D	1	0.68-1.47	p=0.988	1.17	0.66-2.07	p=0.598	1.23	0.75-2.00	p=0.418	0.97	0.43-2.18	p=0.933
TUBA4A	1.83	1.15-2.89	p=0.010	1.05	0.68-1.62	p=0.829	2.56	1.47-4.46	p=0.001	0.71	0.39-1.27	p=0.245
TBCA	1.3	0.90-1.90	p=0.167	1.04	0.58-1.88	p=0.887	1.13	0.69-1.86	p=0.617	0.73	0.35-1.54	p=0.409
TUBB2A	1.28	0.86-1.90	p=0.229	1.19	0.72-1.97	p=0.494	1.46	0.88-2.44	p=0.144	1.24	0.67-2.30	p=0.490
TUBB	1.1	0.66-1.84	p=0.720	0.84	0.54-1.32	p=0.451	1.27	0.65-2.50	p=0.481	1.08	0.61-1.92	p=0.782
TBCC	1.24	0.76-2.03	p=0.392	0.86	0.55-1.35	p=0.521	1.31	0.68-2.51	p=0.420	0.94	0.53-1.66	p=0.837
TUBE1	1.06	0.74-1.52	p=0.736	NA	NA	NA	1.14	0.71-1.85	p=0.587	NA	NA	NA
TUBB4B	1.73	1.02-2.95	p=0.043	0.81	0.55-1.21	p=0.311	1.76	0.90-3.46	p=0.101	0.92	0.56-1.51	p=0.746
TUBB8	1.16	0.79-1.70	p=0.447	3.45	1.89-6.30	p<0.001	1.02	0.61-1.68	p=0.952	3.18	1.60-6.32	p=0.001
TUBAL3	1.26	0.86-1.84	p=0.234	3.62	1.98-6.63	p<0.001	1.13	0.69-1.86	p=0.630	3.34	1.67-6.68	p=0.001
TUBA1A	1.28	0.73-2.25	p=0.391	0.95	0.62-1.45	p=0.806	2	1.00-3.99	p=0.050	0.98	0.56-1.74	p=0.952
TUBA3C	1.17	0.79-1.74	p=0.437	1.02	0.63-1.64	p=0.942	1.32	0.78-2.21	p=0.298	1.16	0.64-2.09	p=0.624
TUBB3	1.06	0.72-1.58	p=0.758	NA	NA	NA	1.03	0.61-1.75	p=0.911	NA	NA	NA
TUBG1	0.95	0.60-1.51	p=0.824	1.59	1.02-2.48	p=0.042	0.85	0.46-1.60	p=0.623	1.42	0.82-2.46	p=0.206
TUBD1	0.96	0.59-1.55	p=0.868	1.64	1.09-2.45	p=0.017	0.82	0.42-1.58	p=0.545	1.5	0.90-2.49	p=0.123
TBCD	0.69	0.39-1.22	p=0.201	1.36	0.93-1.99	p=0.111	0.59	0.26-1.35	p=0.213	1.43	0.89-2.32	p=0.142
TUBB6	1.39	0.87-2.22	p=0.174	1.28	0.85-1.91	p=0.235	0.99	0.53-1.84	p=0.963	1	0.59-1.69	p=0.990
TUBB4A	0.8	0.53-1.21	p=0.297	1.75	1.08-2.84	p=0.023	0.94	0.53-1.67	p=0.835	2.58	1.48-4.52	p=0.001
TBCB	0.82	0.50-1.34	p=0.426	1.61	1.06-2.45	p=0.026	0.84	0.44-1.59	p=0.587	1.64	0.93-2.88	p=0.086
TUBB1	NA	NA	NA	0.84	0.59-1.20	p=0.343	NA	NA	NA	0.92	0.57-1.47	p=0.724
TUBA8	0.68	0.47-0.98	p=0.038	NA	NA	NA	1.09	0.69-1.72	p=0.710	NA	NA	NA

Table 6-3 Univariable analyses survival associations for copy-gains and copy-losses in tubulin genes

Some tubulins were also associated with **survival**, but not response.

- *TUBA4A* (tubulin alpha 4A, chr2 q35, ~18% of cases) copy-losses associated with significantly worse survival, for both OS and RPFS (p=0.01 and 0.001, HRs 1.83 and 2.56).

- Copy-gains of *TUBB8* and *TUBAL3* (both on chr10, p15.3 and p15.1 respectively, ~10% of cases) showed strong associations having significantly shorter OS and RPFS (p<=0.001 and HRs > 3).
- Copy-gains of TUBG1/TUBG2 and TUBD1 (both chr17, q21.2 and 23.1 respectively, ~33% of cases) were also associated with shorter OS (p=0.042 and 0.17, HRs 1.59 and 16) but not RPFS.
- TUBB4A (tubulin beta 4A, chr19 q13.3) copy-gains (~15% of cases) led to a worse OS (p=0.02, HR 1.75) and RPFS (p=0.001, HR 2.58). Similarly, copy-gains of the nearby TBCB (tubulin beta chaperone B, chr19 q13.12, ~24% of cases) were associated with significantly worse OS (p=0.026, HR 1.61), although the RPFS exhibited a non-significant trend (p=0.08, HR 1.64).
- Interestingly, and unlike many of the other survival associations in this study, deletions of TUBA8 (tubulin alpha 8, chr22 q11.21, ~45% of cases) actually resulted in an improved OS (p=0.038, HR 0.68), although RPFS was unaffected.

I went on to perform multivariable analyses of this gene set. I included in the initial model all genes displaying any significance in the univariable analysis, before performing stepwise model optimisation to generate the final model. Several genes were closely located, so to avoid collinearity I removed *TBCC* (same cytoband as *TUBB*) and *TUBAL3* (adjacent to *TUBB8*). The multivariable results are shown below in (Figure 6-3).

In the multivariable analyses, some genes exhibited significant trends for **response** association.

- Copy-gains of *TUBB4B* were again predictive of improved response (p=0.03, OR 3.19), along with a similar trend for OS (p=0.087, HR 0.65) and RPFS (p=0.016, HR 0.45). Decreased expression of *TUBB4B* has been associated with taxane resistance (Nami and Wang, 2018).
- *TUBA3D* copy-gains did not reach significance, although showed a trend for better response (p=0.15, OR 3.89), and had significantly improved RPFS (p=0.017, HR 0.18). Similarly to *TUBB4B*, low *TUBA3D* expression has been linked to paclitaxel resistance (Nami and Wang, 2018).
- The results for *TUBB8/TUBAL3* were counterintuitive: copy-losses were again associated with fewer responses (p=0.01, OR 0.26), but had no

significant association with OS or RPFS; copy-gains however, had no association with response, but showed a substantial association with poorer OS (p=0.009, HR 2.8) and RPFS (p=0.008, HR 3.5).

Changes in these genes may, therefore, offer a taxane-specific predictive clinical factor.

Several genes associated with altered **survival** in multivariable data but no change in response frequency:

- Of particular interest is *TUBB*, with copy-gains showing significantly improved OS (p=0.045, HR 0.44) and a borderline RPFS (p=0.063, HR 0.38).
- Copy-gains of *TUBB2A* were associated with poorer OS (p=0.022, HR 2.83) and RPFS (p=0.003, HR 5.49).
- *TUBB4A* copy-gains had strong correlation with better OS (p=0.029, HR 2.32) and RPFS (p<0.001, HR 14.09).

Changes in these genes may represent processes associated with advanced tumours, but not specific taxane-associated mechanisms.

A Taxane Drug Response

Variable		N	Odds ratio	p
TBCB	Neutral	85	Reference	
	Gain	34	0.82 (0.23, 2.97)	0.76
	Loss	25	0.75 (0.16, 3.40)	0.71
TUBA3D	Neutral	82	Reference	
	Gain	16	3.89 (0.67, 26.71)	0.15
	Loss	46	0.91 (0.32, 2.59)	0.86
TUBA4A	Neutral	84	Reference	
	Gain	34	1.12 (0.34, 3.64)	0.85
	Loss	26	0.79 (0.22, 2.70)	0.71
TUBA8	Neutral	79	Reference	
	Loss	65	1.95 (0.86, 4.55)	0.11
TUBB	Neutral	94	Reference	
	Gain	31	1.36 (0.32, 5.92)	0.67
	Loss	19	2.48 (0.52, 12.49)	0.26
TUBB2A	Neutral	76	Reference	
	Gain	24	2.25 (0.41, 13.26)	0.35
	Loss	44	1.31 (0.41, 4.21)	0.64
TUBB3	Neutral	43	Reference	
	Loss	101	2.19 (0.87, 5.71)	0.10
TUBB4A	Neutral	80	Reference	
	Gain	23	0.42 (0.09, 1.89)	0.26
	Loss	41	1.04 (0.27, 4.10)	0.95
TUBB4B	Neutral	78	Reference	
	Gain	48	3.19 (1.17, 9.33)	0.03
	Loss	18	1.58 (0.38, 6.68)	0.53
TUBB8	Neutral	72	Reference	
	Gain	15	0.40 (0.08, 1.79)	0.23
	Loss	57	0.26 (0.09, 0.71)	0.01
Purity		144	2.23 (0.06, 65.43)	0.65
Study	FIRS	75	Reference	
	PROS	69	0.23 (0.10, 0.50)	<0.001
(Intercept)			0.65 (0.22, 1.84)	0.42

B Overall Survival

Variable		N	Hazard ratio	p
TBCB	Neutral	85	Reference	
	Gain	34	1.06 (0.57, 1.96)	0.862
	Loss	25	0.84 (0.40, 1.73)	0.631
TUBA3D	Neutral	82	Reference	
	Gain	16	0.71 (0.30, 1.66)	0.430
	Loss	46	0.60 (0.36, 1.00)	0.051
TUBA4A	Neutral	84	Reference	
	Gain	34	1.03 (0.59, 1.81)	0.906
	Loss	26	1.76 (0.96, 3.23)	0.068
TUBA8	Neutral	79	Reference	
	Loss	65	0.80 (0.53, 1.21)	0.293
TUBB	Neutral	94	Reference	
	Gain	31	0.44 (0.19, 0.98)	0.045
	Loss	19	0.94 (0.44, 1.99)	0.873
TUBB2A	Neutral	76	Reference	
	Gain	24	2.83 (1.16, 6.90)	0.022
	Loss	44	1.41 (0.82, 2.42)	0.210
TUBB3	Neutral	43	Reference	
	Loss	101	1.26 (0.78, 2.02)	0.345
TUBB4A	Neutral	80	Reference	
	Gain	23	2.32 (1.09, 4.91)	0.029
	Loss	41	0.92 (0.48, 1.79)	0.812
TUBB4B	Neutral	78	Reference	
	Gain	48	0.65 (0.40, 1.06)	0.087
	Loss	18	1.63 (0.87, 3.06)	0.130
TUBB8	Neutral	72	Reference	
	Gain	15	2.80 (1.29, 6.09)	0.009
	Loss	57	1.35 (0.83, 2.19)	0.222
Purity		144	8.09 (1.47, 44.56)	0.016
Study	FIRS	75	Reference	
	PROS	69	1.89 (1.27, 2.80)	0.002

C Radiographic Progression-Free Survival

Variable		N	Hazard ratio	p
TBCB	Neutral	85	Reference	
	Gain	34	0.43 (0.20, 0.95)	0.036
	Loss	25	0.42 (0.16, 1.15)	0.092
TUBA3D	Neutral	82	Reference	
	Gain	16	0.18 (0.04, 0.73)	0.017
	Loss	46	0.67 (0.36, 1.24)	0.202
TUBA4A	Neutral	84	Reference	
	Gain	34	0.54 (0.25, 1.17)	0.118
	Loss	26	3.06 (1.45, 6.45)	0.003
TUBA8	Neutral	79	Reference	
	Loss	65	1.73 (0.97, 3.07)	0.063
TUBB	Neutral	94	Reference	
	Gain	31	0.38 (0.14, 1.06)	0.063
	Loss	19	1.71 (0.65, 4.46)	0.275
TUBB2A	Neutral	76	Reference	
	Gain	24	5.49 (1.79, 16.86)	0.003
	Loss	44	1.65 (0.80, 3.42)	0.178
TUBB3	Neutral	43	Reference	
	Loss	101	1.03 (0.54, 1.99)	0.919
TUBB4A	Neutral	80	Reference	
	Gain	23	14.09 (5.29, 37.53)	<0.001
	Loss	41	2.43 (0.91, 6.46)	0.075
TUBB4B	Neutral	78	Reference	
	Gain	48	0.45 (0.24, 0.86)	0.016
	Loss	18	1.41 (0.58, 3.41)	0.449
TUBB8	Neutral	72	Reference	
	Gain	15	3.50 (1.38, 8.86)	0.008
	Loss	57	0.69 (0.34, 1.38)	0.294
Purity		144	1.81 (0.16, 19.98)	0.629
Study	FIRS	75	Reference	
	PROS	69	1.82 (1.09, 3.03)	0.022

Figure 6-2 Drug response and survival multivariable models for tubulin genes

Forest plots for (A) taxane response status, (B) overall survival and (C) radiographic progression-free survival, showing per-gene CNA associations. Odds/hazard ratios are shown, along with confidence intervals and p-values. The variable 'Study' is included in both models to model the impact of samples drawn from cohorts with different hazards and response rates, and the variable 'Purity' is included to model the impact of high tumour burden. Blue arrows indicate improved response and prognosis; red arrows indicate worse response and prognosis

6.3.3 Key mCRPC Gene CNA Frequency Shifts

I next sought to profile genes that commonly feature copy number alterations in mCRPC (Williams et al., 2014). In addition, I expanded this list with a subset of genes from the PI3K-AKT and cell cycle pathways (Table 6-4). This is not an exhaustive list, but aims to cover a number of common mCRPC CNAs thoughout the genome.

Gene	Group	Chrom	Band
E2F2	Extra Cell Cycle	1	p36.12
CXCR4	Recurrent Del	2	q22.1
FOXP1	Recurrent Del	3	p13
RYBP	Recurrent Del	3	p13
SHQ1	Recurrent Del	3	p13
PIK3CB	Extra Pi3K	3	q22.3
PIK3CA	Recurrent Gains	3	q26.32
ETV5	Recurrent Gains	3	q27.2
PIK3R1	Extra Pi3K	5	q13.1
CDK7	Extra Cell Cycle	5	q13.2
CHD1	Recurrent Del	5	q21.1
APC	Recurrent Del	5	q22.2
E2F3	Extra Cell Cycle	6	p22.3
MAP3K7	Recurrent Del	6	q15
ETV1	Recurrent Gains	7	p21.2
EGFR	Recurrent Gains	7	p11.2
CDK6	Extra Cell Cycle	7	q21.2
MCM7	Recurrent Gains	7	q22.1
BRAF	Recurrent Gains	7	q34
CDK5	Extra Cell Cycle	7	q36.1
NKX3-1	Recurrent Del	8	p21.2
E2F5	Extra Cell Cycle	8	q21.2
MYC	Recurrent Gains	8	q24.21
CDKN2A	Extra Cell Cycle	9	p21.3
CDK9	Extra Cell Cycle	9	q34.11
CDK1	Extra Cell Cycle	10	q21.2
PTEN	Recurrent Del	10	q23.31
CDKN1B	Recurrent Del	12	p13.1
CDK2	Extra Cell Cycle	12	q13.2
CDK4	Extra Cell Cycle	12	q14.1
CDK8	Extra Cell Cycle	13	q12.13
BRCA2	Recurrent Del	13	q13.1
RB1	Recurrent Del	13	q14.2
AKT1	Extra Pi3K	14	q32.33
E2F4	Extra Cell Cycle	16	q22.1
CDH1	Recurrent Del	16	q22.1
TP53	Recurrent Del	17	p13.1
BRCA1	Recurrent Del	17	q21.31
ETV4	Recurrent Del	17	q21.31
CDK3	Extra Cell Cycle	17	q25.1
SMAD4	Recurrent Del	18	q21.2
BCL2	Recurrent Del	18	q21.33
E2F1	Extra Cell Cycle	20	q11.22
AR	Recurrent Gains	X	q12

Table 6-4 Genes assessed for taxane associations

Table of genes selected from four groups: recurrent copy gains/losses in CRPC, along with selected cell-cycle and pi3k/akt genes not featured in the other lists. Also shown are chromosome and cytoband information.

		Copy-Losses				Copy-Gains				Frequency Comparison (Fisher's Test) P-Values					
		Baseline		End of Study		Baseline		End of Study		Baseline FIRS/PROS		FIRS BL/EOS		PROS BL/EOS	
	Gene	FIRS	PROS	FIRS	PROS	FIRS	PROS	FIRS	PROS	Loss	Gain	Loss	Gain	Loss	Gain
1_22500001_23000000	E2F2	0.49	0.48	0.45	0.39	0.04	0.07	0.09	0.12	0.87	0.48	0.84	0.37	0.35	0.52
2_135500001_136000000	CXCR4	0.33	0.33	0.39	0.35	0.09	0.12	0.12	0.12	1.00	0.79	0.66	0.73	1.00	1.00
3_70000001_70500000	FOXP1	0.19	0.19	0.18	0.24	0.39	0.32	0.3	0.24	1.00	0.49	1.00	0.52	0.50	0.42
3_71000001_71500000	RYBP	0.2	0.2	0.18	0.27	0.37	0.32	0.3	0.24	1.00	0.60	1.00	0.52	0.51	0.42
3_71500001_72000000	SHQ1	0.21	0.2	0.18	0.29	0.36	0.32	0.3	0.24	1.00	0.73	0.80	0.66	0.38	0.42
3_137000001_137500000	PIK3CB	0.05	0.04	0	0.06	0.4	0.58	0.42	0.49	1.00	**0.04**	0.31	0.83	0.69	0.35
3_177500001_178000000	PIK3CA	0.08	0.1	0.06	0.06	0.39	0.51	0.36	0.47	0.77	0.18	1.00	1.00	0.52	0.71
3_184500001_185000000	ETV5	0.07	0.1	0.06	0.06	0.43	0.49	0.42	0.47	0.55	0.50	1.00	1.00	0.52	0.85
5_66500001_67000000	PIK3R1	0.37	0.45	0.39	0.51	0.15	0.03	0.15	0.08	0.40	**0.02**	0.83	1.00	0.58	0.23
5_68530622_68573257	CDK7	0.36	0.43	0.42	0.49	0.15	0.03	0.15	0.08	0.40	**0.02**	0.53	1.00	0.58	0.23
5_97000001_97500000	CHD1	0.36	0.43	0.36	0.43	0.19	0.1	0.15	0.14	0.40	0.16	1.00	0.79	1.00	0.57
5_111000001_111500000	APC	0.31	0.32	0.27	0.39	0.24	0.14	0.21	0.12	1.00	0.21	0.82	0.81	0.56	0.79
6_19000001_19500000	E2F3	0.16	0.23	0.27	0.2	0.19	0.2	0.21	0.14	0.30	0.84	0.19	0.80	0.82	0.47
6_90000001_90500000	MAP3K7	0.41	0.61	0.24	0.51	0.09	0	0.21	0.04	**0.02**	**0.01**	0.13	0.12	0.35	0.17
7_12500001_13000000	ETV1	0.13	0.07	0.09	0.08	0.4	0.57	0.45	0.51	0.28	0.07	0.75	0.67	1.00	0.58
7_54000001_54500000	EGFR	0.11	0.01	0.03	0.08	0.4	0.59	0.42	0.51	**0.03**	**0.03**	0.27	0.83	0.16	0.45
7_91000001_91500000	CDK6	0.04	0.03	0	0.04	0.45	0.55	0.45	0.51	1.00	0.32	0.55	1.00	1.00	0.71
7_98500001_99000000	MCM7	0.04	0.04	0	0.06	0.47	0.54	0.45	0.45	1.00	0.50	0.55	1.00	0.69	0.46
7_139000001_139500000	BRAF	0.05	0.06	0	0.04	0.51	0.58	0.55	0.53	1.00	0.41	0.31	0.84	1.00	0.71
7_149500001_150000000	CDK5	0.08	0.09	0	0.14	0.47	0.54	0.55	0.51	1.00	0.50	0.17	0.53	0.38	0.85
8_22500001_23000000	NKX3-1	0.75	0.8	0.64	0.73	0.03	0.03	0.06	0.06	0.55	1.00	0.26	0.58	0.51	0.65
8_85000001_85500000	E2F5	0.03	0	0	0	0.76	0.78	0.76	0.73	0.50	0.84	1.00	1.00	1.00	0.66
8_127500001_128000000	MYC	0	0	0	0	0.77	0.84	0.79	0.78	1.00	0.40	1.00	1.00	1.00	0.47
9_20500001_21000000	CDKN2A	0.23	0.28	0.36	0.33	0.27	0.25	0.09	0.27	0.57	0.85	0.16	**0.04**	0.68	0.83
9_130547958_130553066	CDK9	0.08	0.14	0.18	0.24	0.36	0.39	0.21	0.35	0.29	0.73	0.18	0.18	0.23	0.70
10_61500001_62000000	CDK1	0.27	0.23	0.24	0.2	0.16	0.3	0.15	0.29	0.70	0.05	1.00	1.00	0.82	1.00
10_88500001_89000000	PTEN	0.42	0.52	0.33	0.51	0.07	0.07	0.1	0.07	0.29	1.00	0.50	0.70	1.00	1.00
12_11500001_12000000	CDKN1B	0.25	0.26	0.21	0.27	0.19	0.22	0.21	0.33	1.00	0.68	0.81	0.80	1.00	0.21
12_55000001_55500000	CDK2	0.05	0.12	0.15	0.19	0.27	0.24	0.3	0.3	0.23	0.70	0.13	0.82	0.30	0.52
12_57000001_57500000	CDK4	0.07	0.1	0.12	0.16	0.27	0.25	0.33	0.37	0.55	0.85	0.45	0.50	0.40	0.22
13_26828276_26979375	CDK8	0.35	0.32	0.45	0.27	0.13	0.25	0.09	0.27	0.86	0.09	0.29	0.75	0.55	0.83
13_31500001_32000000	BRCA2	0.41	0.46	0.48	0.37	0.13	0.14	0.06	0.2	0.61	1.00	0.53	0.34	0.35	0.46
13_47500001_48000000	RB1	0.6	0.75	0.55	0.69	0.05	0.01	0.03	0.02	0.05	0.37	0.67	1.00	0.53	1.00
14_104000001_104500000	AKT1	0.13	0.06	0.15	0.1	0.31	0.38	0.39	0.31	0.16	0.39	0.77	0.39	0.49	0.44
16_66000001_66500000	E2F4	0.39	0.42	0.48	0.35	0.16	0.17	0.03	0.18	0.74	0.83	0.40	0.10	0.45	1.00
16_67500001_68000000	CDH1	0.41	0.43	0.48	0.35	0.15	0.14	0.03	0.18	0.87	1.00	0.53	0.10	0.35	0.62
17_6500001_7000000	TP53	0.67	0.52	0.64	0.59	0.01	0.09	0.06	0.06	0.09	0.05	0.83	0.22	0.46	0.73
17_40000001_40500000	BRCA1	0.25	0.13	0.33	0.2	0.15	0.26	0.21	0.22	0.09	0.10	0.49	0.41	0.32	0.67
17_40500001_41000000	ETV4	0.25	0.13	0.33	0.2	0.15	0.26	0.21	0.22	0.09	0.10	0.49	0.41	0.32	0.67
17_72500001_73000000	CDK3	0.15	0.1	0.24	0.12	0.33	0.43	0.3	0.49	0.46	0.23	0.28	0.83	0.77	0.58
18_47000001_47500000	SMAD4	0.37	0.48	0.45	0.49	0.16	0.13	0.12	0.04	0.24	0.64	0.52	0.77	1.00	0.12
18_59500001_60000000	BCL2	0.45	0.46	0.48	0.49	0.15	0.16	0.12	0.08	1.00	1.00	0.84	1.00	0.85	0.27
20_31000001_31500000	E2F1	0.11	0.14	0.06	0.16	0.29	0.35	0.27	0.33	0.62	0.59	0.72	1.00	0.80	0.85
X_65500001_66000000	AR	0.15	0.13	0.21	0.08	0.48	0.46	0.42	0.59	0.81	0.87	0.41	0.68	0.55	0.19

Table 6-5 Frequencies of CNAs in key mCRPC genes between cohorts, with fisher's test p-values shown

I observed significant differences in CNA frequencies between the FIRS (n=75) and PROS (n=69) baseline samples in several genes (Table 6-5):

- *PIK3CB* (chr3 q22.3) copy gains: 40% FIRS to 58% PROS
- *PIK3R1* and *CDK7* (chr5 q13.1) copy gains: 15% to 3%
- *MAP3K7* (chr6 g15) copy losses: 41% to 61%, copy gains: 9% to 0%
- *EGFR* (chr7 p11.2) copy losses: 11% to 1%, copy-gains: 40% to 59%
- *CDK1* (chr10 q21.2) copy gains: 16% to 30%

PIK3R1, CDK7, *MAP3K7*, and *CDK1* trended towards more losses and fewer gains in PROS than FIRS, while *PIK3CB* and *EGFR* exhibited more gains and fewer losses.

The only significant frequency change between FIRS baseline (n=75) to end of study (n=33) was *CDKN2A* (chr9 p21.3), which showed increased deletions (BL 23% to EOS 36%)and significantly fewer copy-gains (BL 27% to EOS 9%)

No genes in the shortlist showed differences in the PROS baseline (n=69) *vs* end of study (n=49) comparison, however.

I went on to examine frequency changes in copy-states in a matched same-patient set of cases (baseline and end of study, n=72). There was a significant shift in the proportion of copy-gains for the *CDK4* (chr12 q14.1) gene in the PROSELICA cohort, with 9 cases displaying new copy-gain events by the end of treatment (out of 43 pairs) (Supplemental Table 10-2).

6.3.4 Key mCRPC Gene Univariable and Multivariable Associations with Taxanes

	Raw Counts (Percentages)				Logistic Regression Results for Response Prediction					
	Losses		Gains		Losses			Gains		
Gene	Responders	Non-Responders	Responders	Non-Responders	OR	CI	p-value	OR	CI	p-value
E2F2	35 (50.0)	35 (50.0)	0 (NaN)	0 (NaN)	1.1	0.56-2.18	p=0.774	NA	NA	NA
CXCR4	23 (47.9)	25 (52.1)	11 (73.3)	4 (26.7)	1.18	0.56-2.50	p=0.667	4.1	1.22-16.51	p=0.030
FOXP1	12 (44.4)	15 (55.6)	30 (58.8)	21 (41.2)	1.05	0.41-2.68	p=0.916	1.85	0.86-4.03	p=0.116
RYBP	13 (44.8)	16 (55.2)	29 (58.0)	21 (42.0)	1.05	0.42-2.62	p=0.918	1.77	0.82-3.86	p=0.149
SHQ1	14 (46.7)	16 (53.3)	28 (57.1)	21 (42.9)	1.12	0.45-2.75	p=0.812	1.72	0.79-3.77	p=0.171
PIK3CB	Insufficient #	Insufficient #	36 (51.4)	34 (48.6)	NA	NA	NA	1.59	0.79-3.25	p=0.196
PIK3CA	Insufficient #	Insufficient #	32 (50.0)	32 (50.0)	NA	NA	NA	1.29	0.65-2.59	p=0.475
ETV5	Insufficient #	Insufficient #	33 (50.0)	33 (50.0)	NA	NA	NA	1.21	0.61-2.40	p=0.591
PIK3R1	29 (49.2)	30 (50.8)	Insufficient #	Insufficient #	1.14	0.57-2.29	p=0.713	NA	NA	NA
CDK7	27 (47.4)	30 (52.6)	Insufficient #	Insufficient #	1	0.50-2.01	p=0.999	NA	NA	NA
CHD1	29 (50.9)	28 (49.1)	15 (71.4)	6 (28.6)	1.73	0.82-3.68	p=0.151	3.55	1.23-11.40	p=0.024
APC	27 (60.0)	18 (40.0)	16 (57.1)	12 (42.9)	2.52	1.15-5.66	p=0.023	1.91	0.76-4.87	p=0.169
E2F3	16 (57.1)	12 (42.9)	19 (67.9)	9 (32.1)	2.58	1.04-6.66	p=0.044	3.87	1.53-10.55	p=0.006
MAP3K7	40 (54.8)	33 (45.2)	Insufficient #	Insufficient #	2.28	1.12-4.82	p=0.026	NA	NA	NA
ETV1	7 (46.7)	8 (53.3)	31 (44.9)	38 (55.1)	0.68	0.21-2.22	p=0.526	0.82	0.40-1.71	p=0.602
EGFR	Insufficient #	Insufficient #	34 (47.9)	37 (52.1)	NA	NA	NA	1.18	0.59-2.39	p=0.637
CDK6	Insufficient #	Insufficient #	36 (50.0)	36 (50.0)	NA	NA	NA	1.26	0.64-2.52	p=0.503
MCM7	Insufficient #	Insufficient #	37 (51.4)	35 (48.6)	NA	NA	NA	1.38	0.70-2.76	p=0.356
BRAF	Insufficient #	Insufficient #	40 (51.3)	38 (48.7)	NA	NA	NA	1.4	0.71-2.81	p=0.334
CDK5	Insufficient #	Insufficient #	35 (48.6)	37 (51.4)	NA	NA	NA	1.08	0.55-2.15	p=0.818
NKX3-1	54 (48.6)	57 (51.4)	Insufficient #	Insufficient #	1.09	0.49-2.47	p=0.830	NA	NA	NA
E2F5	Insufficient #	Insufficient #	52 (46.8)	59 (53.2)	NA	NA	NA	0.74	0.33-1.67	p=0.474
MYC	Insufficient #	Insufficient #	58 (50.0)	58 (50.0)	NA	NA	NA	1.55	0.65-3.77	p=0.326
CDKN2A	17 (47.2)	19 (52.8)	21 (56.8)	16 (43.2)	1.18	0.51-2.74	p=0.698	1.65	0.72-3.84	p=0.239
CDK9	8 (50.0)	8 (50.0)	30 (55.6)	24 (44.4)	1.7	0.54-5.44	p=0.360	1.86	0.89-3.96	p=0.102
CDK1	15 (41.7)	21 (58.3)	14 (42.4)	19 (57.6)	0.58	0.25-1.33	p=0.202	0.75	0.31-1.78	p=0.510
PTEN	28 (46.7)	32 (53.3)	Insufficient #	Insufficient #	0.93	0.47-1.86	p=0.841	NA	NA	NA
CDKN1B	18 (48.6)	19 (51.4)	14 (48.3)	15 (51.7)	1.03	0.45-2.32	p=0.951	1.05	0.43-2.55	p=0.921
CDK2	Insufficient #	Insufficient #	18 (50.0)	18 (50.0)	NA	NA	NA	1.03	0.47-2.25	p=0.946
CDK4	Insufficient #	Insufficient #	19 (51.4)	18 (48.6)	NA	NA	NA	1.14	0.52-2.48	p=0.746
CDK8	27 (56.2)	21 (43.8)	12 (44.4)	15 (55.6)	1.68	0.78-3.67	p=0.189	1.23	0.48-3.17	p=0.666
BRCA2	32 (50.8)	31 (49.2)	9 (45.0)	11 (55.0)	1.24	0.60-2.60	p=0.564	0.96	0.33-2.75	p=0.935
RB1	50 (51.5)	47 (48.5)	Insufficient #	Insufficient #	1.85	0.88-3.99	p=0.109	NA	NA	NA
AKT1	Insufficient #	Insufficient #	22 (44.9)	27 (55.1)	NA	NA	NA	0.86	0.42-1.76	p=0.674
E2F4	29 (50.0)	29 (50.0)	12 (50.0)	12 (50.0)	1.22	0.58-2.58	p=0.608	1.22	0.45-3.27	p=0.696
CDH1	31 (50.8)	30 (49.2)	9 (42.9)	12 (57.1)	1.14	0.55-2.39	p=0.724	0.79	0.28-2.23	p=0.663
TP53	44 (51.2)	42 (48.8)	Insufficient #	Insufficient #	1.11	0.55-2.23	p=0.772	NA	NA	NA
BRCA1	17 (60.7)	11 (39.3)	16 (55.2)	13 (44.8)	1.83	0.75-4.59	p=0.191	2.04	0.84-5.07	p=0.118
ETV4	17 (60.7)	11 (39.3)	16 (55.2)	13 (44.8)	1.83	0.75-4.59	p=0.191	2.04	0.84-5.07	p=0.118
CDK3	11 (61.1)	7 (38.9)	28 (50.9)	27 (49.1)	1.99	0.67-6.21	p=0.218	1.53	0.73-3.24	p=0.260
SMAD4	29 (47.5)	32 (52.5)	10 (47.6)	11 (52.4)	1.02	0.49-2.13	p=0.967	0.9	0.32-2.51	p=0.833
BCL2	29 (43.9)	37 (56.1)	12 (54.5)	10 (45.5)	0.73	0.35-1.53	p=0.404	1.17	0.42-3.34	p=0.759
E2F1	9 (50.0)	9 (50.0)	26 (56.5)	20 (43.5)	1.52	0.52-4.53	p=0.445	1.94	0.90-4.25	p=0.093
AR	10 (50.0)	10 (50.0)	33 (48.5)	35 (51.5)	1.02	0.35-2.97	p=0.971	0.98	0.47-2.05	p=0.957

Table 6-6 Univariable analyses of taxane-response associations for copy-gains and copy-losses in selected genes

To explore the impact on drug response and survival, I performed univariable analyses using logistic regressions and Cox' proportional-hazard models, respectively (Table 6-6 and Table 6-7). Several genes were associated with drug **response**.

- In this analysis, copy-losses in the *APC* (chr5 q22.2) and *MAP3K7* (chr6 q15) loci were associated with improved taxane response status, with 60% and 54% response rates (p=0.02 and 0.026, ORs 2.52 and 2.58).
- Copy-gains of the *CXCR4* (chr2 q22.1) and *CHD1* (chr5 q21.1) loci were also significantly associated with improved drug response, with 68%, 71% responses in these groups (p=0.03 and 0.024, OR 4.1 and 3.55).
- *E2F3* (chr6 p22.3) locus copy-gains were associated with an improved response rate (67%, p=0.006, OR 3.87) – and similarly, losses were

significantly associated with poorer progression-free survival (p=0.04, HR 1.72).

	Cox Proportional Hazard Results for Overall Survival Prognosis						Cox Proportional Hazard Results for Radiographic Progression-Free Survival					
	Losses			Gains			Losses			Gains		
Gene	HR	CI	p-value	HR	CI	p-value	HR	CI	p-value	HR	CI	p-value
E2F2	0.85	0.59-1.21	p=0.357	NA	NA	NA	0.93	0.59-1.45	p=0.739	NA	NA	NA
CXCR4	1.1	0.75-1.61	p=0.631	1.13	0.62-2.05	p=0.685	1.12	0.69-1.81	p=0.646	0.84	0.35-2.00	p=0.695
FOXP1	1.3	0.81-2.09	p=0.277	0.97	0.65-1.44	p=0.877	1.74	0.94-3.23	p=0.079	1.08	0.65-1.79	p=0.763
RYBP	1.36	0.86-2.16	p=0.191	0.98	0.66-1.47	p=0.925	1.74	0.94-3.23	p=0.079	1.08	0.65-1.79	p=0.763
SHQ1	1.41	0.89-2.22	p=0.144	0.96	0.64-1.44	p=0.842	1.81	0.99-3.31	p=0.055	1.05	0.63-1.75	p=0.844
PIK3CB	NA	NA	NA	1.07	0.75-1.53	p=0.697	NA	NA	NA	1.14	0.72-1.79	p=0.584
PIK3CA	NA	NA	NA	1.31	0.92-1.87	p=0.136	NA	NA	NA	1.28	0.81-2.02	p=0.292
ETV5	NA	NA	NA	1.42	0.99-2.02	p=0.054	NA	NA	NA	1.34	0.85-2.11	p=0.210
PIK3R1	1.15	0.81-1.64	p=0.438	NA	NA	NA	1.15	0.72-1.81	p=0.561	NA	NA	NA
CDK7	1.26	0.88-1.80	p=0.201	NA	NA	NA	1.12	0.70-1.79	p=0.628	NA	NA	NA
CHD1	1.28	0.88-1.87	p=0.200	1.06	0.61-1.85	p=0.826	1.07	0.65-1.75	p=0.791	0.86	0.44-1.68	p=0.650
APC	1.39	0.94-2.06	p=0.102	1.21	0.74-1.98	p=0.456	1.24	0.75-2.07	p=0.403	1.08	0.59-1.97	p=0.807
E2F3	1.18	0.76-1.85	p=0.462	0.88	0.54-1.42	p=0.595	1.74	1.00-3.01	p=0.048	1.07	0.58-1.98	p=0.826
MAP3K7	0.94	0.66-1.34	p=0.721	NA	NA	NA	0.9	0.57-1.42	p=0.651	NA	NA	NA
ETV1	1.7	0.93-3.09	p=0.083	1.15	0.79-1.69	p=0.469	2.18	1.02-4.66	p=0.045	1.07	0.65-1.75	p=0.799
EGFR	NA	NA	NA	1.19	0.83-1.71	p=0.336	NA	NA	NA	1	0.63-1.59	p=0.990
CDK6	NA	NA	NA	1.17	0.82-1.66	p=0.391	NA	NA	NA	0.97	0.62-1.53	p=0.905
MCM7	NA	NA	NA	1.14	0.80-1.61	p=0.478	NA	NA	NA	0.97	0.62-1.52	p=0.900
BRAF	NA	NA	NA	1.16	0.82-1.66	p=0.400	NA	NA	NA	1.07	0.67-1.68	p=0.785
CDK5	NA	NA	NA	1.2	0.85-1.71	p=0.302	NA	NA	NA	1.08	0.69-1.69	p=0.739
NKX3-1	0.97	0.64-1.49	p=0.899	NA	NA	NA	0.63	0.37-1.06	p=0.083	NA	NA	NA
E2F5	NA	NA	NA	1.6	1.03-2.51	p=0.038	NA	NA	NA	1.69	0.96-2.95	p=0.067
MYC	NA	NA	NA	1.26	0.78-2.01	p=0.344	NA	NA	NA	1.21	0.68-2.18	p=0.517
CDKN2A	1.62	1.06-2.48	p=0.026	0.95	0.62-1.47	p=0.826	1.52	0.88-2.63	p=0.135	1.23	0.71-2.12	p=0.458
CDK9	1.65	0.95-2.90	p=0.078	1.02	0.70-1.50	p=0.905	1.8	0.86-3.77	p=0.118	1.29	0.80-2.09	p=0.292
CDK1	1.12	0.73-1.71	p=0.613	1.17	0.75-1.83	p=0.486	0.83	0.47-1.46	p=0.521	1.29	0.73-2.28	p=0.380
PTEN	0.89	0.62-1.28	p=0.538	NA	NA	NA	0.73	0.45-1.20	p=0.219	NA	NA	NA
CDKN1B	1.17	0.77-1.79	p=0.457	1.06	0.67-1.66	p=0.814	1.51	0.88-2.59	p=0.137	0.76	0.41-1.40	p=0.373
CDK2	NA	NA	NA	1.02	0.68-1.52	p=0.938	NA	NA	NA	1.02	0.60-1.72	p=0.947
CDK4	NA	NA	NA	0.9	0.61-1.35	p=0.617	NA	NA	NA	0.96	0.57-1.61	p=0.864
CDK8	1.16	0.78-1.73	p=0.459	1.02	0.63-1.65	p=0.946	1.32	0.79-2.22	p=0.291	1.2	0.66-2.15	p=0.553
BRCA2	1.27	0.87-1.86	p=0.217	1.27	0.74-2.18	p=0.390	1.41	0.87-2.30	p=0.165	1.68	0.83-3.38	p=0.150
RB1	1.29	0.86-1.92	p=0.215	NA	NA	NA	0.99	0.60-1.65	p=0.978	NA	NA	NA
AKT1	NA	NA	NA	1.25	0.86-1.82	p=0.235	NA	NA	NA	1.37	0.85-2.21	p=0.190
E2F4	1.59	1.08-2.35	p=0.019	1.73	1.03-2.90	p=0.038	2	1.20-3.34	p=0.008	2.33	1.20-4.51	p=0.012
CDH1	1.58	1.08-2.32	p=0.019	1.8	1.04-3.10	p=0.035	2.03	1.22-3.36	p=0.006	2.11	1.05-4.24	p=0.035
TP53	1.21	0.84-1.76	p=0.311	NA	NA	NA	1.1	0.69-1.76	p=0.675	NA	NA	NA
BRCA1	0.95	0.60-1.51	p=0.824	1.59	1.02-2.48	p=0.042	0.85	0.46-1.60	p=0.623	1.42	0.82-2.46	p=0.206
ETV4	0.95	0.60-1.51	p=0.824	1.59	1.02-2.48	p=0.042	0.85	0.46-1.60	p=0.623	1.42	0.82-2.46	p=0.206
CDK3	0.61	0.34-1.09	p=0.094	1.45	0.99-2.11	p=0.056	0.47	0.20-1.15	p=0.097	1.53	0.95-2.47	p=0.083
SMAD4	1.24	0.84-1.83	p=0.272	1.33	0.78-2.27	p=0.297	1.09	0.65-1.85	p=0.742	1.08	0.56-2.07	p=0.826
BCL2	1.32	0.90-1.95	p=0.156	1.14	0.66-1.96	p=0.636	1.05	0.62-1.78	p=0.858	1.33	0.71-2.49	p=0.380
E2F1	1.5	0.89-2.55	p=0.131	1.08	0.73-1.59	p=0.712	2.29	1.20-4.35	p=0.011	1.03	0.62-1.71	p=0.905
AR	0.93	0.54-1.61	p=0.789	1.2	0.82-1.76	p=0.357	0.7	0.32-1.55	p=0.380	1.25	0.77-2.03	p=0.361

Table 6-7 Univariable analyses of survival associations for copy-gains and copy-losses in selected genes

Some genes were univariably associated with **survival**, but not drug response. In particular:

- Copy-losses in *CDKN2A* (chr9 p21.3) were associated with substantially shorter OS (p=0.02, HR 1.62)
- Copy-losses of *E2F4* and *CDH1* (chr16 q22.1) were associated with a shorter OS (p=0.019, HR 1.59) and RPFS (p=0.008, HR 2).
- Interestingly, in addition to copy-losses, copy-gains of the *E2F4*/*CDH1* locus were also associated with worse OS (p=0.038, HR 1.73) and RPFS (p=0.01, HR 2.3).

I used any statistically significant genes as candidates for multivariable analysis. *BRCA1*/*ETV4* and *E2F4*/*CDH1* are effectively adjacent, so to avoid collinearity in

the multivariable regression models I removed the *ETV4* and *CDH1* variables. Any result for *BRCA1* and *E2F4* can be considered essentially identical for their respective removed variable. I constructed initial models before optimising them using stepwise selection. As I have shown that it is highly prognostic in chapter 4, I included tumour purity in the final multivariable models as a way to account for tumour burden (Figure 6-3).

Several genes remained associated with changes in **response** in the multivariable analysis (Figure 6-3).

- Copy-losses of the *MAP3K7* (chr6 q15) locus were significantly predictive (p=0.0494) of drug response, with an odds ratio of 2.55, and this was supported with a trend for better OS (p=0.066, HR 0.67) and significantly better RPFS (p=0.012, HR 0.49) (Figure 6-3). Interestingly, *MAP3K7* deletions have been implicated in aggressive prostate cancers, but in this study performed better than expected (Rodrigues et al., 2015).
- Copy-gains of *CXCR4* (chr2 q22.1) associated with a non-significant improvement in drug response (p=0.12, OR 3.35) and OS (p=0.388, HR 0.74), but RPFS was significantly longer (p=0.033, HR 0.32).
- Copy-gains of *ETV1* (chr7 p21.2) associated with a significantly lower drug response rate (p=0.02, OR 0.28) but offered no changes to survival.

For changes in **survival**, relatively few genes were multivariably significant:

- Copy-gains of the *E2F4/CDH1* locus (chr16 q22.1) had a non-significant trend for lower drug response (p=0.14, OR 0.38) but a significant association with worse OS (p=0.025, HR 1.96) and RPFS (p=0.004, HR 3.06). Copy-losses of this locus were also associated with poorer RPFS (p=0.009, HR 2.08), but not with a change in response or OS.
- The results for *E2F3* (chr6 p22.3) were counterintuitive, however: both copy-gains and losses were associated with improved response (p=0.01 and 0.05, OR 4.67 and 3.23), with a non-significant impact on OS, and copy-losses associating with worse RPFS (p=0.018, HR 2.33).
- As before, *CDKN2A* losses were associated with significantly worse OS (p=0.026, HR 1.73) but not RPFS.

A Taxane Drug Response

Variable		N	Odds ratio	p
CDKN2A	Neutral	71	Reference	
	Gain	37	0.92 (0.31, 2.68)	0.88
	Loss	36	0.95 (0.34, 2.62)	0.92
CXCR4	Neutral	81	Reference	
	Gain	15	3.35 (0.78, 16.78)	0.12
	Loss	48	0.64 (0.24, 1.63)	0.35
E2F1	Neutral	80	Reference	
	Gain	46	1.85 (0.72, 4.85)	0.20
	Loss	18	1.27 (0.37, 4.47)	0.70
E2F3	Neutral	88	Reference	
	Gain	28	4.67 (1.49, 16.31)	0.01
	Loss	28	3.23 (1.02, 10.94)	0.05
E2F4	Neutral	62	Reference	
	Gain	24	0.38 (0.10, 1.35)	0.14
	Loss	58	0.97 (0.41, 2.32)	0.95
ETV1	Neutral	60	Reference	
	Gain	69	0.32 (0.12, 0.81)	0.02
	Loss	15	0.26 (0.06, 1.02)	0.06
MAP3K7	Neutral	71	Reference	
	Loss	73	2.55 (1.02, 6.73)	0.05
Purity		144	2.82 (0.10, 75.44)	0.53
Study	FIRS	75	Reference	
	PROS	69	0.20 (0.09, 0.45)	<0.001
(Intercept)			1.22 (0.51, 2.96)	0.66

B Overall Survival

Variable		N	Hazard ratio	p
CDKN2A	Neutral	71	Reference	
	Gain	37	0.95 (0.55, 1.64)	0.850
	Loss	36	1.73 (1.07, 2.81)	0.026
CXCR4	Neutral	81	Reference	
	Gain	15	0.74 (0.37, 1.47)	0.388
	Loss	48	0.96 (0.60, 1.54)	0.881
E2F1	Neutral	80	Reference	
	Gain	46	0.80 (0.49, 1.30)	0.367
	Loss	18	1.31 (0.71, 2.42)	0.394
E2F3	Neutral	88	Reference	
	Gain	28	1.10 (0.60, 1.99)	0.759
	Loss	28	1.26 (0.73, 2.19)	0.402
E2F4	Neutral	62	Reference	
	Gain	24	1.96 (1.09, 3.53)	0.025
	Loss	58	1.50 (0.98, 2.30)	0.062
ETV1	Neutral	60	Reference	
	Gain	69	1.12 (0.73, 1.72)	0.618
	Loss	15	1.47 (0.77, 2.79)	0.238
MAP3K7	Neutral	71	Reference	
	Loss	73	0.67 (0.44, 1.03)	0.066
Purity		144	8.68 (1.79, 42.24)	0.007
Study	FIRS	75	Reference	
	PROS	69	1.86 (1.26, 2.73)	0.002

C Radiographic Progression-Free Survival

Variable		N	Hazard ratio	p
CDKN2A	Neutral	71	Reference	
	Gain	37	0.95 (0.47, 1.92)	0.897
	Loss	36	1.60 (0.84, 3.06)	0.156
CXCR4	Neutral	81	Reference	
	Gain	15	0.32 (0.11, 0.91)	0.033
	Loss	48	0.92 (0.50, 1.69)	0.786
E2F1	Neutral	80	Reference	
	Gain	46	1.17 (0.64, 2.16)	0.610
	Loss	18	2.23 (1.08, 4.59)	0.030
E2F3	Neutral	88	Reference	
	Gain	28	1.39 (0.66, 2.96)	0.389
	Loss	28	2.59 (1.27, 5.29)	0.009
E2F4	Neutral	62	Reference	
	Gain	24	3.06 (1.43, 6.57)	0.004
	Loss	58	1.89 (1.08, 3.31)	0.026
ETV1	Neutral	60	Reference	
	Gain	69	0.91 (0.52, 1.59)	0.730
	Loss	15	1.56 (0.67, 3.63)	0.298
MAP3K7	Neutral	71	Reference	
	Loss	73	0.49 (0.28, 0.86)	0.012
Purity		144	5.65 (0.65, 49.09)	0.117
Study	FIRS	75	Reference	
	PROS	69	1.59 (0.96, 2.64)	0.071

Figure 6-3 Refined drug response and survival multivariable models for specific genes

Forest plots for (A) taxane response status, (B) overall survival and (C) radiographic progression-free survival, showing per-gene CNA assocations. Odds/hazard ratios are shown, along with confidence intervals and p-values. The variable 'Study' is included in both models to model the impact of samples drawn from cohorts with different hazards and response rates, and the variable 'Purity' is included to model the impact of high tumour burden. Blue arrows indicate improved response and prognosis; red arrows indicate worse response and prognosis.

6.3.5 *RB1* and Partners and Clinical Outcomes

My analyses above suggested that cell cycle gene alterations may emerge during taxane chemotherapy, with new CDKN2A deletions and CDK4 gains. Some reports have also indicated that *RB1* (chr13 q14.2) deletions may play a role in taxane drug resistance, so I therefore extended the analyses to examine this gene specifically, while further studying the closely linked cell cycle genes *CDK4* (chr12 q14.1), *CDK6* (chr7 21.2) and *CDKN2A* (chr9 p21.3) (Hamid et al., 2019; Liu et al., 2015). Surprisingly, in multivariable analyses, none associated with drug response or RPFS, although *CDKN2A* deletion significantly associated with OS (p=0.022, HR 1.7) (Figure 6-4). The sub-clonal emergence of *CDKN2A* during chemotherapy and castration in this cohort does nonetheless imply a relevance to mCRPC disease evolution.

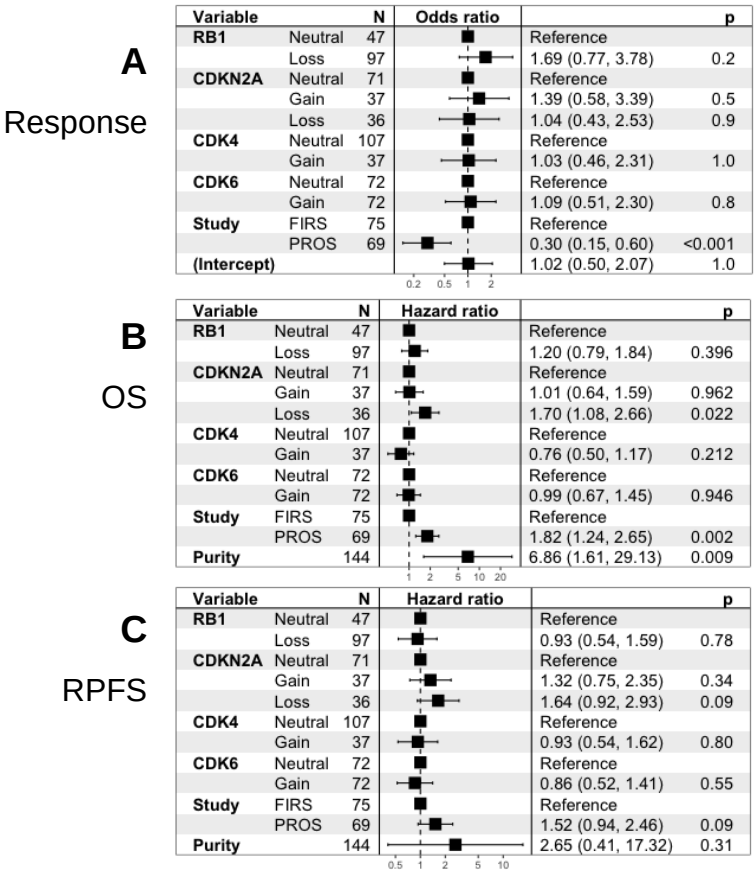


Figure 6-4 Multivariable models of *RB1* gene subset

Forest plots for (A) drug response status, (B) overall survival and (C) radiographic progression-free survival showing associations of *RB1* gene and key pathway partners with clinical outcome. Odds/hazard ratios are shown, along with confidence intervals and p-values.

6.4 Discussion

In this chapter, I initially studied the tubulin genes, encoding the proteins that taxanes bind to, before moving on to genes frequently impacted by CNAs in the mCRPC genome, with some additional important Pi3K/AKT pathway and cell-cycle genes. This list, while not exhaustive, covers a range of genomic loci and serves to test several hypotheses with regards to prostate tumour biology and taxane therapy. Critically, it focused on genes that are key CRPC CNAs and genes with mechanistic links to taxane chemotherapy.

In these analyses, I did not comprehensively test every possible gene, for several reasons:

a) Genes are not evenly distributed in the genome, and many genes would have almost-identical results.
b) Limiting the number of genes in univariable analyses reduced the chance of false positives for multiple testing. These univariable tests were applied only to identify possible candidates for multivariable analyses.

I hypothesised that alterations in tubulin homeostasis *via* CNAs could a) have oncogenic impact, and b) associate with taxane sensitivity. I observed that several tubulin genes altered in frequency between cohorts, and that around half of the studied genes had univariable associations with either drug response or survival, including 6 isoforms of beta-tubulin – which is the binding target of taxanes. Interestingly, several alpha-tubulins isoforms also exhibited clinical association with outcome, possibly due to the close interactions required between alpha and beta tubulins for complete microtubule assembly (Cirillo et al., 2017; Ganguly et al., 2012; Vemu et al., 2017).

Of particular interest are *TUBB4A* and *TUBB4B*, which univariably and multivariably displayed associations with clinical outcome including drug response (*TUBB4A*-gains correlating with worse OS and RPFS, and borderline worse response, respectively; and *TUBB4B*-gains associating with improved responses, OS and RPFS). Interestingly, *TUBB4B* – aside from being downregulated in tumour metastasis – has been reported to be under-expressed in taxane-resistant breast tumours, and has been shown to interact with both *TUBB4A* and *TUBB*, among other tubulin family members (Nami and Wang, 2018; Sobierajska et al., 2019). Similarly, *TUBA3D* downregulation has been

linked to taxane resistance, so the copy-gains in this cohort could plausibly be generating sensitivity.

I also observed significant associations involving alterations of the *TUBB8/TUBAL3* gene locus, which are reported to be the most frequently mutated tubulins in basal breast cancer (Nami and Wang, 2018). These associations appeared counterintuitive however, and require validation to confirm the results. *TUBB3* expression has also been linked to pro-survival pathways in several tumour types, and some (but not all) studies have suggested it to be a biomarker for taxane therapy: my analysis was unable to confirm this, although I did find a modest univariable association with improved drug response and *TUBB3* region copy-loss (Hwang et al., 2013; Loeser et al., 2017; Nami and Wang, 2018; Person et al., 2017). Moreover, there may be some drug-specific interactions – with one report suggesting that cabazitaxel has some binding affinity for the alpha-tubulin *TUBA4A*, in addition to the beta-tubulins that docetaxel binds to (Nami and Wang, 2018). However, in the data I present here the sample sizes limited further analyses, as the majority of individuals received cabazitaxel. Further studies of tubulins in taxane-resistant mCRPC will require expression and mutation analyses to fully characterise this complex set of interactions.

Alterations in the genes *PIK3CA*, *PTEN* and *RB1* have been highlighted as taxane-associated markers in other studies – but in these analyses I found only modest associations between CNAs in these genes and response and survival (Bumbaca and Li, 2018; Hamid et al., 2019; Liu et al., 2015).

One possible explanation for these interesting results may be that such analyses do not fully reflect the clinically relevant changes occurring at a protein level. Additionally, lpWGS data do not permit the identification of somatic mutations. Nevertheless, I found that loss of *CDKN2A*, associates with poor prognosis, as reported in other tumour types (Cao et al., 2018; Horn et al., 2018; Reis et al., 2015; Zeng et al., 2018). Interestingly, *CDKN2A* actually displayed a significant shift towards less copy-gains (27% to 9%) and more copy-losses (23% to 33%) throughout the course of treatment on the FIRSTANA study, indicating the emergence of additional losses over time.

I did also observe an association of *MAP3K7* copy-losses with improved response and survival, which mirrors similar studies that correlated high *MAP3K7* (*TAK1*) expression with taxane resistance (Bo et al., 2016; Ying et al., 2015). Additionally, several genes highlighted in these analyses (including *E2F3*, *MAP3K7*, *TUBB2A/B*, *TUBB*) are present on chromosome 6, which was highlighted several times in chapter 5. Unpicking the genes responsible for taxane resistance here will require careful screening to validate each one. Multiple genes may be contributing to the observed effect, which may explain why some genes had counter-intuitive results.

Low-pass WGS offers robust CNA data, although it is limited by the read depth and resolution used, and in addition can not be used to resolve a) complicated structural events or b) true allele-specific CNAs, including copy-neutral loss of heterozygosity (Adalsteinsson et al., 2017). The baseline CNA frequencies I detected in these studies were broadly in line with those expected from mCRPC tumours, however these increased when I excluded low-purity cases (Robinson et al., 2015). This is an important step in these analyses, particularly when analysing cell-free DNA; the absence of signal (ie. a particular CNA) due to low purity is not truly the absence of signal, and should not be included in biomarker-negative stratification.

Overall, these data show that while common CNAs found in mCRPC may offer few candidates in terms of stratification for taxanes, the tubulin family of genes may provide a solution, with significant changes in aberration frequency along with over half of the genes showing some univariable or multivariable association. Of particular note for further studies is the possibility that taxane efficacy could be modulated by interactions with alpha-tubulins, not just beta-tubulins – as part of the microtubule complex. These studies, however, are limited by the large size of CNAs impacting the genes, and the fact that sub-clonal aberrations are very challenging to resolve with cfDNA studies.

7 Single-Cell CRPC Genomics and Copy-Number Heterogeneity

7.1 Introduction

7.1.1 Specific Hypotheses

- Circulating tumour cells (CTCs) from the peripheral blood of advanced prostate cancer patients are a source of of cancer cell genomic DNA.
- Enrichment through whole-blood apheresis allows for a major improvement in single cell yield.
- Single cell genomic analyses offer a potentially superior alternative to bulk biopsies and cfDNA for identification of tumour subclonal alterations.

7.1.2 Specific Aims

- Optimise methods to access single cell CNA profiles for each patient.
- Illustrate clonal diversity using quantitative methodology.
- Validate copy-number changes with alternative methods.
- Identify sub-clonal CNA events that may be overlooked by bulk sequencing and offer clues to clonal evolution under treatment pressure.

7.1.3 Research in Context

- Tumour heterogeneity remains an intractable issue in terms of identifying clinically relevant alterations, and both solid and liquid biopsies suffer from this limitation due to sampling a mixture of tumour clones.
- Single-cell approaches may alleviate this by treating each cell as a separate sample, allowing for greatly improved resolution of genomic alterations
- Studying single cell genomics can reveal sub-clonal genomic alterations and their relationships, such as mutually exclusive events.

Studies elucidating the genomic landscape of CRPC have been restricted by difficulties acquiring multiple metastatic castration-resistant prostate cancer (mCRPC) biopsies (Attard et al., 2006). The landscape of both localized and advanced prostate cancer has been recently described but bulk tumour biopsy genomics only provides disease snapshots (Robinson et al., 2015). Concerns have been raised regarding the ability of bulk biopsy sequencing to document intra-tumour heterogeneity and clonal evolution, and I have shown earlier in this work that these are limitations to commonly utilized tumour sequencing methods.

In addition, serial tumour biopsies are necessary to evaluate changes imposed by therapeutic selective pressures over time; but the acquisition of repeat biopsies is challenging, invasive, and often not feasible. Less invasive alternatives (such as liquid biopsies) can be an impactful alternative, allowing serial evaluation, and detecting disease evolution that can influence treatment choices (Carreira et al., 2014; Goodall et al., 2017).

Two main forms of liquid biopsy studies have emerged: circulating plasma cell-free DNA (cfDNA) and circulating tumour cell (CTC) analyses. Although cfDNA-based approaches have strong clinical utility, as I have shown earlier in this work, limitations in qualitative analyses to deconvolute intrapatient heterogeneity and accurately call allele-specific copy-number aberrations have been acknowledged (Wan et al., 2017). CTCs, shed from solid tumours and found in the peripheral blood (PB) of patients with advanced disease, are prognostic, allow the early detection of disease dissemination, and changes in serial CTC counts can identify therapeutic benefit (de Bono et al., 2008; Cristofanilli et al., 2004; Premasekharan et al., 2016). Indeed, some data has shown that serial CTC count evaluation may be superior to radiological assessments in determining response to treatment and outcome (Budd et al., 2006; Danila et al., 2007).

Studies of CTCs can be used to molecularly characterise tumours, but a major challenge to this has been their detection in significant numbers to enable true multi-omic analyses. To overcome these limitations, whole-blood apheresis has been suggested to increase CTC yield (Stoecklein et al., 2016). Apheresis involves the processing of an individual's entire blood volume by centrifugation, separating blood components (e.g., red cells, platelets, and leukocytes) based on density. Apheresis has a therapeutic role in the management of hematologic

disorders and is well tolerated with few safety concerns. Previous studies have suggested that CTCs can be collected from an apheresis product from patients with and without metastases (Stoecklein et al., 2016). CTCs can have a similar density to mononuclear cells and apheresis can increase CTC separation from a larger volume of processed blood.

The genomic characterisation of single CTCs from mCRPC patients has so far been limited (Leung et al., 2017). Due to the importance of CNAs to CRPC disease, I sought to profile single CTC data derived from patients suffering from mCRPC. I analysed CTC genome-wide copy number events utilising whole-blood apheresis to increase CTC count yields and initially array comparative genomic hybridisation (aCGH) to detect copy-number events. I first confirmed that these cells are tumour derived CTCs and representative of the mCRPC genome. I went on to characterise the cohort in terms of both key CRPC aberrations (such as *AR*, *PTEN* and *RB1* gene copy-number changes), and the overall copy number burden of individual cells. I calculated measures of per-patient tumour diversity using these. I attempted to identify intra-patient tumour heterogeneity of key genomic events and compared them with orthogonal methods including fluorescent in-situ hybridisation (FISH). To expand on these data, I analysed a separate cohort of single mCRPC CTCs from one patient, utilising low-pass whole genome sequencing, with the goal of more deeply characterising the clonal heterogeneity and development of these cancers.

7.2 Methods

7.2.1 Patient Selection for Apheresis

Eligible patients had histologically confirmed mCRPC. Additional eligibility criteria included detectable PB CTCs (CellSearch), good bilateral antecubital fossa venous access, and no coagulopathy. Clinical assessments included medical history and physical examination, full blood count, biochemical tests, and coagulation. Safety assessments were done during apheresis and after 30 days. All patients provided written informed consent. The study was conducted in accordance with the Declaration of Helsinki, with the ethics committee of the Royal Marsden and The Institute of Cancer Research approving the study.

7.2.2 Circulating Tumour Cell Collection by Apheresis

Apheresis was performed using a Spectra Optia Apheresis System (Terumo, BCT). Patients were connected to this via two peripheral venous catheters in each cubital vein. Whole blood was anticoagulated before entering the rotating centrifuge. Heavier blood elements including erythrocytes migrated to the outside of the channel, plasma to the centre, and the buffy coat (including mononuclear cells and CTCs) to the middle. The mononuclear cell layer was removed and the remaining blood cells and plasma were constantly returned to the patient to the contralateral arm. Granulocyte-colony-stimulating factor was not used. Blood was anticoagulated with citrate dextrose solution A (two to four 500-mL infusion bags were required for each procedure).

7.2.3 CTC enumeration using CellSearch platform

CTC counts were determined in 7.5 mL of peripheral blood (PB) drawn immediately before, and after, the apheresis; an aliquot of apheresis product containing 200 × 106 white blood cell (WBC) was transferred to a CellSave preservative tube (Menarini, Silicon Biosystems) and mixed with CellSearch dilution buffer to a final volume of 8 mL. All samples were processed within 96 hours, and CTC counts were determined by CellSearch (Menarini, Silicon Biosystems). Briefly, cells were subjected to immunomagnetic capture using anti-EpCAM antibodies and stained with antibodies specific for cytokeratin 8, 18, and 19 (CK-PE), CD45 (CD45-APC), and nucleic acid dye (DAPI). Cells were defined as CTCs when positive for cytokeratin and DAPI and negative for CD45. Images

were captured using the CellTracks Analyzer II (Menarini, Silicon Biosystems) and manually examined to determine the presence of CTC. CellSearch cartridges were stored in the dark at 4°C before further analyses.

7.2.4 Single-Cell Isolation and Amplification

CellSearch cartridge contents were transferred into fresh Eppendorf tubes, washed twice with 150 µL of phosphate-buffered saline, and FACS sorted (FACSAria III; Becton, Dickinson and Company) to single CTCs (DAPI+, CK+, CD45−) or WBC (DAPI+, CD45+, CK−). Sorted single CTC or WBC were whole-genome amplified (WGA) using Ampli1 (Menarini, Silicon Biosystems) according to the manufacturer's instructions with minor modifications. Cells were lysed, digested for 30-minutes, adaptor ligated for 3-hours, and PCR amplified. The WGA DNA was purified (MinElute PCR Purification Kit; Qiagen), quantified using Qubit (Invitrogen), and stored at −20°C.

7.2.5 Biopsy DNA Acquisition

DNA from formalin-fixed paraffin-embedded (FFPE) biopsies was extracted using the QIAamp DNA FFPE Tissue kit (Qiagen), quantified using Qubit (Invitrogen), and evaluated by Illumina FFPE QC kit. Whole-genome amplification was carried out on 10 ng of tumour DNA using WGA2 (Sigma-Aldrich). WGA DNA was purified (MinElute PCR Purification Kit; Qiagen), quantified (Qubit; Invitrogen), and stored at −20°C.

7.2.6 Array Comparative Genomic Hybridization (aCGH)

Five hundred nanograms of amplified single CTC DNA was fluorescently labeled with Cy5, and WBC reference DNA labelled with Cy3 (SureTag Complete DNA Labeling Kit; Agilent Technologies). Labeled DNA was purified and hybridized utilizing the Agilent SurePrint G3 Human array CGH Microarray Kit, 4 × 180K. Slides were scanned and ratios of CTC/WBC determined using CytoGenomics Software v 4.0.3.12 (Agilent Technologies).

7.2.7 Copy-Number Data Processing and Statistics

Single cell copy number profiles were segmented using the CytoGenomics Software v 4.0.3.12 (Agilent Technologies). Log_2 ratios of aCGH segments were matched with gene coordinates to assign per-gene values. Copy states of genes

were classified by the assigned $\log_2$ ratio values. Log_2 ratio values < −0.25 were categorized as losses; those > 0.25 as gains; and those in between as unchanged. Amplifications were defined as smoothed $\log_2$ ratio values ≥1.2 and homozygous deletions as the segment $\log_2$ ratio values ≤ −1.2.

Per-sample CNA burden was calculated as the proportion of the human genome (3,000 megabase pairs) affected. Unsupervised hierarchical clustering was performed using R (v3.4) with Ward method and the Euclidean distances of unique copy-number changes. When clustering samples from multiple tissue types, X-chromosome genes were excluded (aside from the *AR* gene and 10 genes on either side) due to different reference X-chromosome ploidies (as a female reference was used). Per-patient functional diversity was derived from cluster dendrograms of CTC samples by calculating the sum of connecting branches in a dendrogram (from the R package vegan v2.4.4) and divided by the number of samples.

7.2.8 Whole-Exome Sequencing

Whole-exome sequencing (WES) was performed using Kapa HyperPlus library prep kits and the Agilent SureSelectXT V6 target enrichment system. Paired-end sequencing was performed using the NextSeq 500 (2 × 150 cycles; Illumina). FASTQ files were generated from the sequencer output using Illumina bcl2fastq2 software (v.2.17.1.14, Illumina) with the default chastity filter to select sequence reads for subsequent analysis. Sequencing reads were aligned to the human genome reference sequence (GRCh37) using the BWA (v. 0.7.12) MEM algorithm, with indels being subsequently realigned using the Stampy (v.1.0.28) package. Picard tools (v.2.1.0) were used to mark PCR duplicates and to calculate sequencing metrics for QC check.

The Genome Analysis Toolkit (GATK, v.3.5.0) was then applied to realign local indels, recalibrate base scores, and identify point mutations and small insertions and deletions. Somatic point mutations and indels were called using MuTect2 by comparing tumor DNA to matched germline control. To obtain copy-number estimates, germline mutation data was converted to a microarray-like data structure (a matrix of allele frequencies with a matching matrix of coverage ratios for each SNP) for input to the ASCAT (v2.5.2) package.

7.2.9 Fluorescence in Situ Hybridization (FISH) Analysis

FISH was performed by FFPE hybridization as previously described (Armstrong et al., 2012). Briefly 3- to 4-µm FFPE sections were deparaffinized, heat pretreated, pepsin digested and hybridized with FISH probe hybridization mix overnight at 37°C. FISH probes used were *BRCA2/CEN13q* (Abnova), *RB1* (Abbott Laboratories), *PTEN* (10q23)/SE10, *MYC* (8q24)/SE8 (Leica Microsystems), and a custom-made *AR/CEPX* probe (Menarini, Silicon Biosystems). Stringency washes were performed on all slides; for *AR*, where the probe was indirectly labelled, a secondary incubation with anti-digoxigenin–fluorescein antibody (Roche Diagnostics) was done. Slides were digitally imaged (BioView Ltd.), and a pathologist (D.N.R.) evaluated a minimum of 100 tumour cells; the ratios between probes of interest and reference probes were recorded. Amplification was reported if the ratio was >2; heterozygous loss and homozygous deletion if at least one of three of the cells showed loss of one copy, or loss of all copies, of the tested probe, respectively.

7.2.10 Pilot Study lpWGS of CTCs

For the pilot study of lpWGS, cells were not isolated with apheresis. Instead, CTC enrichment and enumeration were performed using the CellSearch® Circulating Tumor Cell kit (Veridex) in the CellSearch® system (Veridex) according to the manufacturer's protocol. Isolation of single CTCs, defined as DAPIpos/CKpos/CD45neg cells, and single white blood cells (WBCs), defined as DAPIpos/CKneg/CD45pos, was performed by flow cytometry as previously described (Neves et al., 2014). Whole genome amplification (WGA) was performed using adapter-linker PCR as previously described with the Ampli1^{TM} WGA Kit by Silicon Biosystems (Bologna, Italy) (Klein et al., 1999). Sequencing was performed by Silicon Biosystems using a procedure similar to in chapter 4 on the Illumina HiSeq. NGS reads were aligned using BWA (v0.7). Aligned read data was analysed using the Control-FREEC (v6.0) software to generate segmented CNA data with $\log_2$ ratios.

I merged the segmented CNA data from Control-FREEC into a same-patient dataset with CNA values covering identical 600-kb bins using the GenomicRanges (v.1.36) R library. The data structure was essentially a matrix, with rows corresponding to genomic regions and columns single cell samples.

For CNA segment analyses I performed segmentation using multi-sample segmentation from copynumber (v1.24.0). Missing data points were imputed using the same package. This approach means that CNA boundaries are unified between samples. For phylogenetic analysis I used the raw binned data.

Before downstream analysis, I considered cells which: a) present with extremely noisy CNA data or b) are extremely divergent from all other same-patient cells, uninformative for phylogenetic analyses, so I applied two broad filters to remove them. To identify noisy samples I calculated the median per-sample interquartile range (IQR), (ie. the median spread of $\log_2$ ratios of bins away from the assigned segment value), and excluded samples with a median IQR higher than the 95% of that patient data set. I also calculated a mean per-cell pearson correlation (ie. the mean correlation value of that CTC dataset when compared to every other from that patient), and excluded those below 0.4.

Phylogenetic trees were initially produced using the neighbour-joining algorithm with the NJ() function from the phangorn library (v2.5.5). Maximum parsimony (MP) trees were produced using the initial tree as input, and the optim.parsimony() function from phangorn. MP tree branch lengths were estimated using the acctran() function, also from phangorn. Trees were exported in newick format, and used as input for TreeCluster (v1.0), with the ‘Max’ method and a threshold value of 7000.

7.3 Results

7.3.1 Patient Characteristics of CTC-CNA Cohort

To establish methods for single cell CNA profiling, I first accessed data collected by an enriched sampling method (patients consented and apheresis established by S.S., M.F. and M.L.) using whole-blood apheresis, which enables acquisition of large numbers of viable CTCs for downstream analyses. From November 2015 to July 2017, 14 eligible mCRPC patients with detectable CTCs by CellSearch were enrolled (median age, 70.4 years; range, 60–77); time from prostate cancer diagnosis to procedure ranged from 2–11.6 years (mean, 6.2 years; median, 3.9 years). The median PSA value at sampling was 506 ng/mL (range, 41–6089 ng/mL); all 14 (100%) had metastatic bone disease. Prior to apheresis, patients had received 1 to 5 lines of systemic therapy for CRPC (Supplemental Figure 10-3). At sampling, none of the subjects was receiving active treatment other than androgen deprivation.

Each apheresis procedure lasted between 90 and 160 minutes and the total apheresis product volume ranged from 40 to 100 mL (Supplemental Figure 10-6). The apheresis procedure was well tolerated with no related adverse events recorded during the procedure or in the 30-day follow-up. The observed CTC count did not decrease significantly following apheresis (P = 0.48), however the procedure did substantially increase CTC yields (Supplemental Figure 10-6). Following collection of apheresis product, single cells were stained and identified using the CellSearch platform, and isolated using fluorescence-activated cell sorting (FACS) for prostate epithelial cell surface markers (single cell isolation performed by M.L.).

7.3.2 Single-cell CNAs Represent the CRPC Genomic Landscape

Isolated single CTCs (n=205, from 14 patients) were processed for aCGH analysis (by M.L.) and I assessed the overall genome-wide frequencies of copy-number changes. The aggregate aCGH copy-number profiles of these samples closely matched a previously reported dataset of advanced prostate cancer biopsies analysed by whole-exome sequencing (Figure 7-1, below).

I then analyzed the single CTC aCGH genomic profiles for CNAs with 185 CTC (90%) showing complex genomic copy change profiles and 20 (10%) cells having relatively flat genomic copy-number profiles. Only two of the evaluated 14 patients had cell populations with both flat and cancer-like aCGH profiles, suggesting that these cells are not found across all tumours but are instead a sporadic phenomenon.

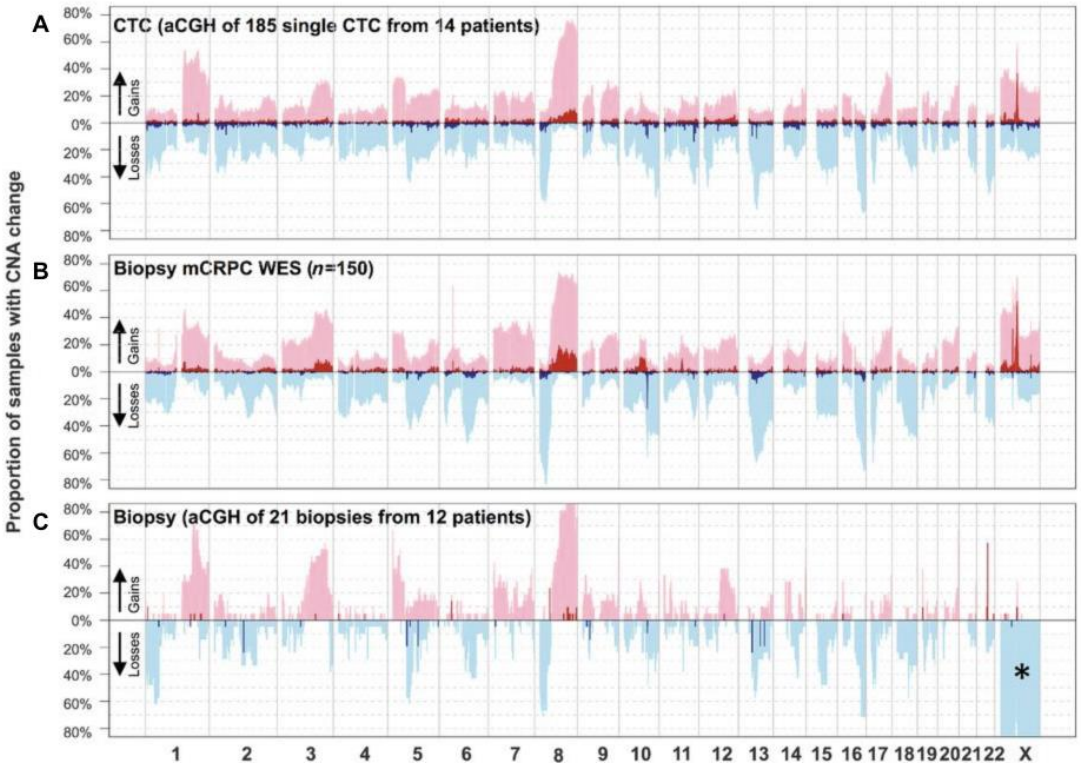


Figure 7-1 Copy number frequencies of CTCs match bulk biopsy tissue

Whole-genome copy number aberration (CNA) frequency plots showing results from 185 single cell aCGH profiles from 14 patients (A), a cohort of 150 WES bulk biopsy samples (B), and a set of 21 biopsies taken from the single cell-analysed cohort (C). aCGH of biopsies (C) was performed using a female reference sample, so deletions across the X chromosome are observed. Chromosomes are shown across the x-axis, whereas the y-axis represents the frequency of gains, losses, amplification, and homozygous deletions. Gains are depicted in light pink, losses are depicted in light blue, amplification in dark red, and homozygous/deep deletions are in dark blue.

To validate the aCGH copy-number calling, comparisons were made between reference samples (male *vs* female) as well as apheresis derived single white blood cells (WBCs). No detectable CNAs were observed, aside from the expected sex chromosome imbalance as female control was used (Figure 7-2). Additionally, single CTC DNA from a patient with known tumour CNAs was

evaluated at variable concentrations of whole-genome amplified (WGA) input material: at 10ng/μL, 1ng/μL, 0.1ng/μL and 0.03 ng/μL, with consistent calling of copy-gains and losses clearly visible and minimal amplification biases.

Of the 14 patients with single CTCs available, 12 had biopsies (a mix of diagnostic and metastatic biopsies) available for analysis, and these also showed similar frequencies of key CRPC events (Figure 7-1) to the overall single cell cohort and the orthogonal WES cohort. Unsupervised hierarchical clustering of all sample CNA profiles showed that patient biopsies invariably clustered alongside same-patient CTCs (Figure 7-3), despite some visible heterogeneity in overall cluster size and position. Differences were frequently observed between treatment-naïve biopsies and castration-resistant CTC samples however, including *AR* gain (chrX), *MYC* gain (chr8q), and *RB1* loss (chr13), which may depict tumour evolution under selective pressures (Supplemental Figure 10-7). High concordance between single CTC genomic profiles and contemporaneous, same patient, metastatic biopsies was seen, although intra-patient genomic variety was discernible from the single CTC analyses.

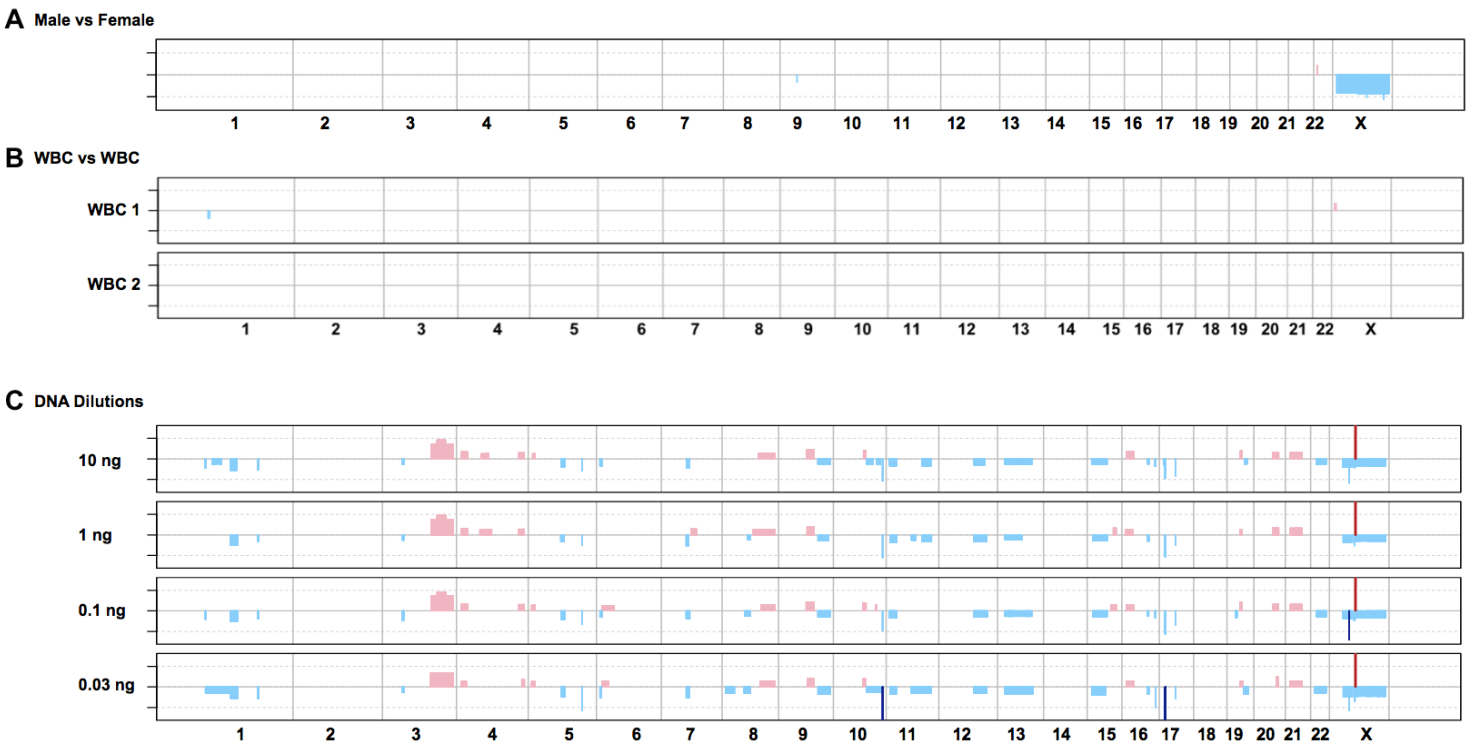


Figure 7-2 Single-cell aCGH validation

Whole-genome copy number profiles of validation samples show now observable amplification biases. Both male/female (A) and WBC/reference (B) analyses show flat (ie. non copy-altered) profiles. Assessing copy-number changes in variable input DNA concentrations (C) shows that the observed copy segments are consistent throughout. Copy gains shown in red, deletions in blue.

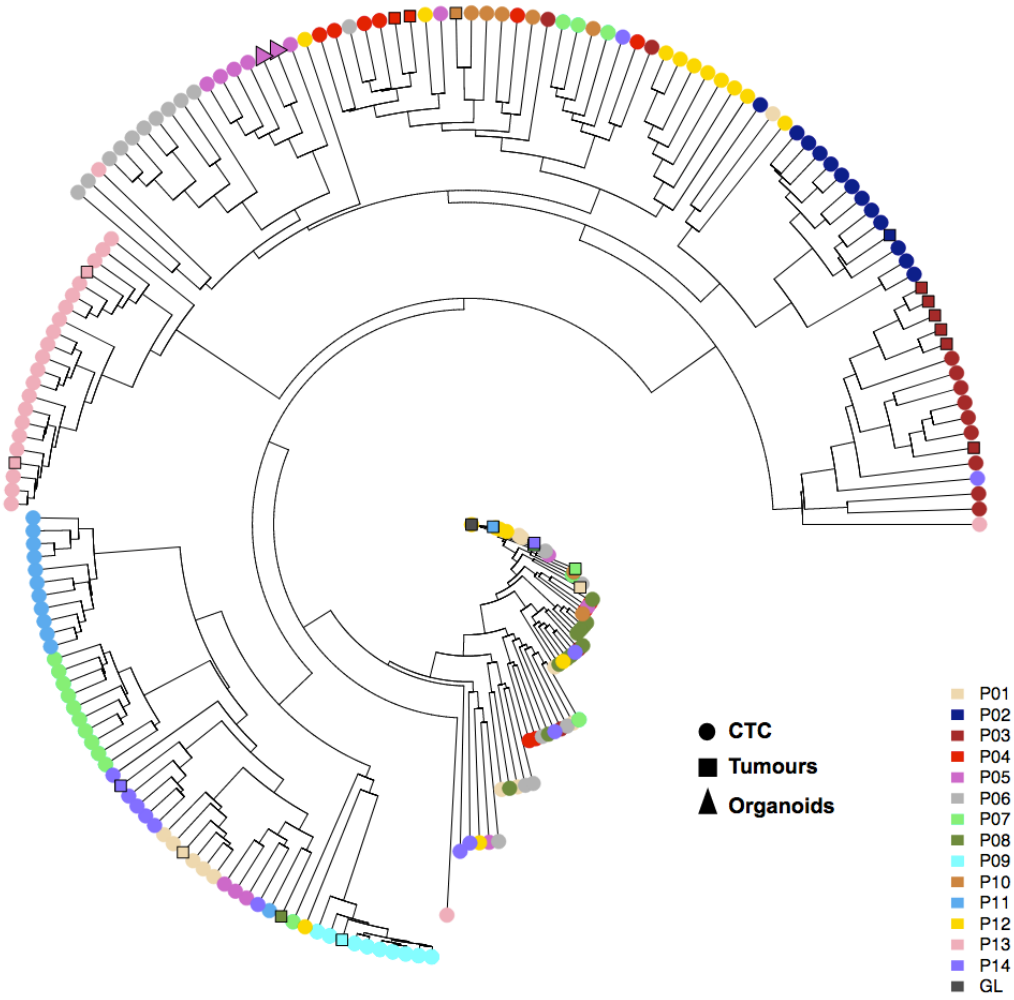


Figure 7-3 Unsupervised hierarchical clustering of CTCs with matched biopsies

Same-patient samples (CTCs, biopsies and CTC-organoids) invariably cluster with other samples from the same individual. Nodes are samples, coloured by patient, and node shape corresponds to sample type.

7.3.3 Single CTC Copy-Number Changes Reveal Diverse Genomes

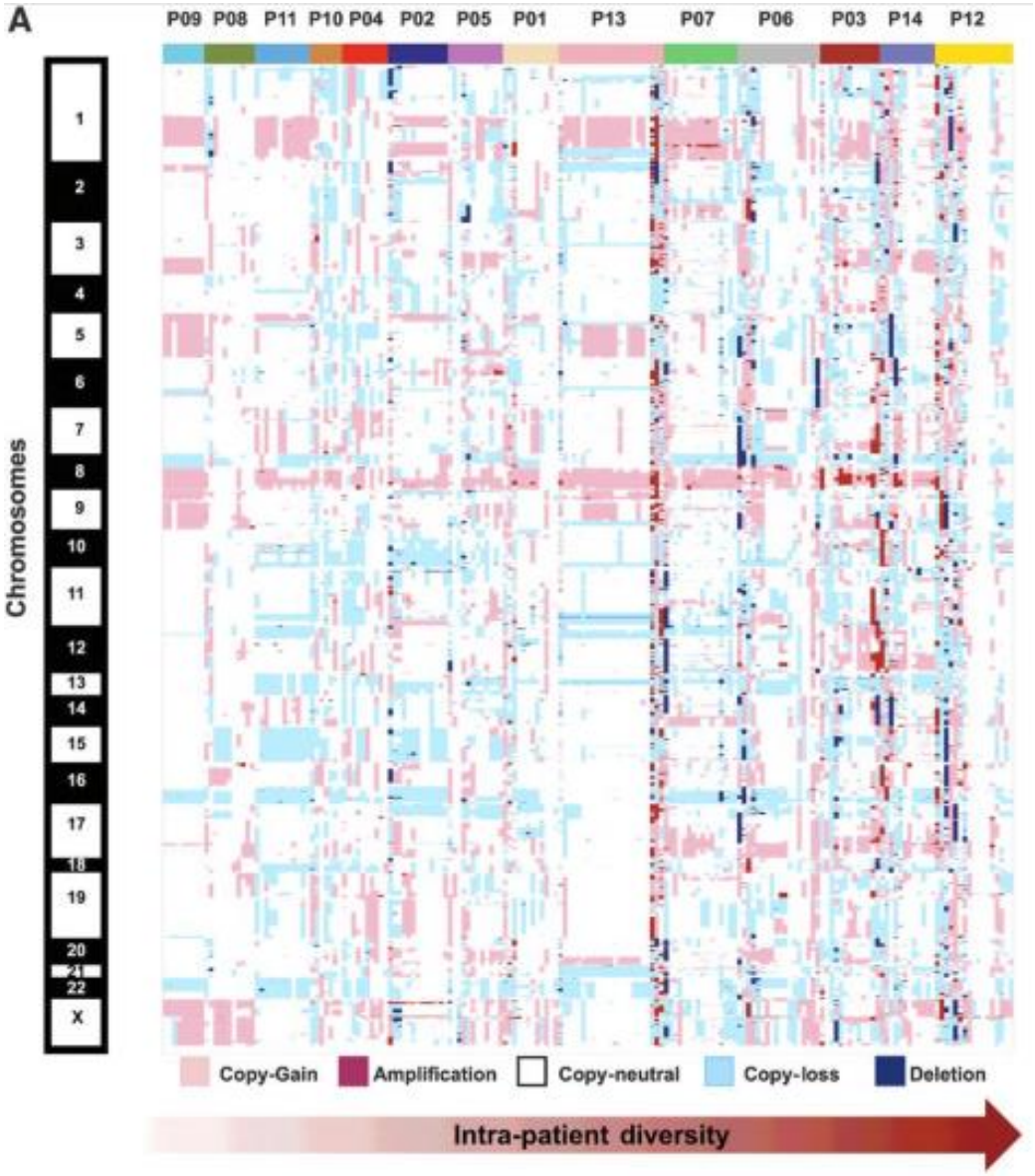


Figure 7-4 Heatmap of single-cell whole-genome CNA profiles

Individual single CTC copy-number data detecting complex intra- and inter-patient genomic heterogeneity. Columns are ordered first by patient and then by unsupervised hierarchical clustering. Patients are organised by the overall intrapatient diversity score, from left to right. Chromosomal coordinates are shown from top to bottom. Copy-number gains are shown in red, with copy-losses in blue.

I scored each patient with a diversity value, derived from the total Euclidean distance of the evaluable cells normalised by total per-patient cell count. Overall, the genomic analyses of 185 single CTCs from 14 patients revealed that some patients had highly homogenous CTC CNA traces, while others had highly

diverse single CTC CNA traces (Figure 7-4). This may be related to disease phenotypes or acquired treatment resistance mechanisms (*AR* and *MYC* gain at chromosomes X and 8q, respectively; *BRCA2/RB1* locus loss at chromosome 13).

I found now no significant correlation between median percentage genome alteration and intrapatient diversity, suggesting that this was due to true clonal diversity rather than accumulation of aberrations over time. Despite this, the unsupervised hierarchical clustering of all the CTC CNA data and same patient biopsies indicated that most samples from one patient clustered together (Figure 7-3).

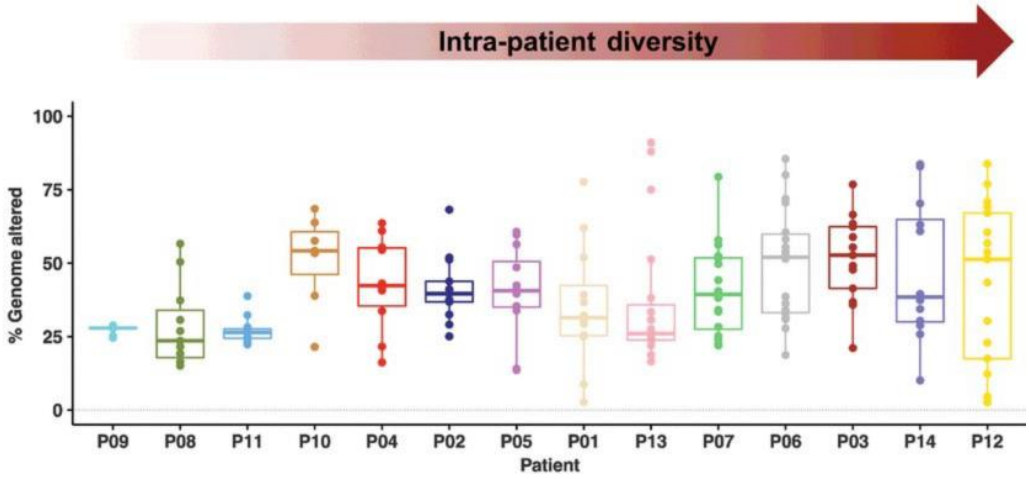


Figure 7-5 Single CNA copy-number burden

Box-plot illustrating the proportion of each genome impacted by a copy-number event in the single CTCs. Patients ordered by the intra-patient heterogeneity value.

Some patients (such as P09) had highly homogenous CTC copy-number profiles, with the contemporaneous mCRPC biopsy bearing a virtually identical set of CNAs to the CTCs (Figure 7-6, below), including characteristic events on chromosome 8 (8p deletion and 8q gain).

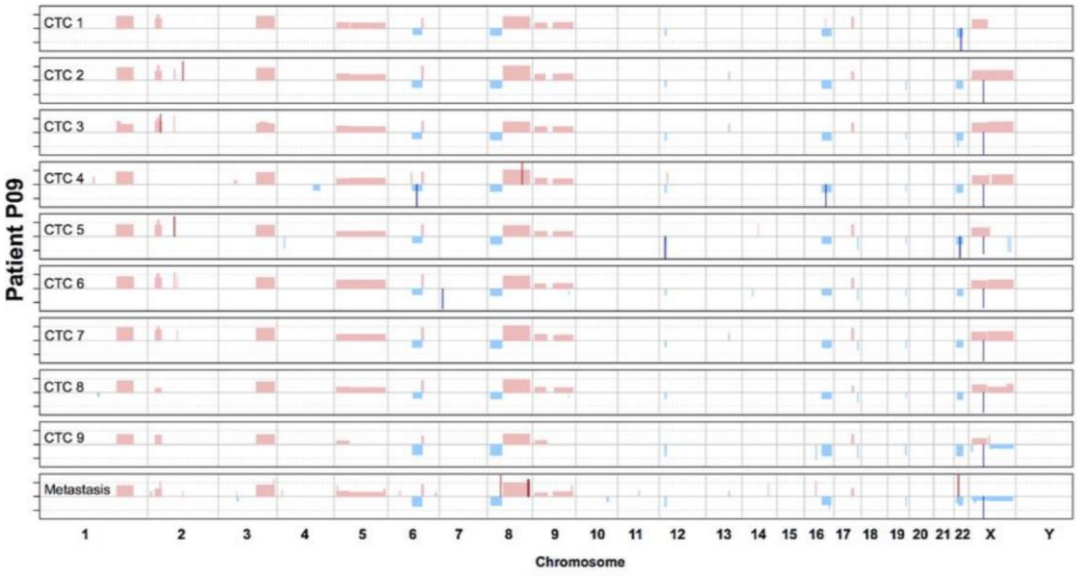


Figure 7-6 Whole-genome copy-number traces for P09

Individual copy-traces for CTCs and a matched metastatic solid biopsy from patient P09. Copy gains/amplifications in pink/red, and copy loss/deletions in light blue/navy.

7.3.4 Clinically Relevant Sub-Clonal Alterations

The majority of evaluated patients bore a heterogeneous set of CTC CNA profiles, with infrequent events that gross biopsy genomic analyses could fail to identify. To further interrogate this intra-patient heterogeneity, I studied additional cells from patient P13, who had heterogeneous CTCs, with CNA data suggesting distinct groups of cells (Figure 7-7). All CTCs were collected at the same time, but some CTCs clustered with his diagnostic prostatectomy sample, while others more closely resembled the mCRPC bone biopsy, with a large copy-number including most of chromosome 5q affecting a proportion of the cells (Figure 7-7). This segment had breakpoint in the *PIK3R1* locus. A third group of cells was also apparent, displaying more complex genomic aberrations.

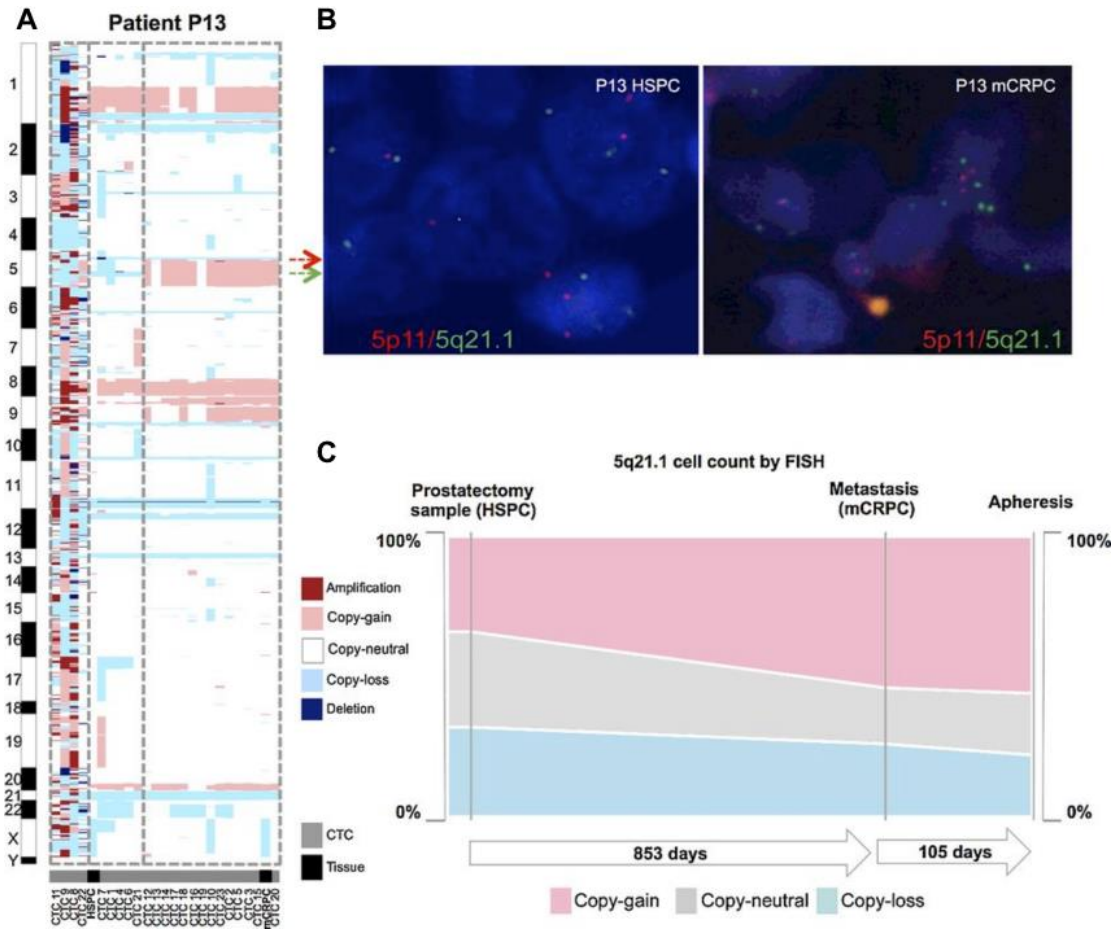


Figure 7-7 Heterogenous cells and subclones in patient P13

Copy-number heatmap illustrating three main groups for Patient P13: prostatectomy-like, metastasis-like, and a set of hyper-altered cells (A). FISH analyses of both biopsies revealed intracellular heterogeneity at the 5q21.1 locus (B). Tracking the changes over time indicates a shift to the copy-gained sub-population (C).

FISH analyses of the 5p11/5q21.1 loci performed (by S.M. and D.N.R.) on both the hormone-sensitive prostatectomy sample and the castration-resistant metastasis revealed the presence of distinct copy-number aberrant cells, with 5q21.1 being either gained, normal, or lost in a mixed cell population (Figure 7-7). This heterogeneity was reflective of the variety observed in the aCGH data. The 5q21.1 region includes the chromatin-remodelling gene *CHD1*. These three copy-states were equally common in the prostatectomy sample, and existed in a broad admixture. Over time and following treatment, the proportion of tumour cells with 5q copy gain increased as shown in the mCRPC biopsy and apheresis CTCs and as confirmed by tissue FISH analysis (Figure 7-7).

I went on to study patient P03 in detail, as multiple biopsies were available: a castration-resistant lymph node metastasis sample, and a transurethral resection

of the prostate (TURP). The TURP was micro-dissected (by D.N.R. and M.L.), and four morphologically distinct regions (A, B, C, and D) were processed for copy-number changes (Figure 7-8).

All four micro-dissected regions featured a homozygous deletion of *BRCA2*, along with copy-gain of the q-arm of chromosome 8. Distinct sub-clonal events were also clearly visible between the regions, including broad losses on chromosome 18 (B and D) and gains on chromosome 7 and 12 (A and C) (Figure 7-8). The CNA profile of a lymph node (LN) biopsy acquired from this patient 6 years later, following treatment with docetaxel P (75 mg/mm^2), enzalutamide (160 mg), and cabazitaxel (25 mg/mm^2), enzalutamide, and cabazitaxel, showed the *BRCA2* homozygous deletion and 8q gain, as well as previously unobserved *AR* amplification and 17q gain.

Whole-exome sequencing (WES) data was also available for this case (processed by W.Y.), with the four TURP regions bearing a distinct pattern of mutations including clonal pathogenic *SPOP* (p.Trp131Cys) and *FOXA1* (p.His168del) variants. Analysis of allele frequencies again grouped the four samples into two pairs: A and C, B and D (Figure 7-9). I clustered the available CNA data from both single cells and tumour biopsies, and these data illustrate a branched evolutionary background for this tumour with heterogenous *PTEN* and *BRCA2* deletions across sub-clones. FISH analysis of the TURP biopsy tissue also showed inter-cell variability in *MYC* amplifcation and *BRCA2* deletion copy-states with some cells bearing concurrent events while others displayed normal *BRCA2*. This illustrates the heterogeneity present at the single-cell level, and likely reflected a subclonal acquisition of the *BRCA2* loss.

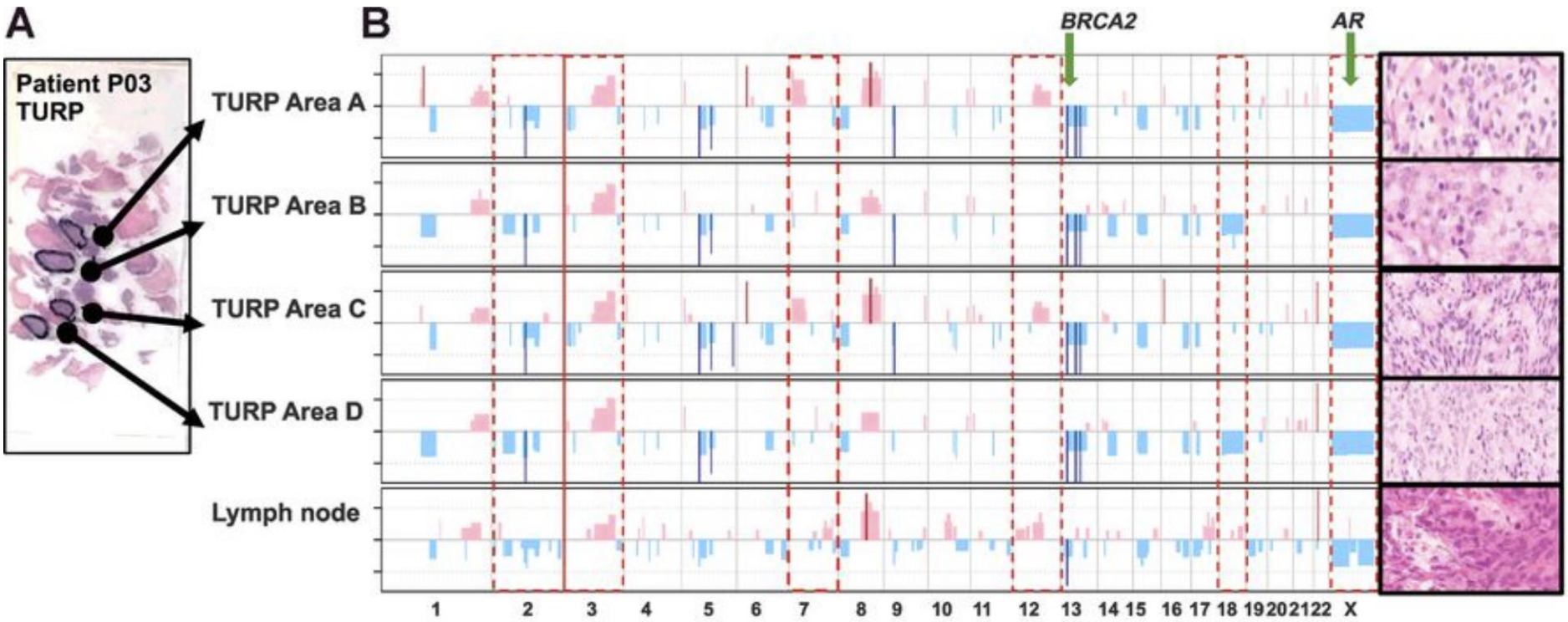


Figure 7-8 Micro-dissection and aCGH of Patient P03

Collating four micro-dissected samples from an archival TURP sample (A), along with a metastatic lymph node biopsy shows both distinct tumour morphology and distinct aCGH profiles (B). Dashed red boxes indicate areas of inter-sample heterogeneity. A clonal deletion of BRCA2 is featured in all samples.

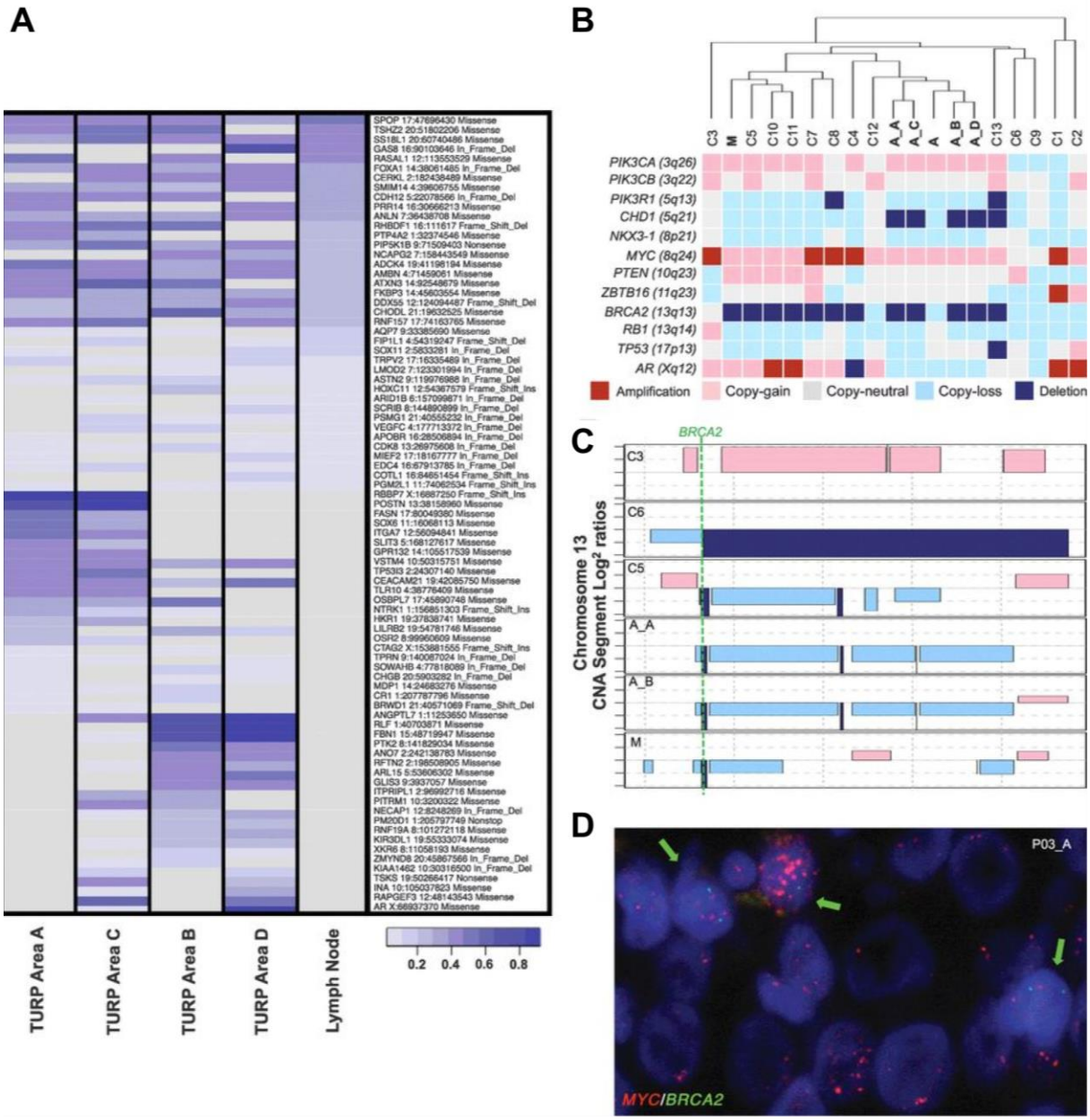


Figure 7-9 Inter-Sample Heterogeneity of Patient P03

Whole exome sequencing (WES) analysis of TURP regions A-C and a lymph node metastasis of shows both clonal and sub-clonal mutations (A). This heterogeneity is reflected in unsupervised clustering of key CRPC genes (B). Individual CTCs are numbered “C#”, “A” indicates bulk TURP biopsy material, the four TURP microdissected regions are listed “A-A”, “A-B”,”A-C”,“A-D”, and the lymph node metastasis represented by “M”. Plot of chromosome 13 showing *BRCA2* gene heterogeneity between both CTCs and tissue samples (C). FISH analysis of *MYC* (red) and *BRCA2* (green) copy-states indicates a admixture of combinations (D).

7.3.5 Pilot Analyses of CTC-lpWGS

To explore the potential of detecting tumour evolution from single CTC analysis, I accessed a dataset from an additional single patient (treated with PARP inhibition, but was not classified as a responder) for which a larger number of

cells were available. These data were generated in collaboration with Dusseldorf University (Professor Nikolas Stoeklein). The cells were not isolated by apheresis, but DNA extraction and WGA was as described earlier in this chapter. Following single-cell isolation, whole genome amplification and sequencing, 55 cells were available at three timepoints: baseline (n=9), on-treatment (n=17) and end-of-treatment (n=29). I converted individual cell lpWGS CNA data (produced by ControlFREEC software) into a merged single-patient dataset with CNA data (ie. log2 ratios) assigned to 600kb bins (see methods 7.2.10). I also removed samples based on filters for CNA data noise and poor overall correlation with same-patient samples, as these are liable to add uninformative data to analysis. Following this quality control step, 46 cells were remaining (baseline=9, on treatment=12, end of treatment=25).

To examine the clonal disease structure of this patient's mCRPC; I established a novel application of existing phylogenetic tools. Instead of treating the CNA data as numeric, I elected to use a character evolution method, which I judge to be better suited to this data source – for example, the magnitude of a homozygous deletion is not simply 'one more' than a hemizygous.

I used the binned CNA data to construct a phylogenetic tree using the optimum parsimony algorithm to find the simplest structure that explains the data, using all the available cells (n=46) across all three timepoints. To convert CNA data to the required format, I treated the analysis similarly to nucleotide sequence phylogenetics, with copy-number states (ie. homdel, hetloss, neutral, gain, amplification) used instead of nucleotides. To account for the differences in CNA magnitudes, I also used a custom 'cost matrix' to apply an arbitrary cost to changes from each state. I judged that some aberrations would be harder or easier to acquire given the background of the cells: ie. a tumour with a homozygous deletion of a region is unlikely to have daughter cells with copy-gain in that same region. I also 'rooted' the generated phylogenetic tree by adding an artificial germline sample – a completely copy-neutral case with no CNAs. The phylogeny suggested a branching structure with three main groups of cells (Figure 7-10 below).

To assign subclones in an unbiased manner, I used the software TreeCluster, which assesses tree structure and branch lengths to generate cluster assignments (Balaban et al., 2019). One cluster (in green) features

predominantly clonal alterations, and likely represented the ‘trunk’ of the tumour evolutionary tree for this patient. The remaining two clones (blue, red) were characterised by both many more (frequently private) additional changes, and could be considered separate subclones. All the clusters feature some degree of deletion on chromosomes 6, 8 and 13. The red cluster depicts numerous large copy-gained regions, (including whole-chromosome gains of chromosomes 4, 5, 10, 14, 16 and 18) and perhaps indicating a shift in overall ploidy. The blue cluster depicts the acquisition of numerous deletions with large areas on chromosomes 1 and 2 lost in almost all cases.

Examining the changes in cluster proportions over time, I found that there was a significant shift in cluster proportions from baseline to end of treatment collections (fisher’s exact test, p=0.001), with one subclone (in blue, Figure 7-10) shifting radically to dominate the others, increasing from 16% at baseline to 76% at end of treatment, In keeping with similar results acquired from the CTCs generated by apheresis. I am now expanding this work to a larger number of mCRPC patients.

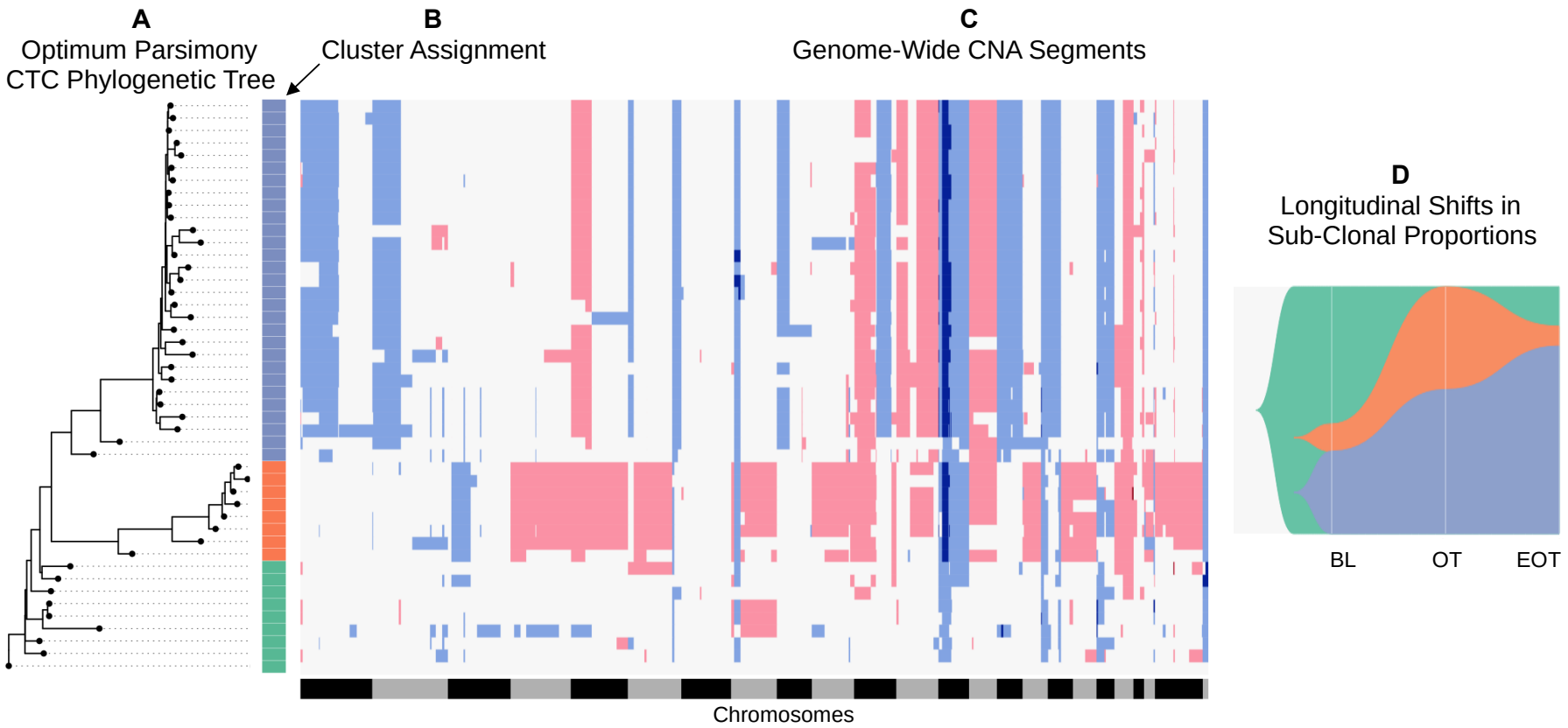


Figure 7-10 Phylogenetic analysis of pilot lpWGS CTC CNAs

Inferred tumour evolution of a single-patient set of single cell CTC CNA data, with a maximum-parsimony phylogenetic tree (A), the associated subclone clusters (B), and raw copy-number data (C). Additionally, as cells were collected at three different timepoints, a plot illustrating the relative proportions of each cluster observed at each timepoint (D). Each subclone is coloured.

7.4 Discussion

I show in this chapter that liquid biopsy by apheresis is a minimally invasive and well-tolerated procedure that increases CTC yield >100-fold from mCRPC patients. Interestingly, these data indicate that apheresis does not significantly impact CTC counts, which suggests either a) a rapid replenishment of cells, or b) inefficient capture. I show here that the genomic landscape of prostate cancer CTCs captured by apheresis mirrors that of mCRPC biopsy exomes, validating this CTC capture methodology and aCGH CNA profiling. Copy-number traces of individual CTCs frequently closely resembled same patient biopsies, with evidence for CNAs changing over time due to therapeutic pressures (including gains in *MYC* and *AR*). Critically, subclonal CNAs are not easily discernible from bulk biopsy analyses but were detected by single CTC.

Yields of evaluable single cells decrease significantly through these experimental procedures; stringent settings in FACS sorting to allow isolation of only pure single cells result in a 60% to 80% retention rate of CTCs from CellSearch cartridges. DNA from approximately another 20% of these cells fails quality control after whole-genome amplification. Therefore, in order to end up with sufficient CTCs for genomic analyses, a high number of cells are required, making the concentrated apheresis product a more efficient source than peripheral blood draws. One limitation of my analyses is that the optimal number of cells required for studying sub-clonality in tumours has not yet been established, with various studies ranging from ten to over a hundred published (Leung et al., 2017; Stoecklein et al., 2016). At least, multiple cells representing each subclone are likely required, and as clonal diversity increases so does the required number of cells to accurately assess this. It may also be the case that the ideal number of cells differs a) between patients and b) between tumour types. Moreover, single-cell RNA sequencing protocols are cheaper and typically use less material than DNA sequencing, and can therefore access more cells, and comparisons between these methodologies will be key in the future.

Interestingly, I found varying degrees of intrapatient heterogeneity in different patients, with some having highly homogenous single CTCs, but many having detectable CTC genomic diversity. Some CTCs resembled diagnostic biopsies, while others more closely mimicked metastatic tissue. Such heterogeneity remains difficult to identify from circulating free DNA, with existing methods

relying on high depth whole-genome sequencing providing large numbers of single nucleotide variants. These are limited by tumour purity and overlapping allele frequency distributions. Analysis of single cells, while challenging, offers a higher resolution for detecting true subclonality and heterogeneity throughout the patient. The relatively limited number of samples available also limits the work in this chapter, along with the extensive variability in the patient backgrounds. Drawing conclusions about overall clonal dynamics in prostate cancer is therefore challenging. I show, however, that by increasing the numbers of cells analysed it is possible to perform a robust analysis of subclonality in mCRPC, although such analyses remain challenging.

Moving forward, studies are needed to identify the optimal number of individual CTCs from one patient to sufficiently interrogate heterogeneity while minimising cost. Low-coverage whole-genome next-generation sequencing with barcoding of DNA from each CTC may allow this, as well as exploration of single-cell RNA sequencing to better understand resistance mechanisms. Direct comparison of CTCs acquired by apheresis with both CTCs and cfDNA from PB, as well as with single cells dissociated from tissue should be pursued. Finally, studies to evaluate the large numbers of immune cells in the apheresis product from these patients are also merited.

There is a clear benefit to studying CTCs from serially taken blood samples. Early identification of resistant clones can be utilized to reverse treatment failure, guiding drug combination administration or the serial utilization of drugs not tolerated when administered together. Ideally, analysing the dynamics of tumour clones through liquid biopsies will offer insights into subclonal eradication/evolution during therapy to further evaluate this strategy, while also generating estimates of CTC counts for monitoring response to therapy. Identifying subclonal events using CTCs also offers a way to monitor the different treatment sensitivity and resistance that multiple metastases display in mCRPC.

In conclusion, I demonstrate here that the analyses of whole-genome CNA profiles from single CTCs permits the identification of intrapatient tumour genomic heterogeneity previously missed by bulk biopsy analyses. Additionally, I show that it is possible to illustrate overall tumour diversity by analysis of these cells, although validating this with orthogonal approaches will be key in the future.

Overall, I show that analyses of large numbers of CTCs allows for more detailed phylogenetic and evolutionary analyses to be performed. Single cells offer a viable liquid biopsy that can display the clinically actionable alterations that I have discussed in previous chapters.

8 Conclusions and Future Perspectives

8.1 Overall Context

Advanced prostate tumours are a complex group of diseases for which it has been challenging to define clinical subtypes. Molecular stratification of advanced prostate cancers is emerging and already impacting patient care, with identification of mismatch repair defects and CDK12 alterations leading to the choice of immunotherapy with anti-PD1/PD-L1 antibodies, while the presence of homologous recombination DNA repair defects – the commonest of which is BRCA2 loss – results in sensitization to PARP inhibition and treatment with drugs such as olaparib.

However, the majority of men suffering from mCRPCs do not yet have genomic results that generate tumour molecular stratification impacting treatment decisions, and it is therefore important to develop new understandings of prostate cancer genomics that can guide therapies. In addition, dissection of inter- and intra-patient heterogeneity will be key to improve patient care, clinical trial design, and ultimately disease outcomes.

The chief conclusions that I draw in this work are as follows:

8.1.1 Targeted Gene Panel NGS Can Define Key mCRPC Subtypes

- Between 20 and 40% of CRPC tumours can be identified as part of various DNA-repair defective subtypes and these can be classified using targeted-sequencing protocols. This requires specific methodologies and considerations.
- Procedures for calling mutations from targeted sequencing are well established, but those identifying copy-number changes are not, so I show here an orthogonally validated method for CNA identification from a 113-gene targeted amplicon sequencing panel data, including the detection of key events such as *BRCA2*.
- This method has been used successfully for delivering the TOPARP-B trial, with patients enrolled and treated on the basis of detected deletions in key HR genes. The preliminary results for this trial have been presented and show substantially elevated drug responses. Interestingly, preliminary genomic analyses of TOPARP-B have suggested that not all HR gene deficiencies result in the same sensitivity to PARP inhibition (Mateo et al., 2019).
- Deficiencies in mismatch repair are a clinically important molecular classification because of a targeted vulnerability to immunotherapies. I show that targeted sequencing can produce analysis of microsatellite instability for a measurement of overall deficiency in the MMR pathway.
- There remains, however, substantial heterogeneity even at the single cell level, presenting a biological limitation to panel sequencing of bulk cancer cell DNA. I showed that *RB1* deletions visible at the CNA level are due to marked cell-to-cell heterogeneity, with adjacent cells bearing variable copies of the gene by FISH analysis.
- Tumour biopsy data have a well-known limitation of tumour purity, with detectable CNAs becoming indistinguishable from background noise as tumour concentrations decrease. Additionally, estimating tumour purity from targeted sequencing is unreliable, due to the small number of potential data points. I show, however, that modifying sample collection methods *via* methods such as microdissection can alleviate some of these concerns by enriching tumour purity.

8.1.2 lpWGS of Cell-Free Plasma DNA Provides Accurate Insights Into mCRPC Tumour Burden

- I studied cfDNA from a cohort of patients treated on two Phase 3 clnical trials of taxane chemotherapy (FIRSTANA and PROSELICA), and showed that low-pass whole genome NGS allows for accurate genome wide CNA profiles that include various key mCRPC events, and I found that the distribution and frequency of events in these cohorts of samples matched those described in biopsy studies pursuing exome sequencing.
- Analysis of lpWGS permits the estimation of tumour purity, however, tumour DNA may not always be present at high concentrations in blood plasma, which is a biological limitation of this and similar studies.
- I propose that a hard limit of ~5% tumour purity is a plausible cutoff for excluding low-purity samples from downstream genomic studies, although ideally for accurate detection of small events samples should have 10% or greater.
- A low tumour purity value is not just a limitation, however, as it performs as a measure of tumour burden and can be considered highly prognostic. I show that baseline estimates of tumour purity from lpWGS of cfDNA are associated strongly with poor overall survival.
- I found that tumour purity, and other well established clinical factors for mCRPC, had no association with drug response – which is to be expected.
- As patients respond to therapy and tumours shrink, the concentration of cfDNA shed into the blood decreases. Studying tumour purity estimates from lpWGS by analyses of serial blood samples reveals stark differences between treatment responders and non-responders. This suggests a method for monitoring tumour burden throughout the course of treatment, in addition to potentially identifying new genomic events.

8.1.3 Genomic Copy-Number Signals of Taxane Chemotherapy

- Some data have shown an association between CNA burden) and mCRPC prognosis and taxane efficacy. I found, however, very modest effects, potentially due to the relatively homogenous group of advanced diseases in the patients studied.

- Unsupervised analyses of genome-wide changes in CNA frequency showed that, in particular, copy-altered regions on chromosome 6 (q23) and 15 (q14) appeared to be induced by taxane treatment.
- Further, several loci can be identified in an unsupervised manner that have strong multivariable associations with taxane response. In particular, chr15 (30-63.5Mb) copy-losses were predictive of better response and survival, while chr19 (12.5-24.5Mb) copy-gains were predictive of significantly worse response and survival.
- Functionally screening the genes found in these regions may be necessary for truly discerning the key players.

8.1.4 Supervised Analysis of Gene Copy-Number Events Identifies Potential Taxane Clinical Factors

- The cohort of taxane-treated tumours bears a set of characteristic mCRPC CNA changes in baseline cfDNA analyses.
- Despite previous reports, once low-tumour purity samples were excluded, I found only very modest associations of mCRPC CNAs with taxane response, which may explain why no genomic subtyping has been successful yet for these drugs.
- Copy-losses in the tumour suppressor gene *CDKN2A* were multivariably prognostic of poor survival in this cohort, although not predictive of drug response. Surprisingly, *RB1* deletions, however, were not. *CDKN2A* deletions (or perhaps, chr9p deletions) may therefore have been an overlooked prognostic factor in mCRPCs.
- I observed some changes in CNA frequency between cohorts and in longitudinal sample sets, including in *MAP3K7* (chr6), *PIK3R1* (chr5), *PIK3CB* (chr3) and *CDKN2A* (chr9). CNAs covering these loci may be selected for by chemotherapy, although this may not represent a taxane-specific signal.
- I found that copy-gains in several tubulin genes were significantly changed in frequency between the baseline FIRSTANA (taxane naïve) and PROSELICA (docetaxel exposed) cohorts.
- Univariable and multivariable statistics revealed significant associations for several tubulin-beta isoforms (including *TUBB8*, *TUBB4A* and *TUBB4B*) with survival and response, and interestingly, several tubulin-alpha isoforms (*TUBA3D*, *TUBA4A*).

- One limitation of this study is that the precise interactions of tubulin aberrations with taxane chemotherapy cannot be fully identified from this dataset, as CNAs offer only a partial examination of the possible sources of regulation for this gene pathway.
- The associations observed in this study require future validation. Downstream studies of tubulins in mCRPC will benefit from expression analysis and functional validation to confirm the impact of individual gene aberrations. Similarly, performing siRNA/CRISPR screens on the genes found within the unsupervised-screened regions will help to identify the genes driving changes in taxane sensitivity.

8.1.5 Single-cell Analyses Show Sub-Clonal CNAs and Illustrate Tumour Heterogeneity

- Since the presence of multiple subclones within a single biopsy, or multiple metastases generating cfDNA (ie. tumour heterogeneity) is a limiting factor in tumour genome analyses, I pursued single cell genomic analyses.
- The process of whole blood Apheresis increased yields of CTCs substantially.
- I found that single cell aCGH data was highly representative of the mCRPC genome, with characteristic CNA changes clearly visible at expected proportions.
- Some CTCs in this analyses were not well representative of the other same-patient samples, and this may be due to either a) true biological variability with low numbers of truly aberrant tumours generating CTCs that differ substantially from the initial clone but are otherwise viable, or b) a technological limitation, with dying cells, cells captured by the immune system, or cell clusters incorrectly identified as viable CTCs.
- Additionally, a relatively low quantity of input DNA, such as in CTCs, was capable of producing accurate CNA profiles with little visible biases.
- Single cell CNAs can resolve subclonal events that are challenging to observe in NGS of bulk tissue or cfDNA.
- Sampling of a large volume of CTCs can allow for studies that mimic true multi-region sequencing protocols, and potentially access cells from metastatic lesions that are difficult to biopsy – or detect metastases undetected by imaging modalities.

- My analyses indicate that tumour phylogenetics can be studied using this data source. I showed that single cell CNAs can be treated similarly to existing character state evolution analyses, and therefore benefit from a substantial body of work already developed for nucleotide and amino acid evolution studies.
- The single CTC analyses also show that this technology allows for a precise examination of tumour clone dynamics, and measure tumour heterogeneity, which some studies have linked to clinical outcome. The number of observable clusters may suggest underlying biological variability of the disease, or relate to the number of metastases present.

8.2 Future Perspectives

8.2.1 Clinical Genomics

The data presented here, alongside other recent publications, show that while large strides have been made in the genomic characterisations of mCRPC, several unmet needs remain. Identification of DNA repair-based therapies has provided substantial benefit, but resistance inevitably emerges. Additionally, not all patients respond to the same degree and for the same length of time – some patients with multiple metastases show benefit in a subset of tumours, but not all.

Resolving these issues will require multiple approaches. The optimal ordering of therapeutic agents has not been yet established, although such trials are currently ongoing, in particular exploring combinations of hormone therapies with targeted agents. Synergistic combination therapies are under development, with the goal of driving tumours into crises by exploiting multiple approaches of attack. For example, combining anti-androgen therapies with taxanes has shown some effect, with the CHAARTED trial results reporting significantly improved survival in combination-treated patients than monotherapy (Kyriakopoulos et al., 2018; Xu and Qiu, 2019). The PROpel trial currently seeks to evaluate PARP inhibition in combination with anti-androgens, and several studies are examining the potential for priming the immune system for enhancing the effect of immunotherapies. While these preliminary trials have shown some effects, it is not clear to what degree these trials will show true synergistic impact; as opposed to simply loading a patient with additional therapies. The role of genomics will be critical to dissect the results, and explore the potential to stratify patients for such drug regimens.

Genomics is also increasingly being used to screening patients for clinical trials, and although this application is currently limited to a limited number of cancer centres, the decreasing costs of and improved accessibility of NGS mean that this will become widespread. There is a substantial role that genomics can play in monitoring the disease of patients enrolled on trials (and eventually, all individuals suffering from cancer). This thesis shows not only methods to classify tumours for drug sensitivity, but also minimally invasive approaches to track disease progression and evolution. Studying liquid biopsies for data about tumour

burden may prove to be an invaluable prognostic tool for mCRPC patients, particularly in cases where traditional biopsies are inaccessible.

Unfortunately, in the majority of patients without genomic classification and following androgen deprivation, there is a paucity of targeted therapeutic approaches. Defining actionable subtypes for taxanes have been challenging in many tumour types, with a variety of resistance mechanisms and markers of sensitivity reported; and while the results in this thesis are suggestive, more work is needed to unpick markers of sensitivity and mechanisms of resistance. In particular, studying the differences in taxane-resistant tumours (both cell lines and patient samples) may provide insights into how these common therapies can be tailored. Validating the genomic regions with significant associations with taxane response identified in this work will be critical to driving true genomic stratification for taxane therapy. To achieve this, I plan to analyse siRNA/CRISPR screen data from a set of recently acquired (from collaborators) cell lines (C42, LNCaP, 22Rv1), including both taxane resistant and sensitive cultures. Preliminary expression analysis on these data has implicated tubulin alterations in taxane resistance, although this work is still ongoing.

8.2.2 Emergent Technologies

Advanced prostate cancers are complex tumours with numerous functional pathways impacted at various levels of regulation – this thesis was concerned primarily with gross copy-number changes, which, while certainly common in mCRPC, likely represent only a proportion of the oncogenic events that drive tumour growth and evolution. Additionally, truly characterising CNAs (as in, estimates of extra/fewer copies of genomic regions) with short read sequencing presents challenges. CNAs are essentially genomic scars from structural events or gross chromosomal abnormalities, and current sequencing methods and analytical technology limit true reconstruction of the cancer genome – which feature complex events within repeat regions, cryptic deletions and extrachromosomal structures. Methods such as FISH can provide clues – but are challenging to be scaled up to handle complex, aneuploidy tumours, which mCRPC frequently are.

New data sources need to be considered for further studies of mCRPC genomes to more deeply resolve these issues. Structural events are difficult to

automatically detect and characterise, but sequencing tools such as linked-read sequencing and long-read sequencing are capable of surpassing WGS and RNA-seq approaches. These approaches could be, for example, used to explore the repeat regions found in key genes – including the AR – that may include structural variation that is undetectable by conventional technologies. Preliminary data has indicated that these tools identify many more structural changes than short read NGS, including events present in healthy germline tissues.

Studies that not only reliably classify genomic aberrations but truly integrate them into unified analytical pipelines are key to handling the growing repertoire of sequencing tools available. Integrating CNA data (low-pass WGS or otherwise) with other data sources, for example transcriptomics (from RNA) and mutation calling (from exomes) offers the possibility for true multi-omic modelling. Joint analysis of such so-called 'vertical data', paired with clinical annotation, could be key to identify latent subtypes of mCRPCs not apparent from traditional analyses (Richardson et al., 2016). Key to this is, aside from collating such large datasets, an understanding of the relative impacts that given alterations can have, and merging data in ways that reflect the underlying biology.

Improving sampling methods offers a way to develop our understanding of genomics. Aside from multi-region sequencing and microdissection, single cells – both from dissociated fresh biopsy samples and from isolated CTCs – may be the best method to identify tumour subclones. In this thesis, I present only single CNA data, but methods for single cell RNA and methylation have shown promising results, albeit with a different set of caveats.

For further studies of CTC genomic analyses, I propose two simple constraints: 1. A plausible subclone population must be represented by at least x% of the total cells studied. 2. A minimum of two similar cells must be assigned to a group to be considered a subclone. If the value of x is 10%, application of these limits suggests a minimum CTC number of 20. In the cohort studied in chapter 7, using these criteria the majority of patients exhibited multiple clones. I am currently working to characterise more patients with large (>50) numbers of CTCs collected, and preliminary data has shown that these patients again display a mix of homogenous and heterogenous samples.

8.2.3 Closing Comments

In conclusion, in this thesis I have demonstrated how targeted panel NGS data can be used to classify the ~40% of mCRPC cases that are likely to respond to DNA-repair targeting therapies. I went on to explore the utility of cfDNA for calling genome-wide copy-number alterations in a cohort of individuals on two clinical trials of taxane chemotherapy, and showed that while tumour purity represents a liability for cancer sequencing, it also has clinical utility in associating with tumour burden and survival. Using this dataset, I explored several possible predictive biomarkers for taxane use, and pursued supervised analyses to test hypotheses focused on taxane-treatment related genes, including their target the tubulins. I also developed an unsupervised approach that identified several promising copy-altered regions that could provide predictive and prognostic biomarkers in the future. I believe that validation and clarification of these will provide key clinical stratifications for mCRPCs. Finally, to truly explore tumour heterogeneity, and bypass issues of tumour purity, I analysed a large set of single CTCs for copy-number alterations, and showed tumour diversity and sub-clonality at a single-cell resolution. I also demonstrated a novel application for existing tools to transform single-cell CNA data for phylogenetic reconstruction.

I believe the work presented here offers both technical and genomic answers to key issues in clinical practice, and that these approaches can help guide the treatment of advanced prostate cancer to improve patient care.

9 Bibliography

1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. Nature *526*, 68–74.

Adalsteinsson, V.A., Ha, G., Freeman, S.S., Choudhury, A.D., Stover, D.G., Parsons, H.A., Gydush, G., Reed, S.C., Rotem, D., Rhoades, J., et al. (2017). Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. Nat. Commun. *8*, 1324.

Agoulnik, I.U., and Weigel, N.L. (2009). Coactivator selective regulation of androgen receptor activity. Steroids *74*, 669–674.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Campbell, P.J., and Stratton, M.R. (2013). Deciphering signatures of mutational processes operative in human cancer. Cell Rep. *3*, 246–259.

Altman, D.G., McShane, L.M., Sauerbrei, W., and Taube, S.E. (2012). Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. BMC Med. *10*, 51.

Annala, M., Vandekerkhove, G., Khalaf, D., Taavitsainen, S., Beja, K., Warner, E.W., Sunderland, K., Kollmannsberger, C., Eigl, B.J., Finch, D., et al. (2018). Circulating Tumor DNA Genomics Correlate with Resistance to Abiraterone and Enzalutamide in Prostate Cancer. Cancer Discov. *8*, 444–457.

Armstrong, A.J., Eisenberger, M.A., Halabi, S., Oudard, S., Nanus, D.M., Petrylak, D.P., Sartor, A.O., and Scher, H.I. (2012). Biomarkers in the management and treatment of men with metastatic castration-resistant prostate cancer. Eur. Urol. *61*, 549–559.

Attard, G., Sarker, D., Reid, A., Molife, R., Parker, C., and de Bono, J.S. (2006). Improving the outcome of patients with castration-resistant prostate cancer through rational drug development. Br. J. Cancer *95*, 767–774.

Attard, G., Parker, C., Eeles, R.A., Schröder, F., Tomlins, S.A., Tannock, I., Drake, C.G., and de Bono, J.S. (2016). Prostate cancer. Lancet *387*,

70–82.

Aucamp, J., Bronkhorst, A.J., Badenhorst, C.P.S., and Pretorius, P.J. (2018). The diverse origins of circulating cell-free DNA in the human body: a critical re-evaluation of the literature. Biol. Rev. Camb. Philos. Soc. *93*, 1649–1683.

Baca, S.C., Prandi, D., Lawrence, M.S., Mosquera, J.M., Romanel, A., Drier, Y., Park, K., Kitabayashi, N., MacDonald, T.Y., Ghandi, M., et al. (2013). Punctuated evolution of prostate cancer genomes. Cell *153*, 666–677.

Bai, Y., Zhang, Z., Cheng, L., Wang, R., Chen, X., Kong, Y., Feng, F., Ahmad, N., Li, L., and Liu, X. (2019). Inhibition of enhancer of zeste homolog 2 (EZH2) overcomes enzalutamide resistance in castration-resistant prostate cancer. J. Biol. Chem. *294*, 9911–9923.

Bajrami, I., Frankum, J.R., Konde, A., Miller, R.E., Rehman, F.L., Brough, R., Campbell, J., Sims, D., Rafiq, R., Hooper, S., et al. (2014). Genome-wide profiling of genetic synthetic lethality identifies CDK12 as a novel determinant of PARP1/2 inhibitor sensitivity. Cancer Res. *74*, 287–297.

Balaban, M., Moshiri, N., Mai, U., and Mirarab, S. (2019). TreeCluster: clustering biological sequences using phylogenetic trees. BioRxiv.

Bao, S., Jiang, R., Kwan, W., Wang, B., Ma, X., and Song, Y.-Q. (2011). Evaluation of next-generation sequencing software in mapping and assembly. J. Hum. Genet. *56*, 406–414.

Barbieri, C.E., Baca, S.C., Lawrence, M.S., Demichelis, F., Blattner, M., Theurillat, J.-P., White, T.A., Stojanov, P., Van Allen, E., Stransky, N., et al. (2012). Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat. Genet. *44*, 685–689.

Bargiela-Iparraguirre, J., Prado-Marchal, L., Pajuelo-Lozano, N., Jiménez, B., Perona, R., and Sánchez-Pérez, I. (2014). Mad2 and BubR1 modulates tumourigenesis and paclitaxel response in MKN45 gastric cancer cells. Cell Cycle *13*, 3590–3601.

Bartley, A.N., Hamilton, S.R., Alsabeh, R., Ambinder, E.P., Berman, M., Collins, E., Fitzgibbons, P.L., Gress, D.M., Nowak, J.A., Samowitz, W.S., et al. (2014). Template for reporting results of biomarker testing of specimens from patients with carcinoma of the colon and rectum. Arch.

Pathol. Lab. Med. *138*, 166–170.

Baselga, J., Campone, M., Piccart, M., Burris, H.A., Rugo, H.S., Sahmoud, T., Noguchi, S., Gnant, M., Pritchard, K.I., Lebrun, F., et al. (2012). Everolimus in postmenopausal hormone-receptor-positive advanced breast cancer. N. Engl. J. Med. *366*, 520–529.

Beer, T.M., Kwon, E.D., Drake, C.G., Fizazi, K., Logothetis, C., Gravis, G., Ganju, V., Polikoff, J., Saad, F., Humanski, P., et al. (2017). Randomized, Double-Blind, Phase III Trial of Ipilimumab Versus Placebo in Asymptomatic or Minimally Symptomatic Patients With Metastatic Chemotherapy-Naive Castration-Resistant Prostate Cancer. J. Clin. Oncol. *35*, 40–47.

Beltran, H., Prandi, D., Mosquera, J.M., Benelli, M., Puca, L., Cyrta, J., Marotz, C., Giannopoulou, E., Chakravarthi, B.V.S.K., Varambally, S., et al. (2016). Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. Nat. Med. *22*, 298–305.

Benjamini, Y., and Speed, T.P. (2012). Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Res. *40*, e72.

Berger, M.F., Lawrence, M.S., Demichelis, F., Drier, Y., Cibulskis, K., Sivachenko, A.Y., Sboner, A., Esgueva, R., Pflueger, D., Sougnez, C., et al. (2011). The genomic complexity of primary human prostate cancer. Nature *470*, 214–220.

Berthold, D.R., Pond, G.R., Soban, F., de Wit, R., Eisenberger, M., and Tannock, I.F. (2008). Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cancer: updated survival in the TAX 327 study. J. Clin. Oncol. *26*, 242–245.

Bo, L., Cui, H., Fang, Z., Qun, T., and Xia, C. (2016). Inactivation of transforming growth factor-β-activated kinase 1 promotes taxol efficacy in ovarian cancer cells. Biomed. Pharmacother. *84*, 917–924.

de Bono, J.S., Scher, H.I., Montgomery, R.B., Parker, C., Miller, M.C., Tissing, H., Doyle, G.V., Terstappen, L.W.W.M., Pienta, K.J., and Raghavan, D. (2008). Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer. Clin. Cancer Res. *14*, 6302–6309.

de Bono, J.S., Oudard, S., Ozguroglu, M., Hansen, S., Machiels, J.-P., Kocak, I., Gravis, G., Bodrogi, I., Mackenzie, M.J., Shen, L., et al. (2010). Prednisone plus cabazitaxel or mitoxantrone for metastatic castration-resistant prostate cancer progressing after docetaxel treatment: a randomised open-label trial. Lancet *376*, 1147–1154.

de Bono, J.S., De Giorgi, U., Rodrigues, D.N., Massard, C., Bracarda, S., Font, A., Arranz Arija, J.A., Shih, K.C., Radavoi, G.D., Xu, N., et al. (2019). Randomized Phase II Study Evaluating Akt Blockade with Ipatasertib, in Combination with Abiraterone, in Patients with Metastatic Prostate Cancer with and without PTEN Loss. Clin. Cancer Res. *25*, 928–936.

Borgel, J., Tyl, M., Schiller, K., Pusztai, Z., Dooley, C.M., Deng, W., Wooding, C., White, R.J., Warnecke, T., Leonhardt, H., et al. (2017). KDM2A integrates DNA and histone modification signals through a CXXC/PHD module and direct interaction with HP1. Nucleic Acids Res. *45*, 1114–1129.

Boysen, G., Rodrigues, D.N., Rescigno, P., Seed, G., Dolling, D., Riisnaes, R., Crespo, M., Zafeiriou, Z., Sumanasuriya, S., Bianchini, D., et al. (2018). SPOP-Mutated/CHD1-Deleted Lethal Prostate Cancer and Abiraterone Sensitivity. Clin. Cancer Res. *24*, 5585–5593.

Brochier, C., and Langley, B. (2013). Chromatin modifications associated with DNA double-strand breaks repair as potential targets for neurological diseases. Neurotherapeutics *10*, 817–830.

Bromberg, J.S., Brennan, D.C., Poggio, E., Bunnapradist, S., Langone, A., Sood, P., Matas, A.J., Mannon, R.B., Mehta, S., Sharfuddin, A., et al. (2017). Biological Variation of Donor-Derived Cell-Free DNA in Renal Transplant Recipients: Clinical Implications. Jrnl App Lab Med *2*, 309–321.

Bronkhorst, A.J., Ungerer, V., and Holdenrieder, S. (2019). The emerging role of cell-free DNA as a molecular marker for cancer management. Biomolecular Detection and Quantification *17*, 100087.

Brooke, G.N., Parker, M.G., and Bevan, C.L. (2008). Mechanisms of androgen receptor activation in advanced prostate cancer: differential co-activator recruitment and gene expression. Oncogene *27*, 2941–2950.

Buchanan, G., Yang, M., Cheong, A., Harris, J.M., Irvine, R.A., Lambert, P.F., Moore, N.L., Raynor, M., Neufing, P.J., Coetzee, G.A., et al. (2004).

Structural and functional consequences of glutamine tract variation in the androgen receptor. Hum. Mol. Genet. *13*, 1677–1692.

Budd, G.T., Cristofanilli, M., Ellis, M.J., Stopeck, A., Borden, E., Miller, M.C., Matera, J., Repollet, M., Doyle, G.V., Terstappen, L.W.M.M., et al. (2006). Circulating tumor cells versus imaging--predicting overall survival in metastatic breast cancer. Clin. Cancer Res. *12*, 6403–6409.

Bumbaca, B., and Li, W. (2018). Taxane resistance in castration-resistant prostate cancer: mechanisms and therapeutic strategies. Acta Pharm. Sin. B *8*, 518–529.

Cancer Genome Atlas Research Network (2015). The molecular taxonomy of primary prostate cancer. Cell *163*, 1011–1025.

Canzar, S., and Salzberg, S.L. (2017). Short read mapping: an algorithmic tour. Proc IEEE Inst Electr Electron Eng *105*, 436–458.

Cao, Z., Wei, L., Zhu, W., and Yao, X. (2018). Meta-analysis of CDKN2A methylation to find its role in prostate cancer development and progression, and also to find the effect of CDKN2A expression on disease-free survival (PRISMA). Medicine *97*, e0182.

Carlson, P., Dasgupta, A., Grzelak, C.A., Kim, J., Barrett, A., Coleman, I.M., Shor, R.E., Goddard, E.T., Dai, J., Schweitzer, E.M., et al. (2019). Targeting the perivascular niche sensitizes disseminated tumour cells to chemotherapy. Nat. Cell Biol. *21*, 238–250.

Carnero, A., and Paramio, J.M. (2014). The PTEN/PI3K/AKT Pathway in vivo, Cancer Mouse Models. Front. Oncol. *4*, 252.

Carreira, S., Romanel, A., Goodall, J., Grist, E., Ferraldeschi, R., Miranda, S., Prandi, D., Lorente, D., Frenel, J.-S., Pezaro, C., et al. (2014). Tumor clone dynamics in lethal prostate cancer. Sci. Transl. Med. *6*, 254ra125.

Carver, B.S., Chapinski, C., Wongvipat, J., Hieronymus, H., Chen, Y., Chandarlapaty, S., Arora, V.K., Le, C., Koutcher, J., Scher, H., et al. (2011). Reciprocal feedback regulation of PI3K and androgen receptor signaling in PTEN-deficient prostate cancer. Cancer Cell *19*, 575–586.

Chang, H.H.Y., Pannunzio, N.R., Adachi, N., and Lieber, M.R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nat. Rev. Mol. Cell Biol. *18*, 495–506.

Chen, C.D., Welsbie, D.S., Tran, C., Baek, S.H., Chen, R., Vessella, R.,

Rosenfeld, M.G., and Sawyers, C.L. (2004). Molecular determinants of resistance to antiandrogen therapy. Nat. Med. *10*, 33–39.

Chen, E.J., Sowalsky, A.G., Gao, S., Cai, C., Voznesensky, O., Schaefer, R., Loda, M., True, L.D., Ye, H., Troncoso, P., et al. (2015). Abiraterone treatment in castration-resistant prostate cancer selects for progesterone responsive mutant androgen receptors. Clin. Cancer Res. *21*, 1273–1280.

Chen, S., Xu, Y., Yuan, X., Bubley, G.J., and Balk, S.P. (2006). Androgen receptor phosphorylation and stabilization in prostate cancer by cyclin-dependent kinase 1. Proc. Natl. Acad. Sci. USA *103*, 15969–15974.

Cheng, H.H., Pritchard, C.C., Montgomery, B., Lin, D.W., and Nelson, P.S. (2017). Prostate cancer screening in a new era of genetics. Clin. Genitourin. Cancer *15*, 625–628.

Cirillo, L., Gotta, M., and Meraldi, P. (2017). The elephant in the room: the role of microtubules in cancer. Adv. Exp. Med. Biol. *1002*, 93–124.

Compiani, M., and Capriotti, E. (2013). Computational and theoretical methods for protein folding. Biochemistry *52*, 8601–8624.

Comstock, C.E.S., Augello, M.A., Goodwin, J.F., de Leeuw, R., Schiewer, M.J., Ostrander, W.F., Burkhart, R.A., McClendon, A.K., McCue, P.A., Trabulsi, E.J., et al. (2013). Targeting cell cycle and hormone receptor pathways in cancer. Oncogene *32*, 5481–5491.

Coutinho, I., Day, T.K., Tilley, W.D., and Selth, L.A. (2016). Androgen receptor signaling in castration-resistant prostate cancer: a lesson in persistence. Endocr. Relat. Cancer *23*, T179–T197.

Crea, F., Sun, L., Mai, A., Chiang, Y.T., Farrar, W.L., Danesi, R., and Helgason, C.D. (2012). The emerging role of histone lysine demethylases in prostate cancer. Mol. Cancer *11*, 52.

Cristofanilli, M., Budd, G.T., Ellis, M.J., Stopeck, A., Matera, J., Miller, M.C., Reuben, J.M., Doyle, G.V., Allard, W.J., Terstappen, L.W.M.M., et al. (2004). Circulating tumor cells, disease progression, and survival in metastatic breast cancer. N. Engl. J. Med. *351*, 781–791.

Crumbaker, M., Khoja, L., and Joshua, A.M. (2017). AR signaling and the PI3K pathway in prostate cancer. Cancers (Basel) *9*.

Cutress, M.L., Whitaker, H.C., Mills, I.G., Stewart, M., and Neal, D.E. (2008). Structural basis for the nuclear import of the human androgen

receptor. J. Cell Sci. *121*, 957–968.

Danila, D.C., Heller, G., Gignac, G.A., Gonzalez-Espinoza, R., Anand, A., Tanaka, E., Lilja, H., Schwartz, L., Larson, S., Fleisher, M., et al. (2007). Circulating tumor cell number and prognosis in progressive castration-resistant prostate cancer. Clin. Cancer Res. *13*, 7053–7058.

Davey, R.A., and Grossmann, M. (2016). Androgen receptor structure, function and biology: from bench to bedside. Clin Biochem Rev *37*, 3–15.

De Gendt, K., Verhoeven, G., Amieux, P.S., and Wilkinson, M.F. (2014). Genome-wide identification of AR-regulated genes translated in Sertoli cells in vivo using the RiboTag approach. Mol. Endocrinol. *28*, 575–591.

Dehm, S.M., Schmidt, L.J., Heemers, H.V., Vessella, R.L., and Tindall, D.J. (2008). Splicing of a novel androgen receptor exon generates a constitutively active androgen receptor that mediates prostate cancer therapy resistance. Cancer Res. *68*, 5469–5477.

Di Mitri, D., Toso, A., Chen, J.J., Sarti, M., Pinton, S., Jost, T.R., D'Antuono, R., Montani, E., Garcia-Escudero, R., Guccini, I., et al. (2014). Tumour-infiltrating Gr-1+ myeloid cells antagonize senescence in cancer. Nature *515*, 134–137.

Duran, G.E., Wang, Y.C., Francisco, E.B., Rose, J.C., Martinez, F.J., Coller, J., Brassard, D., Vrignaud, P., and Sikic, B.I. (2015). Mechanisms of resistance to cabazitaxel. Mol. Cancer Ther. *14*, 193–201.

Edwards, J., Krishna, N.S., Grigor, K.M., and Bartlett, J.M.S. (2003). Androgen receptor gene amplification and protein expression in hormone refractory prostate cancer. Br. J. Cancer *89*, 552–556.

Eeles, R., Goh, C., Castro, E., Bancroft, E., Guy, M., Al Olama, A.A., Easton, D., and Kote-Jarai, Z. (2014). The genetic epidemiology of prostate cancer and its clinical implications. Nat. Rev. Urol. *11*, 18–31.

Efstathiou, E., Titus, M., Wen, S., Hoang, A., Karlou, M., Ashe, R., Tu, S.M., Aparicio, A., Troncoso, P., Mohler, J., et al. (2015). Molecular characterization of enzalutamide-treated bone metastatic castration-resistant prostate cancer. Eur. Urol. *67*, 53–60.

Eisenberger, M., Hardy-Bessard, A.-C., Kim, C.S., Géczi, L., Ford, D., Mourey, L., Carles, J., Parente, P., Font, A., Kacso, G., et al. (2017). Phase III Study Comparing a Reduced Dose of Cabazitaxel (20 mg/m2)

and the Currently Approved Dose (25 mg/m2) in Postdocetaxel Patients With Metastatic Castration-Resistant Prostate Cancer-PROSELICA. J. Clin. Oncol. *35*, 3198–3206.

Ekumi, K.M., Paculova, H., Lenasi, T., Pospichalova, V., Bösken, C.A., Rybarikova, J., Bryja, V., Geyer, M., Blazek, D., and Barboric, M. (2015). Ovarian carcinoma CDK12 mutations misregulate expression of DNA repair genes via deficient formation and function of the Cdk12/CycK complex. Nucleic Acids Res. *43*, 2575–2589.

El-Khattouti, A., Selimovic, D., Haïkel, Y., Megahed, M., Gomez, C.R., and Hassan, M. (2014). Identification and analysis of CD133(+) melanoma stem-like cells conferring resistance to taxol: An insight into the mechanisms of their resistance and response. Cancer Lett. *343*, 123–133.

Ferraldeschi, R., Nava Rodrigues, D., Riisnaes, R., Miranda, S., Figueiredo, I., Rescigno, P., Ravi, P., Pezaro, C., Omlin, A., Lorente, D., et al. (2015). PTEN protein loss and clinical outcome from castration-resistant prostate cancer treated with abiraterone acetate. Eur. Urol. *67*, 795–802.

Filson, C.P., Marks, L.S., and Litwin, M.S. (2015). Expectant management for men with early stage prostate cancer. CA Cancer J Clin *65*, 265–282.

Fitzpatrick, J.M., and de Wit, R. (2014). Taxane mechanisms of action: potential implications for treatment sequencing in metastatic castration-resistant prostate cancer. Eur. Urol. *65*, 1198–1204.

Fong, P.C., Boss, D.S., Yap, T.A., Tutt, A., Wu, P., Mergui-Roelvink, M., Mortimer, P., Swaisland, H., Lau, A., O'Connor, M.J., et al. (2009). Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. N. Engl. J. Med. *361*, 123–134.

Forbes, S.A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., Cole, C.G., Ward, S., Dawson, E., Ponting, L., et al. (2017). COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. *45*, D777–D783.

Fraser, M., Sabelnykova, V.Y., Yamaguchi, T.N., Heisler, L.E., Livingstone, J., Huang, V., Shiah, Y.-J., Yousif, F., Lin, X., Masella, A.P., et al. (2017). Genomic hallmarks of localized, non-indolent prostate cancer. Nature *541*, 359–364.

Frescas, D., Guardavaccaro, D., Kuchay, S.M., Kato, H., Poleshko, A., Basrur, V., Elenitoba-Johnson, K.S., Katz, R.A., and Pagano, M. (2008). KDM2A represses transcription of centromeric satellite repeats and maintains the heterochromatic state. Cell Cycle *7*, 3539–3547.

Fumoleau, P., Trigo, J.M., Isambert, N., Sémiond, D., Gupta, S., and Campone, M. (2013). Phase I dose-finding study of cabazitaxel administered weekly in patients with advanced solid tumours. BMC Cancer *13*, 460.

Galletti, G., Leach, B.I., Lam, L., and Tagawa, S.T. (2017). Mechanisms of resistance to systemic therapy in metastatic castration-resistant prostate cancer. Cancer Treat. Rev. *57*, 16–27.

Gandaglia, G., Abdollah, F., Schiffmann, J., Trudeau, V., Shariat, S.F., Kim, S.P., Perrotte, P., Montorsi, F., Briganti, A., Trinh, Q.-D., et al. (2014). Distribution of metastatic sites in patients with prostate cancer: A population-based analysis. Prostate *74*, 210–216.

Gandaglia, G., Karakiewicz, P.I., Briganti, A., Passoni, N.M., Schiffmann, J., Trudeau, V., Graefen, M., Montorsi, F., and Sun, M. (2015). Impact of the Site of Metastases on Survival in Patients with Metastatic Prostate Cancer. Eur. Urol. *68*, 325–334.

Ganguly, A., Yang, H., Sharma, R., Patel, K.D., and Cabral, F. (2012). The role of microtubules and their dynamics in cell migration. J. Biol. Chem. *287*, 43359–43369.

Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. BMC Bioinformatics *11*, 367.

Gehring, J.S., Fischer, B., Lawrence, M., and Huber, W. (2015). SomaticSignatures: inferring mutational signatures from single-nucleotide variants. Bioinformatics *31*, 3673–3675.

Gerhauser, C., Favero, F., Risch, T., Simon, R., Feuerbach, L., Assenov, Y., Heckmann, D., Sidiropoulos, N., Waszak, S.M., Hübschmann, D., et al. (2018). Molecular Evolution of Early-Onset Prostate Cancer Identifies Molecular Risk Markers and Clinical Trajectories. Cancer Cell *34*, 996–1011.e8.

Giannakakou, P., Sackett, D.L., Kang, Y.K., Zhan, Z., Buters, J.T., Fojo, T., and Poruchynsky, M.S. (1997). Paclitaxel-resistant human ovarian

cancer cells have mutant beta-tubulins that exhibit impaired paclitaxel-driven polymerization. J. Biol. Chem. *272*, 17118–17125.

Goldgar, D.E., Easton, D.F., Cannon-Albright, L.A., and Skolnick, M.H. (1994). Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. J. Natl. Cancer Inst. *86*, 1600–1608.

Gonzalez-Garay, M.L., Chang, L., Blade, K., Menick, D.R., and Cabral, F. (1999). A beta-tubulin leucine cluster involved in microtubule assembly and paclitaxel resistance. J. Biol. Chem. *274*, 23875–23882.

Goodall, J., Mateo, J., Yuan, W., Mossop, H., Porta, N., Miranda, S., Perez-Lopez, R., Dolling, D., Robinson, D.R., Sandhu, S., et al. (2017). Circulating Cell-Free DNA to Guide Prostate Cancer Treatment with PARP Inhibition. Cancer Discov. *7*, 1006–1017.

Graff, J.N., Alumkal, J.J., Drake, C.G., Thomas, G.V., Redmond, W.L., Farhad, M., Cetnar, J.P., Ey, F.S., Bergan, R.C., Slottke, R., et al. (2016). Early evidence of anti-PD-1 activity in enzalutamide-resistant prostate cancer. Oncotarget *7*, 52810–52817.

Grasso, C., Butler, T., Rhodes, K., Quist, M., Neff, T.L., Moore, S., Tomlins, S.A., Reinig, E., Beadling, C., Andersen, M., et al. (2015). Assessing copy number alterations in targeted, amplicon-based next-generation sequencing data. J Mol Diagn *17*, 53–63.

Grasso, C.S., Wu, Y.-M., Robinson, D.R., Cao, X., Dhanasekaran, S.M., Khan, A.P., Quist, M.J., Jing, X., Lonigro, R.J., Brenner, J.C., et al. (2012). The mutational landscape of lethal castration-resistant prostate cancer. Nature *487*, 239–243.

Greenman, C., Stephens, P., Smith, R., Dalgliesh, G.L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C., et al. (2007). Patterns of somatic mutation in human cancer genomes. Nature *446*, 153–158.

Guo, Z., Yang, X., Sun, F., Jiang, R., Linn, D.E., Chen, H., Chen, H., Kong, X., Melamed, J., Tepper, C.G., et al. (2009). A novel androgen receptor splice variant is up-regulated during prostate cancer progression and promotes androgen depletion-resistant growth. Cancer Res. *69*, 2305–2313.

Haffner, M.C., Mosbruger, T., Esopi, D.M., Fedor, H., Heaphy, C.M.,

Walker, D.A., Adejola, N., Gürel, M., Hicks, J., Meeker, A.K., et al. (2013). Tracking the clonal origin of lethal prostate cancer. J. Clin. Invest.

Halabi, S., Lin, C.-Y., Kelly, W.K., Fizazi, K.S., Moul, J.W., Kaplan, E.B., Morris, M.J., and Small, E.J. (2014). Updated prognostic model for predicting overall survival in first-line chemotherapy for patients with metastatic castration-resistant prostate cancer. J. Clin. Oncol. *32*, 671–677.

Hamid, A., Beltran, H., Choudhury, A.D., and Sweeney, C. (2019). Genomic predictors of benefit of docetaxel (D) and next-generation hormonal therapy (NHT) in metastatic castration resistant prostate cancer (mCRPC). J. Clin. Oncol. *37*, 5018–5018.

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

Hansen, A.R., Massard, C., Ott, P.A., Haas, N.B., Lopez, J.S., Ejadi, S., Wallmark, J.M., Keam, B., Delord, J.P., Aggarwal, R., et al. (2018). Pembrolizumab for advanced prostate adenocarcinoma: findings of the KEYNOTE-028 study. Ann. Oncol. *29*, 1807–1813.

Hara, T., Miyazaki, J., Araki, H., Yamaoka, M., Kanzaki, N., Kusaka, M., and Miyamoto, M. (2003). Novel mutations of androgen receptor: a possible mechanism of bicalutamide withdrawal syndrome. Cancer Res. *63*, 149–153.

Hieronymus, H., Schultz, N., Gopalan, A., Carver, B.S., Chang, M.T., Xiao, Y., Heguy, A., Huberman, K., Bernstein, M., Assel, M., et al. (2014). Copy number alteration burden predicts prostate cancer relapse. Proc. Natl. Acad. Sci. USA *111*, 11139–11144.

Hieronymus, H., Murali, R., Tin, A., Yadav, K., Abida, W., Moller, H., Berney, D., Scher, H., Carver, B., Scardino, P., et al. (2018). Tumor copy number alteration burden is a pan-cancer prognostic factor associated with recurrence and death. Elife *7*.

Horn, S., Leonardelli, S., Sucker, A., Schadendorf, D., Griewank, K.G., and Paschen, A. (2018). Tumor CDKN2A-Associated JAK2 Loss and Susceptibility to Immunotherapy Resistance. J. Natl. Cancer Inst. *110*, 677–681.

Hsu, F.-N., Chen, M.-C., Chiang, M.-C., Lin, E., Lee, Y.-T., Huang, P.-H.,

Lee, G.-S., and Lin, H. (2011). Regulation of androgen receptor and prostate cancer growth by cyclin-dependent kinase 5. J. Biol. Chem. *286*, 33141–33149.

Hu, R., Lu, C., Mostaghel, E.A., Yegnasubramanian, S., Gurel, M., Tannahill, C., Edwards, J., Isaacs, W.B., Nelson, P.S., Bluemn, E., et al. (2012). Distinct transcriptional programs mediated by the ligand-dependent full-length androgen receptor and its splice variants in castration-resistant prostate cancer. Cancer Res. *72*, 3457–3462.

Hwang, J.-E., Hong, J.-Y., Kim, K., Kim, S.-H., Choi, W.-Y., Kim, M.-J., Jung, S.-H., Shim, H.-J., Bae, W.-K., Hwang, E.-C., et al. (2013). Class III β-tubulin is a predictive marker for taxane-based chemotherapy in recurrent and metastatic gastric cancer. BMC Cancer *13*, 431.

Jain, R.K., Mehta, R.J., Nakshatri, H., Idrees, M.T., and Badve, S.S. (2011). High-level expression of forkhead-box protein A1 in metastatic prostate cancer. Histopathology *58*, 766–772.

Jamal-Hanjani, M., A'Hern, R., Birkbak, N.J., Gorman, P., Grönroos, E., Ngang, S., Nicola, P., Rahman, L., Thanopoulou, E., Kelly, G., et al. (2015). Extreme chromosomal instability forecasts improved outcome in ER-negative breast cancer: a prospective validation cohort study from the TACT trial. Ann. Oncol. *26*, 1340–1346.

Johansson, G., Andersson, D., Filges, S., Li, J., Muth, A., Godfrey, T.E., and Ståhlberg, A. (2019). Considerations and quality controls when analyzing cell-free tumor DNA. Biomolecular Detection and Quantification *17*, 100078.

Jordan, M.A., and Wilson, L. (2004). Microtubules as a target for anticancer drugs. Nat. Rev. Cancer *4*, 253–265.

Kampan, N.C., Madondo, M.T., McNally, O.M., Quinn, M., and Plebanski, M. (2015). Paclitaxel and its evolving role in the management of ovarian cancer. Biomed Res. Int. *2015*, 413076.

Kawata, H., Ishikura, N., Watanabe, M., Nishimoto, A., Tsunenari, T., and Aoki, Y. (2010). Prolonged treatment with bicalutamide induces androgen receptor overexpression and androgen hypersensitivity. Prostate *70*, 745–754.

Kibel, A.S., Suarez, B.K., Belani, J., Oh, J., Webster, R., Brophy-Ebbers,

M., Guo, C., Catalona, W.J., Picus, J., and Goodfellow, P.J. (2003). CDKN1A and CDKN1B polymorphisms and risk of advanced prostate carcinoma. Cancer Res. *63*, 2033–2036.

Kim, K.H., and Roberts, C.W.M. (2014). Mechanisms by which SMARCB1 loss drives rhabdoid tumor growth. Cancer Genet *207*, 365–372.

Kim, K.H., and Roberts, C.W.M. (2016). Targeting EZH2 in cancer. Nat. Med. *22*, 128–134.

Kirby, M., Hirst, C., and Crawford, E.D. (2011). Characterising the castration-resistant prostate cancer population: a systematic review. Int. J. Clin. Pract. *65*, 1180–1192.

Klein, C.A., Schmidt-Kittler, O., Schardt, J.A., Pantel, K., Speicher, M.R., and Riethmüller, G. (1999). Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. Proc. Natl. Acad. Sci. USA *96*, 4494–4499.

Krokan, H.E., and Bjørås, M. (2013). Base excision repair. Cold Spring Harb. Perspect. Biol. *5*, a012583.

Kumar, A., Coleman, I., Morrissey, C., Zhang, X., True, L.D., Gulati, R., Etzioni, R., Bolouri, H., Montgomery, B., White, T., et al. (2016). Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. Nat. Med. *22*, 369–378.

Kwon, E.D., Drake, C.G., Scher, H.I., Fizazi, K., Bossi, A., van den Eertwegh, A.J.M., Krainer, M., Houede, N., Santos, R., Mahammedi, H., et al. (2014). Ipilimumab versus placebo after radiotherapy in patients with metastatic castration-resistant prostate cancer that had progressed after docetaxel chemotherapy (CA184-043): a multicentre, randomised, double-blind, phase 3 trial. Lancet Oncol. *15*, 700–712.

Kyriakopoulos, C.E., Chen, Y.-H., Carducci, M.A., Liu, G., Jarrard, D.F., Hahn, N.M., Shevrin, D.H., Dreicer, R., Hussain, M., Eisenberger, M., et al. (2018). Chemohormonal Therapy in Metastatic Hormone-Sensitive Prostate Cancer: Long-Term Survival Analysis of the Randomized Phase III E3805 CHAARTED Trial. J. Clin. Oncol. *36*, 1080–1087.

Lallous, N., Volik, S.V., Awrey, S., Leblanc, E., Tse, R., Murillo, J., Singh, K., Azad, A.A., Wyatt, A.W., LeBihan, S., et al. (2016). Functional analysis

of androgen receptor mutations that confer anti-androgen resistance identified in circulating cell-free DNA from prostate cancer patients. Genome Biol. *17*, 10.

Lambros, M.B., Seed, G., Sumanasuriya, S., Gil, V., Crespo, M., Fontes, M., Chandler, R., Mehra, N., Fowler, G., Ebbs, B., et al. (2018). Single-Cell Analyses of Prostate Cancer Liquid Biopsies Acquired by Apheresis. Clin. Cancer Res. *24*, 5635–5644.

Landrum, M.J., Lee, J.M., Benson, M., Brown, G.R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., et al. (2018). ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. *46*, D1062–D1067.

Längst, G., and Manelyte, L. (2015). Chromatin remodelers: from function to dysfunction. Genes (Basel) *6*, 299–324.

Le, D.T., Uram, J.N., Wang, H., Bartlett, B.R., Kemberling, H., Eyring, A.D., Skora, A.D., Luber, B.S., Azad, N.S., Laheru, D., et al. (2015). PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N. Engl. J. Med. *372*, 2509–2520.

Ledermann, J., Harter, P., Gourley, C., Friedlander, M., Vergote, I., Rustin, G., Scott, C.L., Meier, W., Shapira-Frommer, R., Safra, T., et al. (2014). Olaparib maintenance therapy in patients with platinum-sensitive relapsed serous ovarian cancer: a preplanned retrospective analysis of outcomes by BRCA status in a randomised phase 2 trial. Lancet Oncol. *15*, 852–861.

de Leeuw, R., Berman-Booty, L.D., Schiewer, M.J., Ciment, S.J., Den, R.B., Dicker, A.P., Kelly, W.K., Trabulsi, E.J., Lallas, C.D., Gomella, L.G., et al. (2015). Novel actions of next-generation taxanes benefit advanced stages of prostate cancer. Clin. Cancer Res. *21*, 795–807.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.

Leung, M.L., Davis, A., Gao, R., Casasent, A., Wang, Y., Sei, E., Vilar, E., Maru, D., Kopetz, S., and Navin, N.E. (2017). Single-cell DNA sequencing reveals a late-dissemination model in metastatic colorectal cancer. Genome Res. *27*, 1287–1299.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, M.M., Datto, M., Duncavage, E.J., Kulkarni, S., Lindeman, N.I., Roy, S., Tsimberidou, A.M., Vnencak-Jones, C.L., Wolff, D.J., Younes, A., et al. (2017). Standards and guidelines for the interpretation and reporting of sequence variants in cancer: A joint consensus recommendation of the association for molecular pathology, american society of clinical oncology, and college of american pathologists. J Mol Diagn *19*, 4–23.

Lim, J.T.E., Mansukhani, M., and Weinstein, I.B. (2005). Cyclin-dependent kinase 6 associates with the androgen receptor and enhances its transcriptional activity in prostate cancer cells. Proc. Natl. Acad. Sci. USA *102*, 5156–5161.

Litwin, M.S., and Tan, H.-J. (2017). The diagnosis and treatment of prostate cancer: A review. JAMA *317*, 2532–2542.

Liu, B., Morrison, C.D., Johnson, C.S., Trump, D.L., Qin, M., Conroy, J.C., Wang, J., and Liu, S. (2013). Computational methods for detecting copy number variations in cancer genome using next generation sequencing: principles and challenges. Oncotarget *4*, 1868–1881.

Liu, D., Xu, W., Zhang, Z.-W., Qian, J., Zheng, H., Zhang, J., and Su, B. (2015). RB1 Polymorphism Contributes to the Efficacy of Platinum-Taxanes in Advanced Squamous Cell Lung Cancer. Asian Pac. J. Cancer Prev. *16*, 775–781.

Liu, P., Li, S., Gan, L., Kao, T.P., and Huang, H. (2008). A transcription-independent function of FOXO1 in inhibition of androgen-independent activation of the androgen receptor in prostate cancer cells. Cancer Res. *68*, 10290–10299.

Lockett, K.L., Snowhite, I.V., and Hu, J.J. (2005). Nucleotide-excision repair and prostate cancer risk. Cancer Lett. *220*, 125–135.

Loeser, H., Schallenberg, S., von Winterfeld, M., Tharun, L., Alakus, H., Hölscher, A., Bollschweiler, E., Buettner, R., Zander, T., and Quaas, A.

(2017). High protein and mRNA expression levels of TUBB3 (class III ß-tubulin) are associated with aggressive tumor features in esophageal adenocarcinomas. Oncotarget *8*, 115179–115189.

Lord, C.J., and Ashworth, A. (2017). PARP inhibitors: Synthetic lethality in the clinic. Science *355*, 1152–1158.

Lorente, D., Mateo, J., Perez-Lopez, R., de Bono, J.S., and Attard, G. (2015). Sequencing of agents in castration-resistant prostate cancer. Lancet Oncol. *16*, e279-92.

Malumbres, M., and Barbacid, M. (2007). Cell cycle kinases in cancer. Curr. Opin. Genet. Dev. *17*, 60–65.

Mandal, R.K., Gangwar, R., Kapoor, R., and Mittal, R.D. (2012). Polymorphisms in base-excision & nucleotide-excision repair genes & prostate cancer risk in north Indian population. Indian J. Med. Res. *135*, 64–71.

Marta, G.N., Hanna, S.A., Fernandes da Silva, J.L., and Carvalho, H. de A. (2013). Screening for prostate cancer: an updated review. Expert Rev Anticancer Ther *13*, 101–108.

Marugame, T., and Katanoda, K. (2006). International comparisons of cumulative risk of breast and prostate cancer, from cancer incidence in five continents Vol. VIII. Jpn. J. Clin. Oncol. *36*, 399–400.

Mateo, J., Carreira, S., Sandhu, S., Miranda, S., Mossop, H., Perez-Lopez, R., Nava Rodrigues, D., Robinson, D., Omlin, A., Tunariu, N., et al. (2015). DNA-Repair Defects and Olaparib in Metastatic Prostate Cancer. N. Engl. J. Med. *373*, 1697–1708.

Mateo, J., Boysen, G., Barbieri, C.E., Bryant, H.E., Castro, E., Nelson, P.S., Olmos, D., Pritchard, C.C., Rubin, M.A., and de Bono, J.S. (2017). DNA repair in prostate cancer: biology and clinical implications. Eur. Urol. *71*, 417–425.

Mateo, J., Porta, N., McGovern, U.B., Elliott, T., Jones, R.J., Syndikus, I., Ralph, C., Jain, S., Varughese, M.A., Parikh, O., et al. (2019). TOPARP-B: A phase II randomized trial of the poly(ADP)-ribose polymerase (PARP) inhibitor olaparib for metastatic castration resistant prostate cancers (mCRPC) with DNA damage repair (DDR) alterations. J. Clin. Oncol. *37*, 5005–5005.

McAllister, T.E., England, K.S., Hopkinson, R.J., Brennan, P.E., Kawamura, A., and Schofield, C.J. (2016). Recent progress in histone demethylase inhibitors. J. Med. Chem. *59*, 1308–1329.

Mehra, N., Dolling, D., Sumanasuriya, S., Christova, R., Pope, L., Carreira, S., Seed, G., Yuan, W., Goodall, J., Hall, E., et al. (2018). Plasma Cell-free DNA Concentration and Outcomes from Taxane Therapy in Metastatic Castration-resistant Prostate Cancer from Two Phase III Trials (FIRSTANA and PROSELICA). Eur. Urol. *74*, 283–291.

Melling, N., Thomsen, E., Tsourlakis, M.C., Kluth, M., Hube-Magg, C., Minner, S., Koop, C., Graefen, M., Heinzer, H., Wittmer, C., et al. (2015). Overexpression of enhancer of zeste homolog 2 (EZH2) characterizes an aggressive subset of prostate cancers and predicts patient prognosis independently from pre- and postoperatively assessed clinicopathological parameters. Carcinogenesis *36*, 1333–1340.

Messerschmidt, J.L., Prendergast, G.C., and Messerschmidt, G.L. (2016). How cancers escape immune destruction and mechanisms of action for the new significantly active immune therapies: helping nonimmunologists decipher recent advances. Oncologist *21*, 233–243.

Metzger, E., Wissmann, M., Yin, N., Müller, J.M., Schneider, R., Peters, A.H.F.M., Günther, T., Buettner, R., and Schüle, R. (2005). LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. Nature *437*, 436–439.

Miller, L.M., Xiao, H., Burd, B., Horwitz, S.B., Angeletti, R.H., and Verdier-Pinard, P. (2010). Methods in Tubulin Proteomics. In Microtubules, in Vitro, (Elsevier), pp. 105–126.

Mittal, K., Donthamsetty, S., Kaur, R., Yang, C., Gupta, M.V., Reid, M.D., Choi, D.H., Rida, P.C.G., and Aneja, R. (2017). Multinucleated polyploidy drives resistance to Docetaxel chemotherapy in prostate cancer. Br. J. Cancer *116*, 1186–1194.

Mottet, N., Bellmunt, J., Bolla, M., Briers, E., Cumberbatch, M.G., De Santis, M., Fossati, N., Gross, T., Henry, A.M., Joniau, S., et al. (2017). EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. Eur. Urol. *71*, 618–629.

Mucci, L.A., Hjelmborg, J.B., Harris, J.R., Czene, K., Havelick, D.J., Scheike, T., Graff, R.E., Holst, K., Möller, S., Unger, R.H., et al. (2016). Familial risk and heritability of cancer among twins in nordic countries. JAMA *315*, 68–76.

Mulholland, D.J., Tran, L.M., Li, Y., Cai, H., Morim, A., Wang, S., Plaisier, S., Garraway, I.P., Huang, J., Graeber, T.G., et al. (2011). Cell autonomous role of PTEN in regulating castration-resistant prostate cancer growth. Cancer Cell *19*, 792–804.

Muzzey, D., Evans, E.A., and Lieber, C. (2015). Understanding the basics of NGS: from mechanism to variant calling. Curr. Genet. Med. Rep. *3*, 158–165.

Nami, B., and Wang, Z. (2018). Genetics and expression profile of the tubulin gene superfamily in breast cancer subtypes and its relation to taxane resistance. Cancers (Basel) *10*.

Nava Rodrigues, D., Rescigno, P., Liu, D., Yuan, W., Carreira, S., Lambros, M.B., Seed, G., Mateo, J., Riisnaes, R., Mullane, S., et al. (2018). Immunogenomic analyses associate immunological alterations with mismatch repair defects in prostate cancer. J. Clin. Invest. *128*, 4441–4453.

Neves, R.P.L., Raba, K., Schmidt, O., Honisch, E., Meier-Stiegen, F., Behrens, B., Möhlendick, B., Fehm, T., Neubauer, H., Klein, C.A., et al. (2014). Genomic High-Resolution Profiling of Single CKpos/CD45neg Flow-Sorting Purified Circulating Tumor Cells from Patients with Metastatic Breast Cancer. Clin. Chem. *60*, 1290–1297.

Nygaard, A.D., Holdgaard, P.C., Spindler, K.L.G., Pallisgaard, N., and Jakobsen, A. (2014). The correlation between cell-free DNA and tumour burden was estimated by PET/CT in patients with advanced NSCLC. Br. J. Cancer *110*, 363–368.

Olatubosun, A., Väliaho, J., Härkönen, J., Thusberg, J., and Vihinen, M. (2012). PON-P: integrated predictor for pathogenicity of missense variants. Hum. Mutat. *33*, 1166–1174.

Olshen, A.B., Venkatraman, E.S., Lucito, R., and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics *5*, 557–572.

Ong, M., Carreira, S., Goodall, J., Mateo, J., Figueiredo, I., Rodrigues, D.N., Perkins, G., Seed, G., Yap, T.A., Attard, G., et al. (2014). Validation and utilisation of high-coverage next-generation sequencing to deliver the pharmacological audit trail. Br. J. Cancer *111*, 828–836.

Oprea-Lager, D.E., Bijnsdorp, I.V., VAN Moorselaar, R.J.A., VAN DEN Eertwegh, A.J.M., Hoekstra, O.S., and Geldof, A.A. (2013). ABCC4 Decreases docetaxel and not cabazitaxel efficacy in prostate cancer cells in vitro. Anticancer Res. *33*, 387–391.

Oudard, S., Fizazi, K., Sengeløv, L., Daugaard, G., Saad, F., Hansen, S., Hjälm-Eriksson, M., Jassem, J., Thiery-Vuillemin, A., Caffo, O., et al. (2017). Cabazitaxel Versus Docetaxel As First-Line Therapy for Patients With Metastatic Castration-Resistant Prostate Cancer: A Randomized Phase III Trial-FIRSTANA. J. Clin. Oncol. *35*, 3189–3197.

Owada-Ozaki, Y., Muto, S., Takagi, H., Inoue, T., Watanabe, Y., Fukuhara, M., Yamaura, T., Okabe, N., Matsumura, Y., Hasegawa, T., et al. (2018). Prognostic Impact of Tumor Mutation Burden in Patients With Completely Resected Non-Small Cell Lung Cancer: Brief Report. J. Thorac. Oncol. *13*, 1217–1221.

Parker, A.L., Kavallaris, M., and McCarroll, J.A. (2014). Microtubules and their role in cellular stress in cancer. Front. Oncol. *4*, 153.

Perkins, G., Yap, T.A., Pope, L., Cassidy, A.M., Dukes, J.P., Riisnaes, R., Massard, C., Cassier, P.A., Miranda, S., Clark, J., et al. (2012). Multi-purpose utility of circulating plasma DNA testing in patients with advanced cancers. PLoS One *7*, e47020.

Person, F., Wilczak, W., Hube-Magg, C., Burdelski, C., Möller-Koop, C., Simon, R., Noriega, M., Sauter, G., Steurer, S., Burdak-Rothkamm, S., et al. (2017). Prevalence of βIII-tubulin (TUBB3) expression in human normal tissues and cancers. Tumour Biol. *39*, 1010428317712166.

Pezaro, C., Omlin, A., Lorente, D., Rodrigues, D.N., Ferraldeschi, R., Bianchini, D., Mukherji, D., Riisnaes, R., Altavilla, A., Crespo, M., et al. (2014). Visceral disease in castration-resistant prostate cancer. Eur. Urol. *65*, 270–273.

Ploussard, G., Terry, S., Maillé, P., Allory, Y., Sirab, N., Kheuang, L., Soyeux, P., Nicolaiew, N., Coppolani, E., Paule, B., et al. (2010). Class III

beta-tubulin expression predicts prostate tumor aggressiveness and patient response to docetaxel-based chemotherapy. Cancer Res. *70*, 9253–9264.

Pomerantz, M.M., Li, F., Takeda, D.Y., Lenci, R., Chonkar, A., Chabot, M., Cejas, P., Vazquez, F., Cook, J., Shivdasani, R.A., et al. (2015). The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. Nat. Genet. *47*, 1346–1351.

Poulogiannis, G., Frayling, I.M., and Arends, M.J. (2010). DNA mismatch repair deficiency in sporadic colorectal cancer and Lynch syndrome. Histopathology *56*, 167–179.

Premasekharan, G., Gilbert, E., Okimoto, R.A., Hamirani, A., Lindquist, K.J., Ngo, V.T., Roy, R., Hough, J., Edwards, M., Paz, R., et al. (2016). An improved CTC isolation scheme for pairing with downstream genomics: Demonstrating clinical utility in metastatic prostate, lung and pancreatic cancer. Cancer Lett. *380*, 144–152.

Prensner, J.R., Iyer, M.K., Sahu, A., Asangani, I.A., Cao, Q., Patel, L., Vergara, I.A., Davicioni, E., Erho, N., Ghadessi, M., et al. (2013). The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. Nat. Genet. *45*, 1392–1398.

Pritchard, C.C., Morrissey, C., Kumar, A., Zhang, X., Smith, C., Coleman, I., Salipante, S.J., Milbank, J., Yu, M., Grady, W.M., et al. (2014). Complex MSH2 and MSH6 mutations in hypermutated microsatellite unstable advanced prostate cancer. Nat. Commun. *5*, 4988.

Pritchard, C.C., Mateo, J., Walsh, M.F., De Sarkar, N., Abida, W., Beltran, H., Garofalo, A., Gulati, R., Carreira, S., Eeles, R., et al. (2016). Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. N. Engl. J. Med. *375*, 443–453.

Punnoose, E.A., Ferraldeschi, R., Szafer-Glusman, E., Tucker, E.K., Mohan, S., Flohr, P., Riisnaes, R., Miranda, S., Figueiredo, I., Rodrigues, D.N., et al. (2015). PTEN loss in circulating tumour cells correlates with PTEN loss in fresh tumour tissue from castration-resistant prostate cancer patients. Br. J. Cancer *113*, 1225–1233.

Qin, J., Lee, H.-J., Wu, S.-P., Lin, S.-C., Lanz, R.B., Creighton, C.J., DeMayo, F.J., Tsai, S.Y., and Tsai, M.-J. (2014). Androgen deprivation–

induced NCoA2 promotes metastatic and castration-resistant prostate cancer. J. Clin. Invest.

Rahaman, M.H., Kumarasiri, M., Mekonnen, L.B., Yu, M., Diab, S., Albrecht, H., Milne, R.W., and Wang, S. (2016). Targeting CDK9: a promising therapeutic opportunity in prostate cancer. Endocr. Relat. Cancer *23*, T211–T226.

Ramakrishnan Geethakumari, P., Schiewer, M.J., Knudsen, K.E., and Kelly, W.K. (2017). PARP inhibitors in prostate cancer. Curr. Treat. Options Oncol. *18*, 37.

Ramensky, V., Bork, P., and Sunyaev, S. (2002). Human non-synonymous SNPs: server and survey. Nucleic Acids Res. *30*, 3894–3900.

Rasool, R.U., Natesan, R., Deng, Q., Aras, S., Lal, P., Sander Effron, S., Mitchell-Velasquez, E., Posimo, J.M., Carskadon, S., Baca, S.C., et al. (2019). CDK7 inhibition suppresses Castration-Resistant Prostate Cancer through MED1 inactivation. Cancer Discov.

Reid, A.H.M., Attard, G., Ambroisine, L., Fisher, G., Kovacs, G., Brewer, D., Clark, J., Flohr, P., Edwards, S., Berney, D.M., et al. (2010). Molecular characterisation of ERG, ETV1 and PTEN gene loci identifies patients at low and high risk of death from prostate cancer. Br. J. Cancer *102*, 678–684.

Reis, G.F., Pekmezci, M., Hansen, H.M., Rice, T., Marshall, R.E., Molinaro, A.M., Phillips, J.J., Vogel, H., Wiencke, J.K., Wrensch, M.R., et al. (2015). CDKN2A loss is associated with shortened overall survival in lower-grade (World Health Organization Grades II-III) astrocytomas. J. Neuropathol. Exp. Neurol. *74*, 442–452.

Richardson, S., Tseng, G.C., and Sun, W. (2016). Statistical methods in integrative genomics. Annu. Rev. Stat. Appl. *3*, 181–209.

Rizzo, J.M., and Buck, M.J. (2012). Key principles and clinical applications of “next-generation” DNA sequencing. Cancer Prev Res (Phila Pa) *5*, 887–900.

Robinson, D., Van Allen, E.M., Wu, Y.-M., Schultz, N., Lonigro, R.J., Mosquera, J.-M., Montgomery, B., Taplin, M.-E., Pritchard, C.C., Attard, G., et al. (2015). Integrative clinical genomics of advanced prostate

cancer. Cell *161*, 1215–1228.

Rodrigues, D.N., Boysen, G., Sumanasuriya, S., Seed, G., Marzo, A.M.D., and de Bono, J. (2017). The molecular underpinnings of prostate cancer: impacts on management and pathology practice. J. Pathol. *241*, 173–182.

Rodrigues, L.U., Rider, L., Nieto, C., Romero, L., Karimpour-Fard, A., Loda, M., Lucia, M.S., Wu, M., Shi, L., Cimic, A., et al. (2015). Coordinate loss of MAP3K7 and CHD1 promotes aggressive prostate cancer. Cancer Res. *75*, 1021–1034.

Ruggero, K., Farran-Matas, S., Martinez-Tebar, A., and Aytes, A. (2018). Epigenetic regulation in prostate cancer progression. Curr. Mol. Biol. Rep. *4*, 101–115.

Saeidi, H.R., Lohrasebi, A., and Mahnam, K. (2014). External electric field effects on the mechanical properties of the αβ-tubulin dimer of microtubules: a molecular dynamics study. J Mol Model *20*, 2395.

Salipante, S.J., Scroggins, S.M., Hampel, H.L., Turner, E.H., and Pritchard, C.C. (2014). Microsatellite instability detection by next generation sequencing. Clin. Chem. *60*, 1192–1199.

Samstein, R.M., Lee, C.-H., Shoushtari, A.N., Hellmann, M.D., Shen, R., Janjigian, Y.Y., Barron, D.A., Zehir, A., Jordan, E.J., Omuro, A., et al. (2019). Tumor mutational load predicts survival after immunotherapy across multiple cancer types. Nat. Genet. *51*, 202–206.

Samuels, Y., and Waldman, T. (2010). Oncogenic mutations of PIK3CA in human cancers. Curr. Top. Microbiol. Immunol. *347*, 21–41.

Sarker, D., Reid, A.H.M., Yap, T.A., and de Bono, J.S. (2009). Targeting the PI3K/AKT pathway for the treatment of prostate cancer. Clin. Cancer Res. *15*, 4799–4805.

Sartor, O., and de Bono, J.S. (2018). Metastatic Prostate Cancer. N. Engl. J. Med. *378*, 645–657.

Sauerbrei, W., Taube, S.E., McShane, L.M., Cavenagh, M.M., and Altman, D.G. (2018). Reporting recommendations for tumor marker prognostic studies (REMARK): an abridged explanation and elaboration. J. Natl. Cancer Inst. *110*, 803–811.

Schiewer, M.J., Augello, M.A., and Knudsen, K.E. (2012). The AR dependent cell cycle: mechanisms and cancer relevance. Mol. Cell.

Endocrinol. *352*, 34–45.

Seed, G., Yuan, W., Mateo, J., Carreira, S., Bertan, C., Lambros, M., Boysen, G., Ferraldeschi, R., Miranda, S., Figueiredo, I., et al. (2017). Gene Copy Number Estimation from Targeted Next-Generation Sequencing of Prostate Cancer Biopsies: Analytic Validation and Clinical Qualification. Clin. Cancer Res. *23*, 6070–6077.

Seligson, D.B., Horvath, S., Shi, T., Yu, H., Tze, S., Grunstein, M., and Kurdistani, S.K. (2005). Global histone modification patterns predict risk of prostate cancer recurrence. Nature *435*, 1262–1266.

Sharma, A., Yeow, W.-S., Ertel, A., Coleman, I., Clegg, N., Thangavel, C., Morrissey, C., Zhang, X., Comstock, C.E.S., Witkiewicz, A.K., et al. (2010). The retinoblastoma tumor suppressor controls androgen signaling and human prostate cancer progression. J. Clin. Invest. *120*, 4478–4492.

Sharma, N.L., Massie, C.E., Ramos-Montoya, A., Zecchini, V., Scott, H.E., Lamb, A.D., MacArthur, S., Stark, R., Warren, A.Y., Mills, I.G., et al. (2013). The androgen receptor induces a distinct transcriptional program in castration-resistant prostate cancer in man. Cancer Cell *23*, 35–47.

Shenoy, T.R., Boysen, G., Wang, M.Y., Xu, Q.Z., Guo, W., Koh, F.M., Wang, C., Zhang, L.Z., Wang, Y., Gil, V., et al. (2017). CHD1 loss sensitizes prostate cancer to DNA damaging therapy by promoting error-prone double-strand break repair. Ann. Oncol. *28*, 1495–1507.

Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. *29*, 308–311.

Sissung, T.M., Baum, C.E., Deeken, J., Price, D.K., Aragon-Ching, J., Steinberg, S.M., Dahut, W., Sparreboom, A., and Figg, W.D. (2008). ABCB1 genetic variation influences the toxicity and clinical outcome of patients with androgen-independent prostate cancer treated with docetaxel. Clin. Cancer Res. *14*, 4543–4549.

Sobierajska, K., Ciszewski, W.M., Wawro, M.E., Wieczorek-Szukała, K., Boncela, J., Papiewska-Pajak, I., Niewiarowska, J., and Kowalska, M.A. (2019). TUBB4B downregulation is critical for increasing migration of metastatic colon cancer cells. Cells *8*.

Sobue, S., Mizutani, N., Aoyama, Y., Kawamoto, Y., Suzuki, M., Nozawa,

Y., Ichihara, M., and Murate, T. (2016). Mechanism of paclitaxel resistance in a human prostate cancer cell line, PC3-PR, and its sensitization by cabazitaxel. Biochem. Biophys. Res. Commun. *479*, 808–813.

Steinkamp, M.P., O'Mahony, O.A., Brogley, M., Rehman, H., Lapensee, E.W., Dhanasekaran, S., Hofer, M.D., Kuefer, R., Chinnaiyan, A., Rubin, M.A., et al. (2009). Treatment-dependent androgen receptor mutations in prostate cancer exploit multiple mechanisms to evade therapy. Cancer Res. *69*, 4434–4442.

Stoecklein, N.H., Fischer, J.C., Niederacher, D., and Terstappen, L.W.M.M. (2016). Challenges for CTC-based liquid biopsies: low CTC frequency and diagnostic leukapheresis as a potential solution. Expert Rev Mol Diagn *16*, 147–164.

Stratmann, A., and Haendler, B. (2012). Histone demethylation and steroid receptor function in cancer. Mol. Cell. Endocrinol. *348*, 12–20.

Swanton, C., Nicke, B., Schuett, M., Eklund, A.C., Ng, C., Li, Q., Hardcastle, T., Lee, A., Roy, R., East, P., et al. (2009). Chromosomal instability determines taxane response. Proc. Natl. Acad. Sci. USA *106*, 8671–8676.

Talevich, E., Shain, A.H., Botton, T., and Bastian, B.C. (2016). CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS Comput. Biol. *12*, e1004873.

Tan, H.-L., Sood, A., Rahimi, H.A., Wang, W., Gupta, N., Hicks, J., Mosier, S., Gocke, C.D., Epstein, J.I., Netto, G.J., et al. (2014). Rb loss is characteristic of prostatic small cell neuroendocrine carcinoma. Clin. Cancer Res. *20*, 890–903.

Tan, J., Sharief, Y., Hamil, K.G., Gregory, C.W., Zang, D.Y., Sar, M., Gumerlock, P.H., deVere White, R.W., Pretlow, T.G., Harris, S.E., et al. (1997). Dehydroepiandrosterone activates mutant androgen receptors expressed in the androgen-dependent human prostate cancer xenograft CWR22 and LNCaP cells. Mol. Endocrinol. *11*, 450–459.

Tan, M.H.E., Li, J., Xu, H.E., Melcher, K., and Yong, E. (2015). Androgen receptor: structure, role in prostate cancer and drug discovery. Acta Pharmacol Sin *36*, 3–23.

Tang, H., and Thomas, P.D. (2016). Tools for predicting the functional

impact of nonsynonymous genetic variation. Genetics *203*, 635–647.

Tannock, I.F., de Wit, R., Berry, W.R., Horti, J., Pluzanska, A., Chi, K.N., Oudard, S., Théodore, C., James, N.D., Turesson, I., et al. (2004). Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cancer. N. Engl. J. Med. *351*, 1502–1512.

Tarish, F.L., Schultz, N., Tanoglidi, A., Hamberg, H., Letocha, H., Karaszi, K., Hamdy, F.C., Granfors, T., and Helleday, T. (2015). Castration radiosensitizes prostate cancer tissue by impairing DNA double-strand break repair. Sci. Transl. Med. *7*, 312re11.

Taylor, A.C. (1997). Titration of heparinase for removal of the PCR-inhibitory effect of heparin in DNA samples. Mol. Ecol. *6*, 383–385.

Telli, M.L., Timms, K.M., Reid, J., Hennessy, B., Mills, G.B., Jensen, K.C., Szallasi, Z., Barry, W.T., Winer, E.P., Tung, N.M., et al. (2016). Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. Clin. Cancer Res. *22*, 3764–3773.

Teply, B.A., and Antonarakis, E.S. (2017). Treatment strategies for DNA repair-deficient prostate cancer. Expert Rev Clin Pharmacol *10*, 889–898.

Tosoian, J.J., Mamawala, M., Epstein, J.I., Landis, P., Wolf, S., Trock, B.J., and Carter, H.B. (2015). Intermediate and Longer-Term Outcomes From a Prospective Active-Surveillance Program for Favorable-Risk Prostate Cancer. J. Clin. Oncol. *33*, 3379–3385.

Treangen, T.J., and Salzberg, S.L. (2011). Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat. Rev. Genet. *13*, 36–46.

Turajlic, S., Litchfield, K., Xu, H., Rosenthal, R., McGranahan, N., Reading, J.L., Wong, Y.N.S., Rowan, A., Kanu, N., Al Bakir, M., et al. (2017). Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. Lancet Oncol. *18*, 1009–1021.

Turner, K.M., Deshpande, V., Beyter, D., Koga, T., Rusert, J., Lee, C., Li, B., Arden, K., Ren, B., Nathanson, D.A., et al. (2017). Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. Nature *543*, 122–125.

Ulz, P., Belic, J., Graf, R., Auer, M., Lafer, I., Fischereder, K., Webersinke, G., Pummer, K., Augustin, H., Pichler, M., et al. (2016). Whole-genome plasma sequencing reveals focal amplifications as a driving force in metastatic prostate cancer. Nat. Commun. *7*, 12008.

Underhill, C., Toulmonde, M., and Bonnefoi, H. (2011). A review of PARP inhibitors: from bench to bedside. Ann. Oncol. *22*, 268–279.

Valpione, S., Gremel, G., Mundra, P., Middlehurst, P., Galvani, E., Girotti, M.R., Lee, R.J., Garner, G., Dhomen, N., Lorigan, P.C., et al. (2018). Plasma total cell-free DNA (cfDNA) is a surrogate biomarker for tumour burden and a prognostic biomarker for survival in metastatic melanoma patients. Eur. J. Cancer *88*, 1–9.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics *11*, 11.10.1-11.10.33.

Van Hoeck, A., Tjoonk, N.H., van Boxtel, R., and Cuppen, E. (2019). Portrait of a cancer: mutational signature analyses for cancer diagnostics. BMC Cancer *19*, 457.

Vemu, A., Atherton, J., Spector, J.O., Moores, C.A., and Roll-Mecak, A. (2017). Tubulin isoform composition tunes microtubule dynamics. Mol. Biol. Cell *28*, 3564–3572.

Viswanathan, S.R., Ha, G., Hoff, A.M., Wala, J.A., Carrot-Zhang, J., Whelan, C.W., Haradhvala, N.J., Freeman, S.S., Reed, S.C., Rhoades, J., et al. (2018). Structural Alterations Driving Castration-Resistant Prostate Cancer Revealed by Linked-Read Genome Sequencing. Cell *174*, 433–447.e19.

Vivanco, I., and Sawyers, C.L. (2002). The phosphatidylinositol 3-Kinase AKT pathway in human cancer. Nat. Rev. Cancer *2*, 489–501.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., and Kinzler, K.W. (2013). Cancer genome landscapes. Science *339*, 1546–1558.

Wadosky, K.M., and Koochekpour, S. (2017). Androgen receptor splice variants and prostate cancer: From bench to bedside. Oncotarget *8*,

18550–18576.

Wan, J.C.M., Massie, C., Garcia-Corbacho, J., Mouliere, F., Brenton, J.D., Caldas, C., Pacey, S., Baird, R., and Rosenfeld, N. (2017). Liquid biopsies come of age: towards implementation of circulating tumour DNA. Nat. Rev. Cancer *17*, 223–238.

Wang, X., Pan, L., Mao, N., Sun, L., Qin, X., and Yin, J. (2013). Cell-cycle synchronization reverses Taxol resistance of human ovarian cancer cell lines. Cancer Cell Int. *13*, 77.

Watson, P.A., Chen, Y.F., Balbas, M.D., Wongvipat, J., Socci, N.D., Viale, A., Kim, K., and Sawyers, C.L. (2010). Constitutively active androgen receptor splice variants expressed in castration-resistant prostate cancer require full-length androgen receptor. Proc. Natl. Acad. Sci. USA *107*, 16759–16765.

Wee, Z.N., Li, Z., Lee, P.L., Lee, S.T., Lim, Y.P., and Yu, Q. (2014). EZH2-mediated inactivation of IFN-γ-JAK-STAT1 signaling is an effective therapeutic target in MYC-driven prostate cancer. Cell Rep. *8*, 204–216.

Weischenfeldt, J., Simon, R., Feuerbach, L., Schlangen, K., Weichenhan, D., Minner, S., Wuttig, D., Warnatz, H.-J., Stehr, H., Rausch, T., et al. (2013). Integrative genomic analyses reveal an androgen-driven somatic alteration landscape in early-onset prostate cancer. Cancer Cell *23*, 159–170.

Welti, J., Rodrigues, D.N., Sharp, A., Sun, S., Lorente, D., Riisnaes, R., Figueiredo, I., Zafeiriou, Z., Rescigno, P., de Bono, J.S., et al. (2016). Analytical Validation and Clinical Qualification of a New Immunohistochemical Assay for Androgen Receptor Splice Variant-7 Protein Expression in Metastatic Castration-resistant Prostate Cancer. Eur. Urol. *70*, 599–608.

Wen, X., Li, X., Liao, B., Liu, Y., Wu, J., Yuan, X., Ouyang, B., Sun, Q., and Gao, X. (2009). Knockdown of p21-activated kinase 6 inhibits prostate cancer growth and enhances chemosensitivity to docetaxel. Urology *73*, 1407–1411.

Wen, Y., Hu, M.C., Makino, K., Spohn, B., Bartholomeusz, G., Yan, D.H., and Hung, M.C. (2000). HER-2/neu promotes androgen-independent survival and growth of prostate cancer cells through the Akt pathway.

Cancer Res. *60*, 6841–6845.

Williams, J.L., Greer, P.A., and Squire, J.A. (2014). Recurrent copy number alterations in prostate cancer: an in silico meta-analysis of publicly available genomic data. Cancer Genet *207*, 474–488.

Wu, Y.-M., Cieślik, M., Lonigro, R.J., Vats, P., Reimers, M.A., Cao, X., Ning, Y., Wang, L., Kunju, L.P., de Sarkar, N., et al. (2018). Inactivation of CDK12 delineates a distinct immunogenic class of advanced prostate cancer. Cell *173*, 1770–1782.e14.

Xu, C. (2018). A review of somatic single nucleotide variant calling algorithms for next-generation sequencing data. Comput Struct Biotechnol J *16*, 15–24.

Xu, J., and Qiu, Y. (2019). Current opinion and mechanistic interpretation of combination therapy for castration-resistant prostate cancer. Asian J Androl *21*, 270–278.

Yang, Y.A., and Yu, J. (2015). Current perspectives on FOXA1 regulation of androgen receptor signaling and prostate cancer. Genes Dis. *2*, 144–151.

Yang, L., Xie, S., Jamaluddin, M.S., Altuwaijri, S., Ni, J., Kim, E., Chen, Y.-T., Hu, Y.-C., Wang, L., Chuang, K.-H., et al. (2005). Induction of androgen receptor expression by phosphatidylinositol 3-kinase/Akt downstream substrate, FOXO3a, and their roles in apoptosis of LNCaP prostate cancer cells. J. Biol. Chem. *280*, 33558–33565.

Ying, L., Chunxia, Y., and Wei, L. (2015). Inhibition of ovarian cancer cell growth by a novel TAK1 inhibitor LYTAK1. Cancer Chemother. Pharmacol. *76*, 641–650.

Yoon, H.-G., and Wong, J. (2006). The corepressors silencing mediator of retinoid and thyroid hormone receptor and nuclear receptor corepressor are involved in agonist- and antagonist-regulated transcription by androgen receptor. Mol. Endocrinol. *20*, 1048–1060.

Zare, F., Dow, M., Monteleone, N., Hosny, A., and Nabavi, S. (2017). An evaluation of copy number variation detection tools for cancer using whole exome sequencing data. BMC Bioinformatics *18*, 286.

Zarrei, M., MacDonald, J.R., Merico, D., and Scherer, S.W. (2015). A copy number variation map of the human genome. Nat. Rev. Genet. *16*, 172–

183.

Zeng, H., Jorapur, A., Shain, A.H., Lang, U.E., Torres, R., Zhang, Y., McNeal, A.S., Botton, T., Lin, J., Donne, M., et al. (2018). Bi-allelic Loss of CDKN2A Initiates Melanoma Invasion via BRN2 Activation. Cancer Cell *34*, 56–68.e9.

Zhang, J., and Manley, J.L. (2013). Misregulation of pre-mRNA alternative splicing in cancer. Cancer Discov. *3*, 1228–1237.

Zhao, X.Y., Malloy, P.J., Krishnan, A.V., Swami, S., Navone, N.M., Peehl, D.M., and Feldman, D. (2000). Glucocorticoids can promote androgen-independent growth of prostate cancer cells through a mutated androgen receptor. Nat. Med. *6*, 703–706.

10 Appendices

10.1 Supplemental Figures

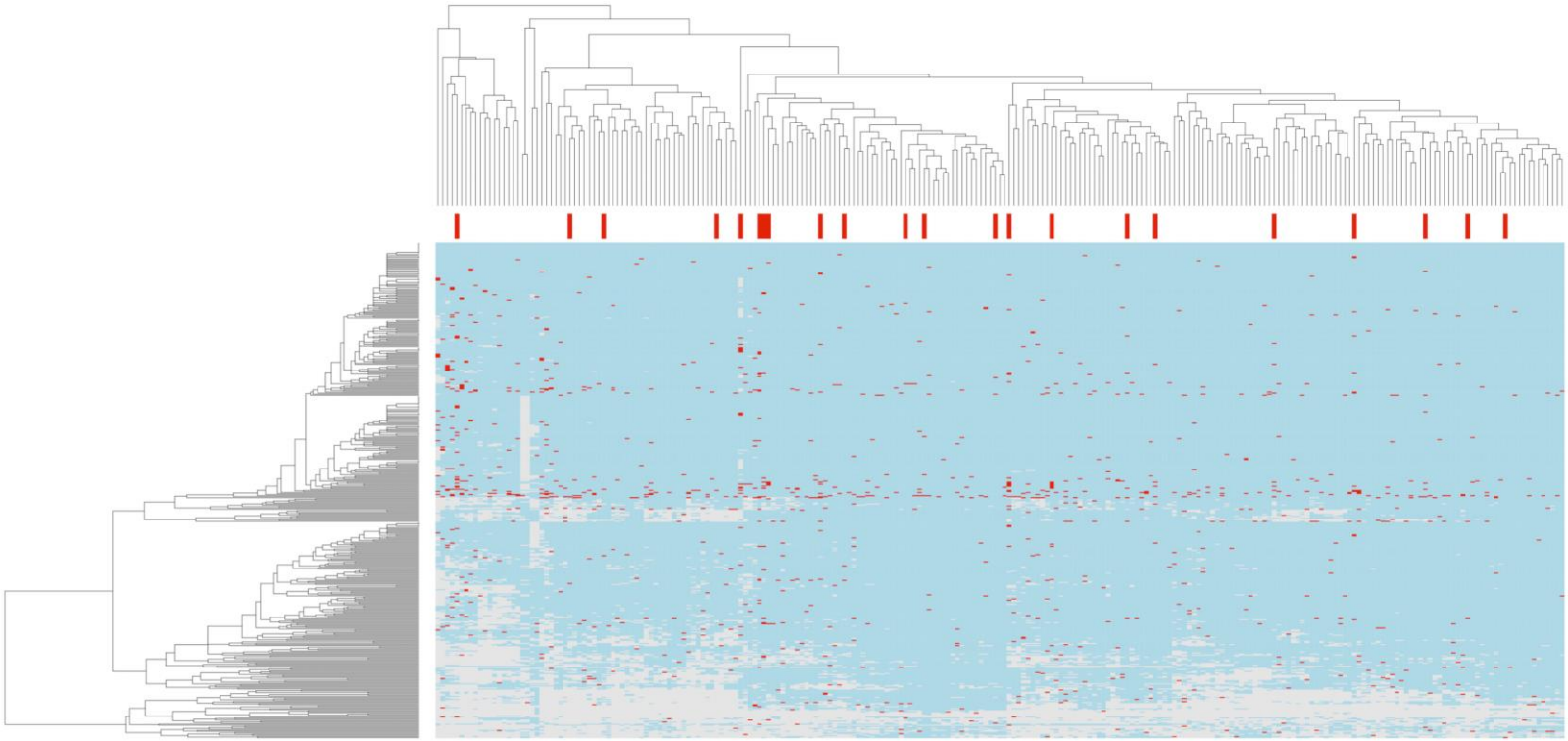


Figure 10-1 MSI loci identified by targeted panel NGS

Unsupervised hierarchical clustering of MSI loci (rows) shows no overall patterns. Cells are coloured by MSI status (blue: normal, red: unstable). Samples (columns) are highlighted if they are flagged as MSI-unstable through an alternative method.

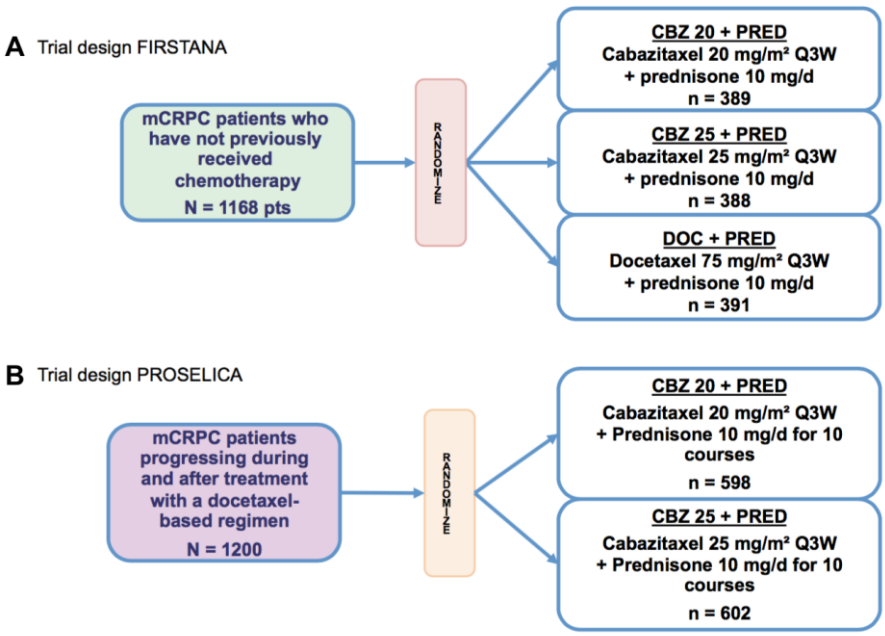


Figure 10-2 FIRSTANA and PROSELICA trial designs

Trial designs of FIRSTANA (A) and PROSELICA (B), both trials of taxane chemotherapy in advanced prostate tumours. Individuals on FIRSTANA have no prior taxane exposure, while those on PROSELICA have been pre-treated with Docetaxel.

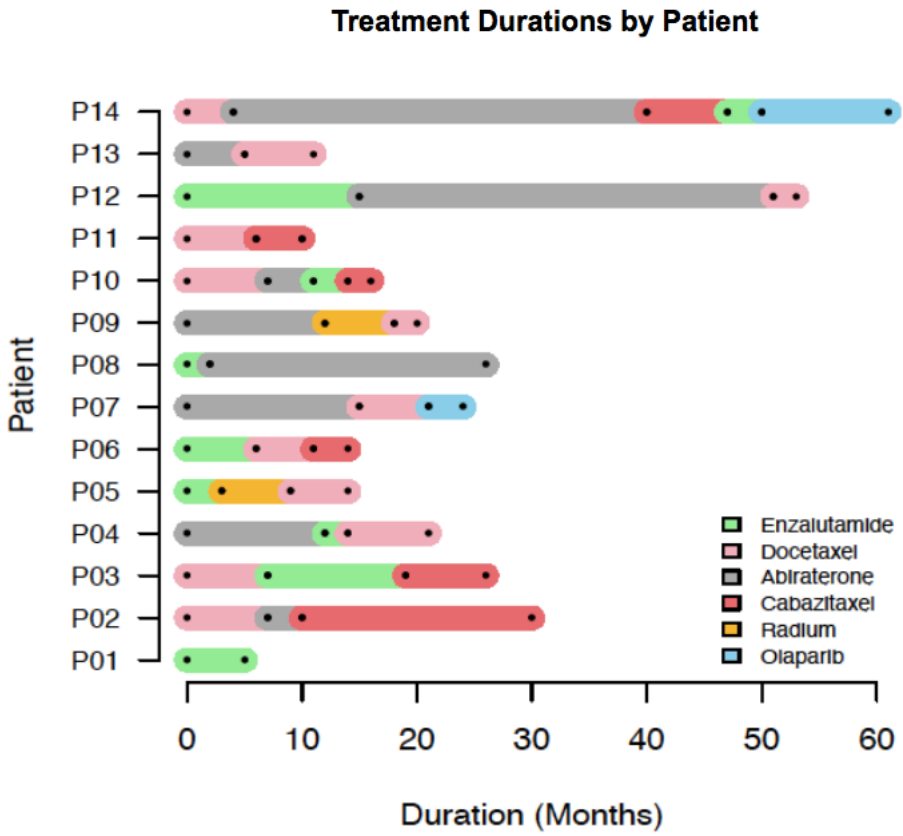


Figure 10-3 Prior treatments of patients included in CTC-CNA validation study

Summarising the range of prior therapies of each individual who participated in the whole-blood apheresis sampling method for CTC collection. Each horizontal bar indicates an individual and the coloured bands mark treatment types.

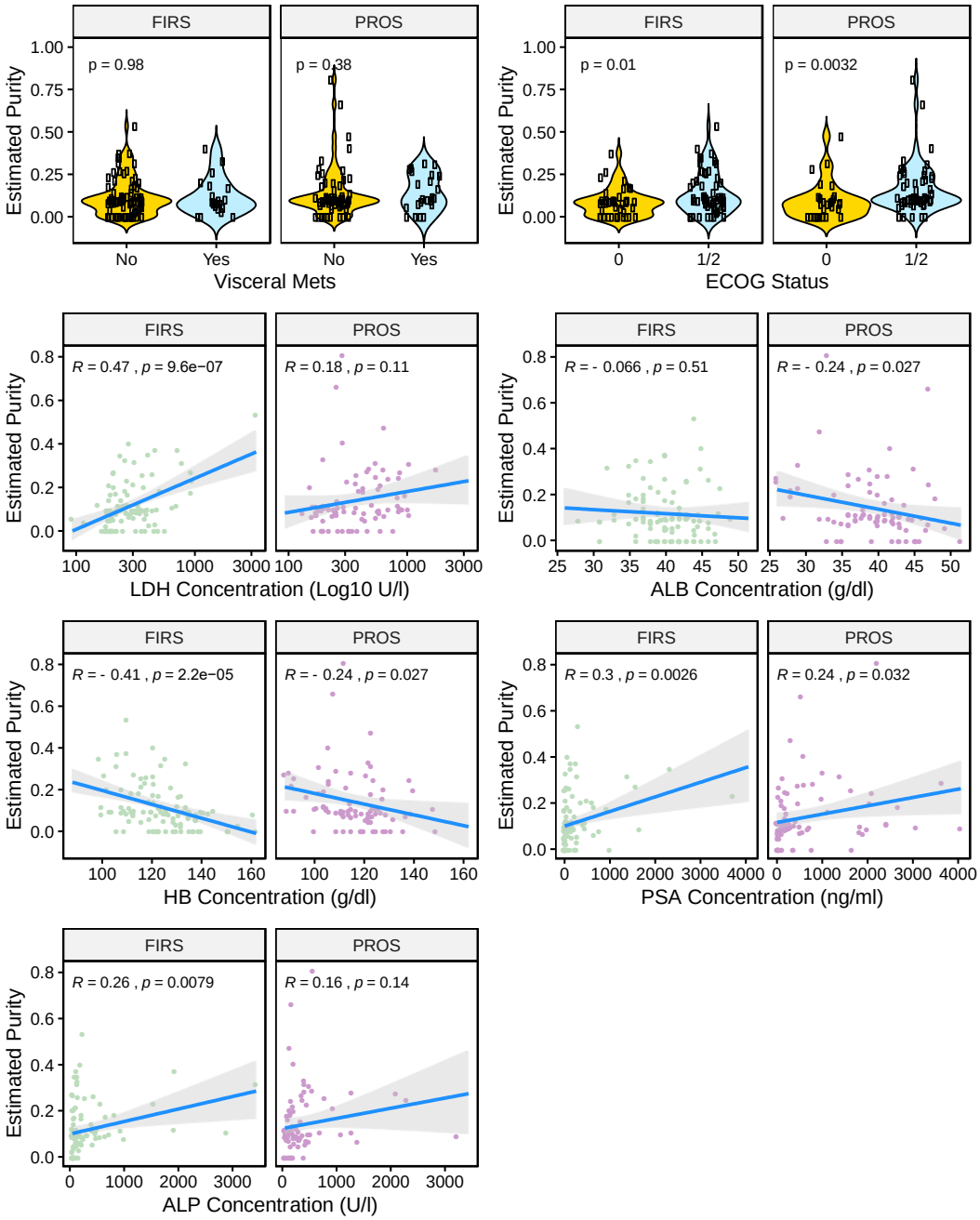


Figure 10-4 Comparing baseline cfDNA tumour purity against clinical factors

Plots showing association of tumour purity values with CRPC clinical variables: the presence of absence of visceral metastases, ECOG performance status, lactate dehydrogenase (LDH), serum albumin (ALB), haemoglobin (HB), prostate specific antigen (PSA), alkaline phosphatase (ALP). Plots and statistical tests split by study (FIRSTANA or PROSELICA). Categorical variables (mets/ECOG) compared using Wilcoxon rank sum test, continuous variables compared using pearson correlation.

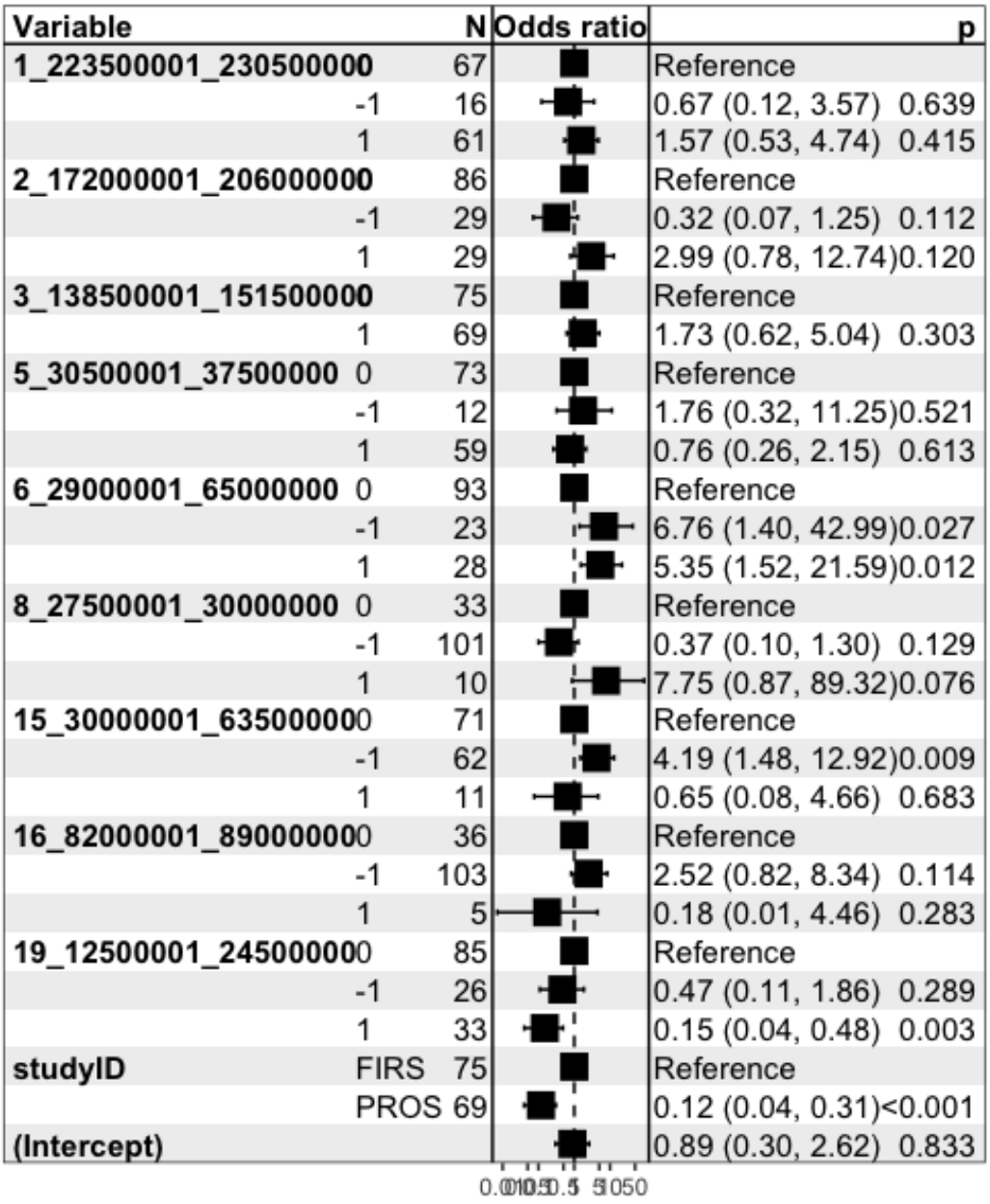


Figure 10-5 Forest model for unsupervised CNA associations with drug response

Logistic regression model for nine genomic regions previously highlighted in unsupervised analysis of taxane-associated genomic regions. Odds ratios, confidence intervals and p-values shown. Membership of the FIRSTANA and PROSELICA cohorts is also included in the model.

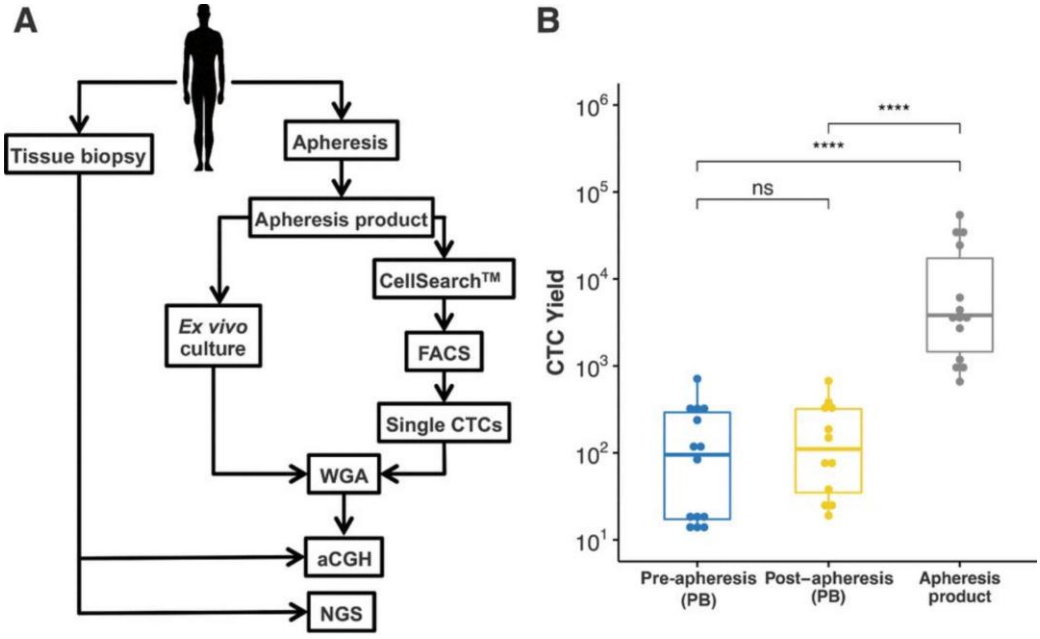


Figure 10-6 Apheresis sample collection workflow and metrics

Whole-blood apheresis allows for enriched numbers of viable single cells for downstream analysis. The sample collection workflow of the apheresis study (A) resulted in sufficient output for a variety of analyses. The overall yield of CTCs (B) is much improved compared to traditional methods.

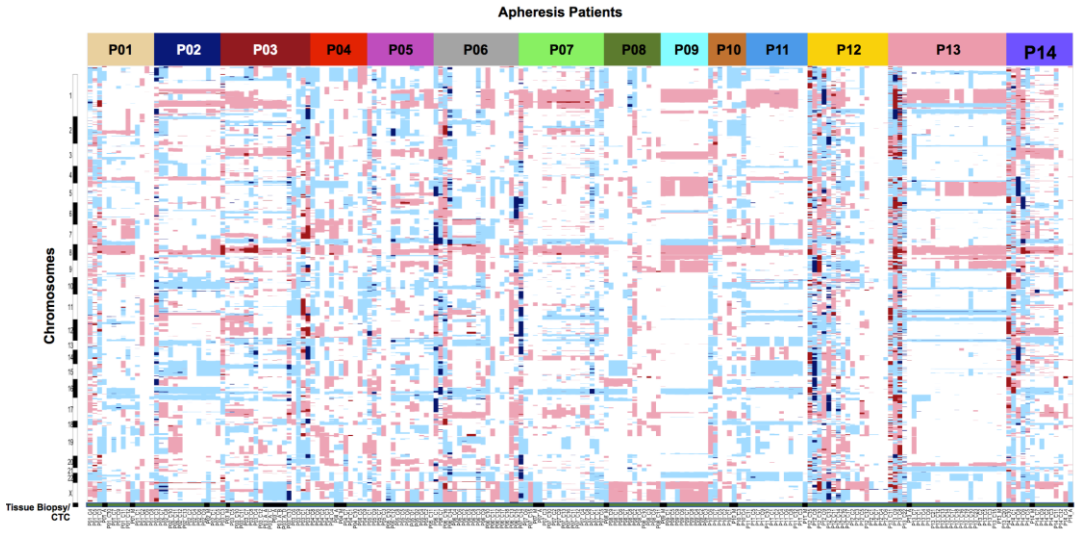


Figure 10-7 Heatmap of genome-wide CNAs across the CTC-CNA cohort

Heatmaps presenting unsupervised hierarchical clustering based on CNA and Euclidean distance, of all the samples for each patient. Column from left to right depicts patients, and rows are ordered by chromosome position. Black bars identify tumour biopsies, grey bars identify CTCs, at the bottom of the heatmap.

10.2 Supplemental Tables

Chrom	Start	End	Gene	Bins
chr1	2488044	2494762	TNFRSF14	9
chr1	11166624	11319601	MTOR	57
chr1	27023543	27107293	ARID1A	22
chr1	45794948	45805950	MUTYH	9
chr1	65300216	65352007	JAK1	24
chr1	115251107	115258811	NRAS	4
chr1	118165465	118166700	FAM46C	1
chr1	120457873	120612065	NOTCH2	36
chr1	156785597	156851536	NTRK1	24
chr2	29416015	30143142	ALK	28
chr2	39477721	39605325	MAP4K3	32
chr2	47600571	47613838	EPCAM	8
chr2	47630332	47710179	MSH2	16
chr2	48018028	48034034	MSH6	9
chr2	58386846	58468473	FANCL	13
chr2	128015124	128051392	ERCC3	14
chr2	212248277	213403289	ERBB4	28
chr2	215593373	215661923	BARD1	10
chr3	10070299	10143006	FANCD2	42
chr3	10183416	10191716	VHL	6
chr3	14187336	14220171	XPC	16
chr3	37035016	37092170	MLH1	19
chr3	38180100	38182862	MYD88	5
chr3	41265457	41280910	CTNNB1	10
chr3	142168225	142297653	ATR	47
chr3	178916573	178952180	PIK3CA	17
chr4	1800897	1809071	FGFR3	18
chr4	55124897	55161535	PDGFRA	23
chr5	56116542	56189532	MAP3K1	19
chr5	67522481	67593505	PIK3R1	18
chr5	79950793	80171748	MSH3	26
chr5	112043349	112179927	APC	16
chr5	131892967	131978081	RAD50	25
chr6	35420573	35434153	FANCE	10
chr6	138192314	138202512	TNFAIP3	8
chr7	6026415	6048684	PMS2	11
chr7	55209957	55273347	EGFR	30
chr7	106507921	106545907	PIK3CG	12
chr7	116335776	116436202	MET	20
chr7	140434376	140550097	BRAF	17
chr7	148504716	148547656	EZH2	20
chr8	30915930	31030684	WRN	34
chr8	48686686	48872585	PRKDC	84
chr8	90947766	90995199	NBN	15
chr8	128748815	128753281	MYC	5
chr8	145736777	145742863	RECQL4	19
chr9	5021949	5126830	JAK2	23
chr9	21968185	21994411	CDKN2A	7
chr9	35074040	35079557	FANCG	14
chr9	97863903	98011615	FANCC	15
chr9	100437667	100456095	XPA	5
chr9	135771597	135804355	TSC1	20
chr9	139390613	139438574	NOTCH1	54
chr10	43595874	43623817	RET	21
chr10	50666825	50741101	ERCC6	20
chr10	89624161	89725308	PTEN	10
chr10	123239046	123353351	FGFR2	20
chr11	532652	534317	HRAS	5
chr11	22646176	22647416	FANCF	3
chr11	32410508	32456399	WT1	11
chr11	47236647	47260486	DDB2	9
chr11	94153206	94226024	MRE11A	20
chr11	108098341	108236301	ATM	60
chr11	125496642	125525285	CHEK1	11
chr12	12870774	12871995	CDKN1B	2
chr12	25362711	25398361	KRAS	5
chr12	46123588	46298889	ARID2	19
chr12	56474034	56495900	ERBB3	25
chr12	58142278	58145569	CDK4	7
chr12	69202894	69233681	MDM2	10
chr12	121416694	121439019	HNF1A	11
chr13	32890544	32972943	BRCA2	25
chr13	48881353	49054295	RB1	25
chr13	103498580	103528287	ERCC5	15
chr14	35871128	35873123	NFKBIA	4
chr14	45605206	45669259	FANCM	22
chr14	68290194	69149691	RAD51B	15
chr14	75483756	75516409	MLH3	12
chr14	104165077	104177508	XRCC3	8
chr14	105236653	105259078	AKT1	15
chr15	40453354	40513013	BUB1B	24
chr15	66679673	66783018	MAP2K1	12
chr15	89790837	89859765	FANCI	34
chr15	91290584	91358577	BLM	22
chr16	338053	397131	AXIN1	16
chr16	2097726	2138715	TSC2	44
chr16	14013974	14042252	ERCC4	11
chr16	23614752	23652489	PALB2	13
chr16	68772146	68867433	CDH1	16
chr16	89805016	89882428	FANCA	41
chr17	7572846	7579964	TP53	7
chr17	11958147	12044610	MAP2K4	10
chr17	29482952	29701201	NF1	62
chr17	33427876	33446720	RAD51D	12
chr17	37618334	37687625	CDK12	14
chr17	37855748	37884425	ERBB2	26
chr17	41197616	41276226	BRCA1	23
chr17	47677717	47700284	SPOP	9
chr17	56769935	56811677	RAD51C	9
chr17	63526045	63554838	AXIN2	13
chr19	1206853	1226541	STK11	9
chr19	4090533	4117657	MAP2K2	12
chr19	11095034	11172529	SMARCA4	41
chr19	40740892	40771261	AKT2	13
chr19	45854812	45873568	ERCC2	20
chr20	36014396	36031804	SRC	11
chr22	24133868	24176430	SMARCB1	8
chr22	29083879	29130783	CHEK2	14
chr22	29999919	30090862	NF2	17
chrX	14861624	14883724	FANCB	7
chrX	15808569	15841424	ZRSR2	11
chrX	66764935	66943789	AR	13
chrX	76763771	77041588	ATRX	36

Table 10-1 Genes and genomic ranges in the DNA-repair defect targeted panel

		Paired Data New CNAs						Paired Comparisons (McNemar's Test) P-Values					
		Overall (n=72)		FIRS (n=29)		PROS (n=43)		Overall BL vs Overall EOS		FIRS BL vs FIRS EOS		PROS BL vs PROS EOS	
		Gains	Losses	Gains	Losses	Gains	Losses	Gains	Losses	Gains	Losses	Gains	Losses
1_22500001_23000000	E2F2	4	6	2	2	2	4	1	0.453	1	1	1	0.546
2_135500001_136000000	CXCR4	3	6	2	3	1	3	1	0.752	1	1	1	1
3_70000001_70500000	FOXP1	7	6	3	3	4	3	0.803	1	1	1	1	1
3_71000001_71500000	RYBP	7	6	3	3	4	3	0.803	1	1	1	1	1
3_71500001_72000000	SHQ1	7	7	3	3	4	4	0.803	0.773	1	1	1	1
3_137000001_137500000	PIK3CB	6	1	3	0	3	1	0.789	0.371	0.617	0.48	0.343	1
3_177500001_178000000	PIK3CA	5	1	3	1	2	0	0.423	0.221	1	1	0.182	0.134
3_184500001_185000000	ETV5	6	1	4	1	2	0	0.606	0.371	0.683	1	0.182	0.248
5_66500001_67000000	PIK3R1	5	9	2	3	3	6	0.221	1	1	1	0.248	1
5_97000001_97500000	CHD1	5	7	2	3	3	4	0.45	0.628	1	1	0.248	0.752
5_111000001_111500000	APC	6	9	4	3	2	6	1	1	1	1	0.48	1
6_19000001_19500000	E2F3	6	7	4	4	2	3	1	1	0.683	0.371	0.45	0.724
6_90000001_90500000	MAP3K7	5	7	4	2	1	5	0.45	0.359	0.683	1	1	0.302
7_12500001_13000000	ETV1	8	2	4	1	4	1	1	1	1	1	1	1
7_54000001_54500000	EGFR	8	2	3	0	5	2	1	1	1	0.48	1	1
7_91000001_91500000	CDK6	7	1	3	0	4	1	1	1	1	NA	1	1
7_98500001_99000000	MCM7	6	0	4	0	2	0	0.789	NA	1	NA	0.683	NA
7_139000001_139500000	BRAF	7	0	4	0	3	0	1	0.48	1	NA	1	0.48
7_149500001_150000000	CDK5	11	1	4	0	7	1	0.332	0.371	0.683	0.48	0.546	1
8_22500001_23000000	NKX3-1	2	2	1	0	1	2	0.48	0.114	1	0.074	1	1
8_85000001_85500000	E2F5	4	0	3	0	1	0	0.752	NA	1	NA	0.617	NA
8_127500001_128000000	MYC	6	0	4	0	2	0	0.789	NA	1	NA	0.45	NA
9_20500001_21000000	CDKN2A	2	7	0	5	2	2	0.182	0.546	0.134	0.221	1	1
10_61500001_62000000	CDK1	4	3	2	1	2	2	1	0.149	1	0.221	0.683	0.683
10_88500001_89000000	PTEN	0	4	0	3	0	1	1	0.181	NA	0.343	1	0.617
12_11500001_12000000	CDKN1B	8	3	2	1	6	2	0.789	1	1	1	0.505	1
12_55000001_55500000	CDK2	10	4	2	2	8	2	0.181	0.683	1	1	0.114	1
12_57000001_57500000	CDK4	13	3	4	1	9	2	0.052	1	1	1	0.027	1
13_31500001_32000000	BRCA2	3	2	1	1	2	1	0.617	0.114	1	0.617	1	0.221
13_47500001_48000000	RB1	2	2	1	1	1	1	1	0.114	1	0.617	1	0.221
14_104000001_104500000	AKT1	6	5	4	3	2	2	0.606	0.724	1	0.617	0.45	1
16_66000001_66500000	E2F4	3	8	0	6	3	2	0.343	1	1	1	0.505	0.683
16_67500001_68000000	CDH1	3	8	0	6	3	2	0.505	1	1	1	0.724	0.683
17_6500001_7000000	TP53	2	8	2	2	0	6	1	1	0.48	1	0.48	1
17_40000001_40500000	BRCA1	9	6	4	3	5	3	0.803	0.752	1	1	1	1
17_40500001_41000000	ETV4	9	6	4	3	5	3	0.803	0.752	1	1	1	1
17_72500001_73000000	CDK3	8	5	3	3	5	2	0.579	1	1	1	0.45	1
18_47000001_47500000	SMAD4	2	5	1	4	1	1	1	0.773	1	1	1	0.617
18_59500001_60000000	BCL2	1	5	1	1	0	4	1	0.579	1	0.371	1	1
20_31000001_31500000	E2F1	11	2	6	1	5	1	0.838	1	1	1	0.579	1
X_65500001_66000000	AR	10	5	4	4	6	1	0.302	0.423	1	1	0.289	0.371

Table 10-2 Matched pair analysis of selected gene CNA changes

		Copy-Losses				Copy-Gains				Frequency Comparison (Fisher's Test) P-Values					
		Baseline		End of Study		Baseline		End of Study		FIRS BL vs PROS BL		FIRS BL vs FIRS EOS		PROS BL vs PROS EOS	
		FIRS	PROS	FIRS	PROS	FIRS	PROS	FIRS	PROS	Losses	Gains	Losses	Gains	Losses	Gains
1_235000001_235500000	TBCE	0.16	0.09	0.15	0.08	0.35	0.53	0.24	0.48	0.21	0.04	1.00	0.37	1	0.706
2_131500001_132000000	TUBA3D_TUBA3E	0.32	0.32	0.39	0.31	0.11	0.12	0.12	0.12	1.00	1.00	0.51	1.00	1	1
2_219500001_220000000	TUBA4A	0.17	0.19	0.12	0.27	0.24	0.23	0.39	0.16	0.83	1.00	0.58	0.11	0.371	0.487
5_76000001_76500000	TBCA	0.33	0.41	0.42	0.45	0.17	0.07	0.15	0.12	0.39	0.08	0.39	1.00	0.707	0.522
6_2500001_3000000	TUBB2A_TUBB2B	0.28	0.33	0.33	0.33	0.16	0.17	0.12	0.12	0.59	0.83	0.65	0.77	1	0.605
6_30000001_30500000	TUBB	0.11	0.16	0.24	0.18	0.23	0.2	0.27	0.18	0.46	0.84	0.08	0.63	0.805	1
6_42000001_42500000	TBCC	0.12	0.17	0.24	0.18	0.24	0.17	0.27	0.18	0.48	0.41	0.15	0.81	1	1
6_111500001_112000000	TUBE1	0.33	0.48	0.27	0.39	0.12	0.04	0.24	0.08	0.09	0.13	0.65	0.15	0.353	0.447
9_139500001_140000000	TUBB4B	0.11	0.14	0.21	0.24	0.33	0.33	0.18	0.29	0.62	1.00	0.23	0.17	0.23	0.688
10_500001_1000000	TUBB8	0.39	0.41	0.27	0.47	0.07	0.14	0.18	0.08	0.87	0.17	0.28	0.09	0.573	0.391
10_4500001_5000000	TUBAL3	0.41	0.42	0.3	0.47	0.07	0.14	0.15	0.1	1.00	0.17	0.39	0.17	0.707	0.582
12_49000001_49500000	TUBA1A_TUBA1B_TUBA1C	0.07	0.14	0.12	0.2	0.25	0.2	0.3	0.33	0.17	0.55	0.45	0.64	0.459	0.139
13_19500001_20000000	TUBA3C	0.33	0.33	0.45	0.35	0.16	0.22	0.12	0.18	1.00	0.40	0.28	0.77	1	0.817
16_89000001_89500000	TUBB3	0.68	0.72	0.55	0.57	0.04	0.03	0	0.06	0.59	1.00	0.20	0.55	0.114	0.648
17_40000001_40500000	TUBG1_TUBG2	0.25	0.13	0.33	0.2	0.15	0.26	0.21	0.22	0.09	0.10	0.49	0.41	0.317	0.672
17_57000001_57500000	TUBD1	0.21	0.14	0.33	0.16	0.24	0.35	0.24	0.43	0.39	0.20	0.23	1.00	0.8	0.443
17_80000001_80500000	TBCD	0.13	0.12	0.24	0.12	0.35	0.42	0.3	0.49	0.81	0.39	0.17	0.83	1	0.573
18_11500001_12000000	TUBB6	0.16	0.23	0.12	0.2	0.32	0.35	0.39	0.22	0.30	0.73	0.77	0.51	0.823	0.16
19_5500001_6000000	TUBB4A	0.25	0.32	0.24	0.43	0.2	0.12	0.18	0.06	0.46	0.18	1.00	1.00	0.248	0.358
19_36000001_36500000	TBCB	0.15	0.2	0.21	0.18	0.25	0.22	0.27	0.29	0.39	0.70	0.41	0.82	1	0.516
20_57000001_57500000	TUBB1	0.08	0.01	0.06	0.04	0.39	0.58	0.48	0.55	0.12	0.03	1.00	0.40	0.569	0.851
22_18000001_18500000	TUBA8	0.45	0.45	0.48	0.43	0.08	0.07	0.09	0.12	1.00	1.00	0.84	1.00	0.853	0.522

Table 10-3 Frequency data and statistical tests of tubulin gene set

		Paired Data New CNAs						Paired Comparisons (McNemar's Test) P-Values					
		Overall (n=72)		FIRS (n=29)		PROS (n=43)		Overall BL vs Overall EOS		FIRS BL vs FIRS EOS		PROS BL vs PROS EOS	
		Gains	Losses	Gains	Losses	Gains	Losses	Gains	Losses	Gains	Losses	Gains	Losses
1_235000001_235500000	TBCE	7	2	3	1	4	1	0.48	1	0.724	1	0.752	1
2_131500001_132000000	TUBA3D_TUBA3E	3	5	2	3	1	2	1	1	1	1	1	1
2_219500001_220000000	TUBA4A	8	4	7	1	1	3	0.789	1	0.182	0.617	0.371	0.617
5_76000001_76500000	TBCA	7	9	3	3	4	6	0.343	1	1	1	0.371	1
6_2500001_3000000	TUBB2A_TUBB2B	5	6	3	3	2	3	0.579	0.789	1	1	0.45	0.505
6_30000001_30500000	TUBB	9	9	6	4	3	5	1	0.07	0.505	0.134	0.505	0.45
6_42000001_42500000	TBCC	9	9	6	4	3	5	0.606	0.149	0.505	0.371	1	0.45
6_111500001_112000000	TUBE1	6	7	4	3	2	4	0.289	0.803	0.371	1	1	0.546
9_139500001_140000000	TUBB4B	7	8	3	3	4	5	1	1	1	1	1	1
10_500001_1000000	TUBB8	4	6	4	1	0	5	1	0.606	0.134	0.131	0.248	0.724
10_4500001_5000000	TUBAL3	3	5	3	0	0	5	1	0.423	0.248	0.041	0.248	0.724
12_49000001_49500000	TUBA1A_TUBA1B_TUBA1C	12	3	3	1	9	2	0.08	1	1	1	0.027	1
13_19500001_20000000	TUBA3C	6	4	3	1	3	3	0.505	0.546	0.248	0.617	1	1
16_89000001_89500000	TUBB3	2	7	0	4	2	3	1	0.359	1	1	1	0.343
17_40000001_40500000	TUBG1_TUBG2	9	6	4	3	5	3	0.803	0.752	1	1	1	1
17_57000001_57500000	TUBD1	7	6	2	4	5	2	1	0.752	1	0.371	0.724	1
17_80000001_80500000	TBCD	8	5	3	3	5	2	0.579	1	1	1	0.45	0.683
18_11500001_12000000	TUBB6	8	5	6	2	2	3	0.814	0.302	0.752	0.683	0.289	0.505
19_5500001_6000000	TUBB4A	2	6	2	2	0	4	0.289	0.332	1	0.289	0.248	1
19_36000001_36500000	TBCB	7	5	3	3	4	2	0.546	0.302	1	0.724	0.683	0.45
20_57000001_57500000	TUBB1	11	2	7	1	4	1	1	1	0.546	1	0.386	1
22_18000001_18500000	TUBA8	4	7	1	3	3	4	1	0.803	1	1	1	1

Table 10-4 Matched pair analysis of tubulin gene-set CNA changes

Pathology Variables	Overall Total (n=246)	Cohort Subsets HR (n=35)	HR + MMR (n=75)	MMR (n=136)
T				
1	3 (1 %)	1 (3 %)	0 (0 %)	2 (1 %)
2	15 (6 %)	1 (3 %)	7 (9 %)	7 (5 %)
3	91 (37 %)	6 (17 %)	34 (45 %)	51 (38 %)
4	30 (12 %)	2 (6 %)	9 (12 %)	19 (14 %)
Missing	107 (43.5%)	25 (71.4%)	25 (33.3%)	57 (41.9%)
N				
0	54 (22 %)	3 (9 %)	18 (24 %)	33 (24 %)
1	65 (26 %)	6 (17 %)	20 (27 %)	39 (29 %)
2	6 (2 %)	0 (0 %)	2 (3 %)	4 (3 %)
Missing	121 (49.2%)	26 (74.3%)	35 (46.7%)	60 (44.1%)
M				
0	80 (33 %)	5 (14 %)	29 (39 %)	46 (34 %)
1	118 (48 %)	5 (14 %)	36 (48 %)	77 (57 %)
Missing	48 (19.5%)	25 (71.4%)	10 (13.3%)	13 (9.6%)
Gleason				
5	1 (0 %)	0 (0 %)	1 (1 %)	0 (0 %)
6	13 (5 %)	1 (3 %)	5 (7 %)	7 (5 %)
7	30 (12 %)	3 (9 %)	10 (13 %)	17 (12 %)
8	32 (13 %)	6 (17 %)	7 (9 %)	19 (14 %)
9	101 (41 %)	1 (3 %)	35 (47 %)	65 (48 %)
10	11 (4 %)	0 (0 %)	5 (7 %)	6 (4 %)
Missing	58 (23.6%)	24 (68.6%)	12 (16.0%)	22 (16.2%)

Other Variables	Overall Total (n=246)	Cohort Subsets HR (n=35)	HR + MMR (n=75)	MMR (n=136)
PSA (ng/ml)				
Median (IQR)	75 (± 78)	83 (± 47)	58 (± 89)	80 (± 75)
Missing	52 (21.1%)	28 (80.0%)	10 (13.3%)	14 (10.3%)
Prostatectomy				
N	190 (77 %)	9 (26 %)	68 (91 %)	113 (83 %)
Y	27 (11 %)	3 (9 %)	7 (9 %)	17 (12 %)
Missing	29 (11.8%)	23 (65.7%)	0 (0%)	6 (4.4%)
Radiotherapy				
N	142 (58 %)	6 (17 %)	48 (64 %)	88 (65 %)
Y	75 (30 %)	6 (17 %)	27 (36 %)	42 (31 %)
Missing	29 (11.8%)	23 (65.7%)	0 (0%)	6 (4.4%)
Age (years)				
Median (IQR)	62 (± 9.0)	59 (± 8.0)	62 (± 8.9)	62 (± 8.3)
Missing	25 (10.2%)	23 (65.7%)	0 (0%)	2 (1.5%)

Υνσαµπλαβλε/νον-εϖαλυαβλε δαταλιστεδ ασΜισσινγ

Table 10-5 Patient Clinical Characteristics for Targeted Sequencing Dataset

Characteristic	FIRSTANA n =103 N (%)	PROSELICA n = 85 N (%)	*p* value[a]
ECOG PS[b]			
0-1	100 (97)	75 (88)	0.02
2	3 (3)	10 (12)	
RECIST measurable[b]			
No	49 (48)	40 (47)	0.944
Yes	54 (52)	45 (53)	
Visceral disease			
No	83 (81)	60	0.11
Yes	20 (19)	25	
Pain at baseline[c]			
No	19 (18)	21 (25)	0.48
Yes	69 (67)	59 (69)	
Gleason score at diagnosis[d]			
< 8	42 (41)	45 (53)	0.051
≥ 8	56 (54)	33 (39)	
Trial arm			
Cabazitaxel (20mg/m^2)	34 (33)	41 (48)	
Cabazitaxel (25mg/m^2)	28 (27)	44 (52)	<0.001
Docetaxel (75mg/m^2)	41 (40)	0 (0)	
	Median (IQR)	**Median (IQR)**	***p* value[e]**
Age (yr)	68 (62.5 - 72)	67 (64 - 71)	0.254
LDH (U/l)	266.95 (204.25 - 359.5)	365.5 (234 - 605)	0.003
ALP (U/l)	129.39 (81 - 241.49)	214 (118 - 413)	0.002
Haemoglobin (g/dl)	121 (110.69 - 128.24)	112 (105 - 122.33)	<0.001
Albumin (g/dl)	40.2 (37.5 - 43.35)	40 (36 - 43)	0.23
PSA (ng/ml)	60.45 (18.16 - 187.67)	160.6 (64.92 - 622.75)	<0.001
PSA doubling time (mo)	2.04 (1.19 - 3.3)	1.7 (1.17 - 2.82)	0.374
NLR	2.25 (1.53 - 4.1)	2.79 (1.69 - 3.89)	0.166
Outcome	**N (%)**	**N (%)**	***p* value**
>50% PSA response at 12 weeks			
No	48 (47)	62 (73)	
Yes	55 (53)	23 (27)	<0.001
>50% PSA response at any time			
No	35 (34)	52 (61)	
Yes	68 (66)	33 (39)	<0.001
	Median (IQR)	**Median (IQR)**	**HR (95% CI) p value[f]**
rPFS (mo)	6.97 (4.58 - 11.37)	4.37 (2.83 - 8.08)	1.83 (1.22-2.74) 0.003
OS (mo)	21.26 (11.33 - 30.08)	13.27 (7.39 - 19.42)	2.36 (1.68-3.30) < 0.001

ALP = alkaline phosphatase; ECOG PS = Eastern Cooperative Oncology Group performance status; IQR = interquartile range; LDH = lactate dehydrogenase; mo = months; PSA = prostate-specific antigen; RECIST = Response Evaluation Criteria in Solid Tumours; U = unit, yr = years

a χ^2 test.

b Stratification parameters

c Twenty assessments missing (15 in FIRSTANA and 5 in PROSELICA)

d Twelve assessments missing (5 in FIRSTANA and 7 in PROSELICA)

e Wilcoxon rank-sum test

f Proportional Hazards Cox model, FIRSTANA cohort as reference

Table 10-6 Patient Clinical Characteristics for FIRSTANA and PROSELICA (cfDNA) Studies

	FIRSTANA			PROSELICA		
Characteristic	Biomarker subset n = 105 N (%)	Non-biomarker subset n = 1063 N (%)	p-value[a]	Biomarker subset n = 85 N (%)	Non-biomarker subset n = 1115 N (%)	p-value[a]
ECOG PS[b]						
0–1	102 (7)	1018 (96)	0.5	75 (88)	1005 (90)	0.6
2	3 (3)	45 (4)		10 (1)	1110 (10)	
RECIST measurable[b]						
No	51 (49)	502 (47)	0.79	40 (47)	571 (51)	0.46
Yes	54 (51)	561 (53)		45 (53)	544 (49)	
Visceral disease						
No	85 (81)	833 (78)	0.54	60 (71)	805 (72)	0.75
Yes	20 (19)	230 (22)		25 (29)	310 (28)	
Pain at baseline[c]						
No	19 (18)	355 (33)	0.04	21 (25)	289 (26)	0.7
Yes	71 (68)	624 (59)		59 (69)	734 (66)	
	Median (IQR)	Median (IQR)	p-value[d]	Median (IQR)	Median (IQR)	p-value[d]
Age (yr)	68 (62 - 72)	68 (63 - 74)	0.14	67 (64-71)	69 (63 - 74)	0.01
LDH (U/L)	263 (207 - 368)	239 (190 - 374)	0.12	350 (222 - 588)	325 (220 - 498)	0.32
ALP (U/L)	128.5 (80.5 - 240.8)	125 (79 - 264)	0.35	209 (118 - 415)	163 (92 - 346)	0.02
Haemoglobin (g/dl)	123 (114 - 132.1)	128 (117 - 137)	0.01	116.8 (107.8 - 124)	120 (108 - 130)	0.02
Albumin (g/dl)	40.1 (37.3 - 43)	41 (38 - 44)	0.05	40 (36 - 43)	40 (36.7 - 43)	0.4
PSA (ng/ml)	77.9 (22.4 - 236.5)	76 (29.9 - 196.4)	0.43	247.1 (93.7 - 740.8)	158.3 (53.2 - 412.85)	0.001
PSA-doubling time (mo)	2 (1.2 - 3.2)	2 (1.3 - 3.4)	0.14	1.7 (1.2 - 2.8)	1.9 (1.2 - 3.1)	0.32
Outcome	N (%)	N (%)	p-value	N (%)	N (%)	p-value
>50% PSA response at 12 weeks						
No	49 (47)	484 (46)	0.82	62 (73)	765 (69)	0.40
Yes	56 (53)	579 (54)		23 (27)	351 (31)	
>50% PSA response at any time						
No	36 (34)	309 (30)	0.26	52 (61)	660 (59)	0.71
Yes	69 (66)	754 (70)		33 (39)	456 (41)	

ALP = alkaline phosphatase; cfDNA = cell-free DNA; ECOG PS = Eastern Cooperative Oncology Group performance status; IQR = interquartile range; LDH = lactate dehydrogenase; mo = months; PSA = prostate-specific antigen; RECIST = Response Evaluation Criteria in Solid Tumours; U = unit, yr = years.
a = χ^2 test.
b = Stratification parameters
c = For FIRSTANA 99 assessments were missing (15 in the sub-study and 84 in the main study), For PROSELICA 97 assessments were missing (5 in the sub-study and 92 in the main study).
d = Wilcoxon rank sum test

Table 10-7 Comparing the cfDNA Study Subset to the Overall Trial Populations

Characteristics		Value
Age (yr)		
	Median	70.4
	Range	60-77
Time from prostate cancer to apheresis procedure (yr)		
	Median	3.9
	Range	2-11.6
No. of metastatic sites		
	Median	3
	Range	1-3
No. Bone metastases (%)		14 (1005)
PSA Level (ng/ml)		
	Median	506
	Range	41-6089
ECOG Performance Status No. (%)		
	0	0
	1	14 (100%)
Received Prior CRPC Treatment No. (100%)		
	1	1 (7%)
	2	4 (29%)
	3	7 (50%)
	≥4	2 (14%)

Table 10-8 Patient Characteristics for CTC-Apheresis Study