Mitomycin C Induced Genetic Alterations and Tumour Evolution in Non-Muscle-Invasive Bladder Cancer

Rebecca Jane Astley

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds

Faculty of Medicine and Health

January 2019

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

I would like to extend an enormous thank you to my primary supervisor, Dr Carolyn Hurst for all the help, advice, support and patience throughout this project and write up. I could not have asked for a nicer, more thoughtful supervisor with such an infectious enthusiasm for science. I would also like to thank Professor Margaret Knowles for her guidance and unparalleled expertise in the field of bladder cancer.

I would also like to thank Dr Guo Cheng for her bioinformatics expertise and for performing the initial variant calling and the rest of Lab 5 for their help and support.

A big thank you goes to my friends and fellow PhD students Abi Bloy, Aarren Mannion and Rebecca Ellerington, for the moral support throughout the years. Having you guys with me for the journey helped in so many ways. I would like to thank my family for their endless support and encouragement throughout. And a special thank you goes to Aaron McDonagh for keeping me fed and sane throughout this writeup.

Finally, I would like to thank Cancer Research UK for funding this project and my studentship.

Thank you all of you, I couldn't have done it without you.

Abstract

Bladder cancer (BC) is the ninth most common cancer in the world. There are two main forms of the disease; non-muscle-invasive BC (NMIBC) and muscle-invasive BC (MIBC). NMIBC commonly recurs and multiple tumours may be resected from the same patient over many years. This provides a unique opportunity to study the molecular events that occur during disease evolution. Some patients receive intravesical courses of mitomycin-C (MMC) chemotherapy, which may provide a potent selective advantage during disease evolution.

67 tumours from 23 patients with recurrent NMIBC were analysed for copy number alterations (CNAs) using shallow-pass whole genome sequencing. Recurrent tumours from each patient tended to share CNAs and common changes seen in BC, such as loss of chromosome 9, suggestive of a monoclonal origin. Some patients received a course of MMC. However no CNAs that specifically related to treatment were identified.

To better understand chemotherapy related events in NMIBC, 18 tumours from 8 patients who received a course of MMC were analysed using whole exome sequencing. A significant increase in non-synonymous mutations was identified post-treatment. Unique mutations post-treatment displayed a significant increase in the number of transversions, specifically C:G>A:T substitutions, as well as an increase in the number of tandem substitutions, specifically at CC or GG dinucleotides. This is consistent with the types of mutations induced by MMC experimentally. Analysis of the sequence context shows that MMC produces a signature similar to that seen by aflatoxin. Analysis of the clonality of the MMC-induced mutations demonstrates that these unique mutations tend to be subclonal.

Overall this work shows that MMC can cause DNA damage that can be identified in post-treatment tumours and this could influence the evolutionary trajectory of the cancer. Further work is required to confirm this mutational signature and fully discern the effect of MMC treatment on the clonal dynamics of NMIBC.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

aCGH	Array comparative genomic hybridisation
ALL	Acute lymphoblastic leukaemia
ANOVA	Analysis of Variance
APOBEC	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
ASCN	Allele-specific copy number
BAC	Bacteria artificial chromosome
BAM	Binary alignment map
BC	Bladder Cancer
BCG	Bacillus Calmette–Guérin
bp	Base pair
CBS	Circular binary segmentation
CCF	Cancer cell fraction
ccRCC	Clear-cell renal cell carcinoma
CGH	Comparative genomic hybridisation
CIS	Carcinoma *in situ*
CLL	Chronic Lymphocytic leukaemia
CN	Copy number
CNA	copy number alterations
COSMIC	Catalogue Of Somatic Mutations In Cancer
cpb	Cycles per burst
ddNTP	dideoxynucleoside triphosphate
DMC	10-decarbamoyl mitomycin C
DNA	Deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DREAM	Dialogue for Reverse Engineering Assessments and Methods
EAU	European Association of Urology
EORTC	European Organisation for Research and Treatment of Cancer
ETS	E-twenty-six
ExoI	Exonuclease I
FFPE	Formalin fixed paraffin embedded
FGA	Fraction genome altered
FGF	Fibroblast growth factor
FISH	Fluorescence *in situ* hybridisation
G1	Grade 1
G2	Grade 2
G3	Grade 3
gDNA	genomic DNA
GWFP	Genome wide frequency plots
GS1	Genomic Subtype 1
GS2	Genomic Subtype 2
GWFP	Genome wide frequency plot
H&E	Haematoxylin and Eosin
HAL	hexylaminolevulinate

HD	Homozygous deletion
HG	High-grade
HMM	Hidden Markov Model
ICGC	International Cancer Genome Consortium
IGV	Interactive genome viewer
ILS	Internal lane standard
IQR	Inter-quartile range
JR	Jo-an Roulson
LCM	Laser Capture Microdissection
LG	Low-grade
LogOR	Log-odds ratio
LogR	Log ratio
LOH	Loss of heterozygosity
M	Molar
MAF	Mutation annotation file
MAPK	Mitogen-activated protein kinase
Mb	Megabase
MIBC	Muscle-invasive bladder cancer
ml	Millilitre
mM	Minimolar
MMC	Mitomycin C
MSKCC	Memorial Sloan Kettering Cancer Center
Mtr	Months to recurrence
NEB	New England Biolabs
NFW	Nuclease free water
ng	Nanogram
NGS	Next Generation Sequencing
NMF	Non-negative matrix factorisation
NMIBC	Non-muscle-invasive bladder cancer
NSCLC	Non-small cell lung cancer
OCT	Optimal Cutting Temperature compound
PCAWG	Pan Cancer Analysis of Whole Genomes
PCR	Polymerase Chain Reaction
PDD	Photodynamic diagnosis
PEN	polyethylene napthalate
PI3K	Phosphoinositide 3-kinase
PoN	Panel of Normal
QC	Quality control
r.t.	Room temperature
RJA	Rebecca Jane Astley
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RTB	Research Tissue Bank
SAM	Sequence alignment map
SAP	Shrimp alkaline phosphatase

siRNA	Small interfering RNA
SNV	Single nucleotide variant
STR	Short tandem repeat
TBE	Tris-Borate-EDTA
TE	Tris-EDTA
TGCA	The Cancer Genome Atlas
TMZ	Temozolomide
TNM	Tumour-node-metastasis
TRACERx	TRAcking Cancer Evolution through therapy [Rx]
Ts/Tv	Transition/Transversion
TURBT	Transurethral resection of the bladder tumour
UV	Ultraviolet
VAF	Variant allele frequency
VCF	Variant call format
VEP	Variant effect predictor
w/v	Weight per volume
WES	Whole Exome Sequencing
WGA	Whole genome amplification
WGP	Whole genome plots
WGS	Whole Genome Sequencing
WHO	World health organisation
WT	Wild-type
ºC	Degrees Celsius
µg	Microgram
µl	Microliter
µm	Micrometre
µM	Micromolar

Chapter 1
Introduction

1.1 Bladder Cancer

Bladder cancer is the most common cancer of the urinary tract and is the ninth most common cancer in the world with around 430,000 new cases and 165,000 deaths recorded in 2012[1]. Bladder cancer affects three times more men than women[2]. There are multiple risk factors for the disease including occupational exposure to aromatic amines and polycyclic hydrocarbons as well as smoking. Smoking is the highest risk factor and accounts for approximately 50% of cases[3].

Bladder cancer is divided into two main disease forms; non-muscle-invasive bladder cancer (NMIBC) and muscle-invasive bladder cancer (MIBC) (Figure 1.1). MIBCs are those which have invaded the muscle and consist of tumours that are of a high-grade and are of stage T2 to T4. At diagnosis, MIBCs make up about 20% of all bladder cancer cases. These patients have a poor outcome with a five-year survival rate of <50% and around 50% of patients will progress to metastatic disease[4]. NMIBC makes up 60-70% of cases and is limited to the urothelium and lamina propria[5]. NMIBC has a more favourable outcome with a five-year survival rate of around 90%[4]. Carcinoma *in situ* (CIS) is a flat, high-grade lesion and, although limited to the mucosa, it is believed to be the precursor to invasive disease[4,6].

Histological examination of tumours is used to assess stage and grade. The staging of tumours describes the degree of invasion using the Tumour-Node-Metastasis (TNM) system[7] whilst the grade defines the degree of differentiation of the tumour cells. Grading follows the World Health Organisation (WHO) 1973[8] or 2004[9] systems; low-grade (LG) tumours (Grade 1 and 2) are well differentiated whilst high-grade (HG) tumours (Grade 2 and Grade 3) are poorly differentiated (Figure 1.1).

Treatment of bladder cancer depends primarily on the stage and grade of the disease. Multifocality and tumour size are risk factors considered prior to treatment. Treatment of low-stage low-grade NMIBC patients involves transurethral resection of the bladder tumour (TURBT) followed by an intravesical instillation of mitomycin C (MMC) (or other chemotherapy) or, for higher-grade specimens, Bacillus Calmette–Guérin (BCG) therapy. Some patients will also undergo further instillations of chemotherapy[10].

Patients are then monitored by cystoscopy at regular intervals[6]. High-risk NMIBC patients (stage T1 and/or grade 3 tumours) are particularly difficult to manage. They have a reduced recurrence-free survival compared to other NMIBC and an increased mortality rate[11]. Clinicians and patients have to choose between bladder conservative therapy (TURBT plus BCG therapy) or a cystectomy, where the whole bladder and nearby lymph nodes are removed. In some cases, cystectomy may be a potential over-treatment and has its own risks associated with it[6,11]. Treatment of MIBC involves the removal of the bladder via cystectomy where possible as this can provide long-term disease free survival in over 70% of patients with organ-confined disease[12]. This is often combined with neoadjuvant systemic chemotherapy such as cisplatin-based combination therapies. For patients with non-organ confined disease, adjuvant chemotherapy or radiotherapy is often considered[6]. In the past few years, five immune check point inhibitors, known as immunotherapies, have been approved by the FDA for the second-line treatment of metastatic bladder cancer[13]. Additionally, two of the inhibitors have been approved for frontline use, expanding the treatment options for metastatic bladder cancers.

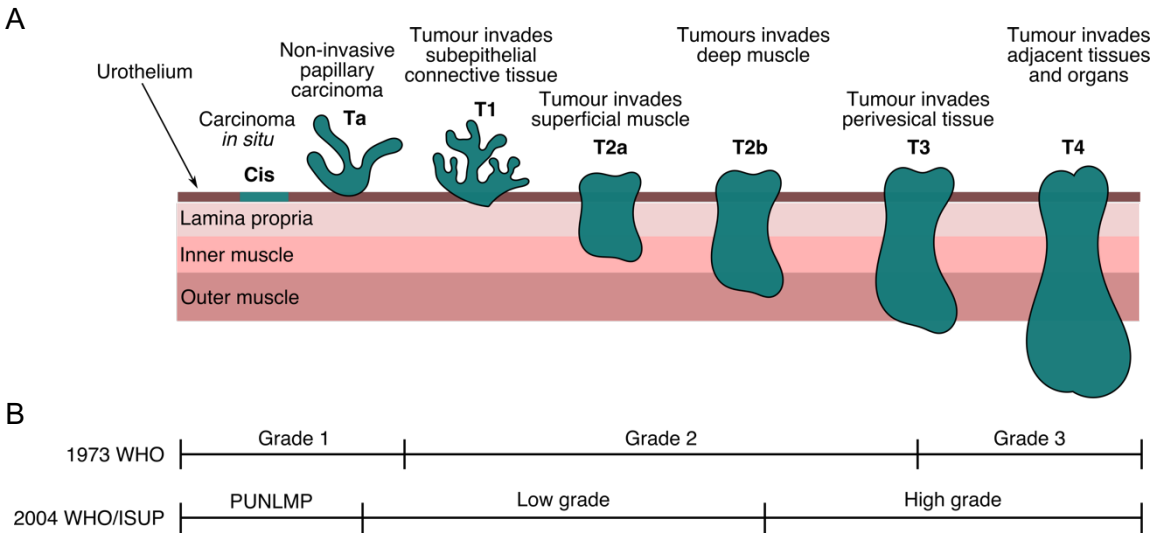


Figure 1.1: Bladder cancer staging and grading.
A) Staging of bladder cancer according to the TNM system[7]. Non-muscle-invasive bladder cancer consists of stages Tis, Ta and T1 whilst muscle-invasive bladder cancer encompasses stages T2 to T4. B) Grading according to the 1973 World Health Organization (WHO) and 2004 WHO/International Society of Urological Pathology (ISUP) criteria. PUNLMP = papillary urothelial neoplasm of low malignant potential. Figure adapted from Knowles and Hurst[4].

Despite being associated with a good outcome, around 50-70% of NMIBCs will recur and 10-15% will progress to muscle-invasive disease. At present, predictive models for the recurrence and progression of NMIBC are unsatisfactory[14] and there are currently no molecular biomarkers identified that can predict progression[4]. Consequently, patients require continued surveillance[6], making bladder cancer one of the most expensive cancers to treat[15]. This highlights the need for a greater understanding of the spatial and temporal dynamics of bladder cancer evolution; from diagnosis and throughout the disease history including treatment, recurrence and progression. An understanding of this at a molecular level will enable the identification of biomarkers for the prediction of recurrence and progression. This may also identify new therapeutic targets or provide knowledge to guide therapeutic regimes for the maximal effect for each individual patient. The recurrent and often multifocal nature of NMIBC, combined with the relative ease of tumour sampling, makes it an ideal candidate for the type of longitudinal analysis required.

1.2 Evolutionary theory, Cancer and Clonality

1.2.1 Cancer evolutionary theory

Charles Darwin's theory of evolution by natural selection is probably one of the most well-known scientific theories of all time. The theory contains three main requirements:

I. Individuals in a species show a wide range of variation
II. These variations are heritable
III. Variations lead to differential survival and reproduction in individuals[16]

The variations that provide an advantage to an individual and enhance their survival will be positively selected whilst less well-adapted individuals will be negatively selected. Peter Nowell was one of the first to describe tumour progression as an evolutionary process in his clonal evolution theory[17]. In his theory, mutations are the variation that leads to differential survival of tumour cells. Some mutations are evolutionary dead ends and will be detrimental to the survival of the associated cells, whereas other mutations may confer a selective advantage to a tumour cell. This cell can then proliferate and its progeny cells can undergo further diversification and selection[17]. This means that all cells should contain the original mutation(s) but they may differ in the subsequent mutations creating heterogeneity within and between tumours.

Within evolutionary theory there is a debate on whether macro evolutionary trends and rules exist or whether complex life as we know it is the culmination of an unrepeatable series of possibilities[18,19]. These opposing arguments are termed convergence and contingency respectively. Convergence suggests that there is a limited set of potentially repeatable outcomes due to constraints to evolution[20] and is presented as the development of a particular trait in independent lineages. The common evolutionary examples given to this argument are the evolution of wings (seen in bats, birds and insects) or the streamlined aquatic shape of some fish, dolphins and whales. Contingency on the other hand is best described by Gould's famous assertion that a different evolutionary outcome would result if the tape of life were to be rewound and replayed[21] i.e. evolution occurs by chance. These debates are also relevant to tumour evolution. If we can determine if there are spatial and temporal trends and patterns in tumour evolution, and we can decipher these, then this would be advantageous from a therapeutic perspective[20].

1.2.2 Historical methods for investigating clonality and heterogeneity

The idea that cancers are clonally derived has been investigated thoroughly throughout the years. Early studies into clonality often focused on investigating the clonal relationships between multiple tumours from the same patient including multifocal tumours[22] or primary and metastatic tumour-pairs[23], whilst early studies into heterogeneity focussed on assessing different regions from the same tumour[24,25]. This section provides a brief overview of methods used to investigate clonality and heterogeneity before the genomics era.

Loss of heterozygosity (LOH) analysis was a common method used to detect alleles that had been somatically lost in cancer cells. LOH is commonly assessed using polymerase chain reaction (PCR)-based analysis of polymorphic microsatellite markers in cancer cells and control germline samples from the same patient and has been used in the assessment of clonality of multifocal bladder tumours[22,26]. However, LOH analysis can be obscured by contaminating normal cells, homozygous deletions, karyotypic complexity and PCR artefacts[27] meaning that data often requires confirmation by independent techniques[28].

Other focused methods for looking at clonal relationships include the sequencing of specific genes that are known to be mutated in a particular cancer[28] such as *TP53* analysis[22,29]. A problem of these two approaches was that they were limited to analysis

of a particular locus or set of loci[28]. Shared markers may be missed whilst the presence of subclones and intratumour heterogeneity could lead to an incorrect assessment of non-clonality. Additionally, using highly mutable hotspots could indicate clonality when actually the events occurred independently.

Other studies have used karyotypic complexity to assess clonality and heterogeneity within many cancers such as glioma[24] and breast cancer[25]. Karyotype analysis uses ploidy and chromosomal banding patterns to identify differences in chromosome number and structure[30]. The main caveat to this technique is that it requires the culturing of the tumour cells which could alter clonal dynamics[28]. Fluorescence *in situ* hybridisation (FISH) has been used to study clonal heterogeneity in several cancers[22,31]. FISH uses fluorescently tagged probes complimentary to DNA regions to identify and label DNA with different types of probes used depending on the experiment. Probes can be targeted at centromeres for the specific detection of certain chromosomes or identification of aneuploidy[31]. Alternatively, probes can be designed to label specific genes to look for amplifications and deletions[32]. Additionally, probes have been designed to label entire chromosomes, known as whole chromosome painting and this can be used to identify chromosome translocations and large structural alterations[33] but sensitivity is an issue as smaller structural alterations may be missed. One of the advantages of FISH is that it can be performed on cells in interphase, allowing for the detection of alterations in non-dividing cells[34]. However FISH is labor intensive and less suitable than other methods for high-throughput studies[27].

Copy number alterations (CNAs) have also been used to investigate clonality in some cancers[23,35]. Comparative genomic hybridisation (CGH) was one of the first techniques to generate a genome-wide estimate of CNAs. Relative copy numbers are estimated using differential labelling of normal and tumour DNA mixed in a 1:1 ratio[36]. These are hybridised to a normal metaphase spread and gains in the tumour DNA will result in more of that DNA binding, giving a higher readout of the tumour fluorescent label. Likewise, loss of a region in the tumour will lead to a higher read of the normal fluorescent label. This provides a global overview of gains and losses in the genome of a tumour[36] and has the benefit of potentially many markers from which to identify relatedness. Despite not requiring culturing of the tumour cells this method still requires the generation of a metaphase spread of a normal cell[37]. The use of a metaphase spread limits sensitivity, as closely spaced events and events less than 20 Mb are difficult to detect[38]. CGH was later improved by hybridizing to an array of mapped sequences instead of a metaphase spread. This is known as array-CGH[38]. This has an improved resolution compared to regular CGH as each region is spatially separated on

a chip. This method has been widely used in the deciphering of clonality using CNAs[23,39].

Many of the technologies used in these studies have numerous limitations. The recent technological advances of next-generation sequencing (NGS) have allowed the analysis of structural rearrangements, CN alterations and somatic mutations of entire genomes to be studied at the single nucleotide level[40]. This unprecedented resolution has revolutionised the study of clonality and inter- and intra-tumour heterogeneity.

1.3 The next-generation sequencing era

Advances in sequencing technology over the past decade have revolutionised the field of cancer genetics. The introduction of massively parallel sequencing, known as next-generation sequencing, has reduced the cost of genome sequencing substantially: the human genome project was published in full in 2003, cost around $2.7 billion US dollars and took 13 years to complete, whilst today a human genome costs in the region of $1,000 US dollars and sequencing takes as little as a day although analysis can take much longer[41] (Figure 1.2). NGS has the ability to sequence heterogeneous mixtures of genomes simultaneously[42] and has enabled the analysis of structural rearrangements, copy number alterations and somatic mutations of entire genomes to be studied at the single nucleotide level[40]. This unprecedented resolution has revolutionised the study of clonality and inter- and intra-tumour heterogeneity.

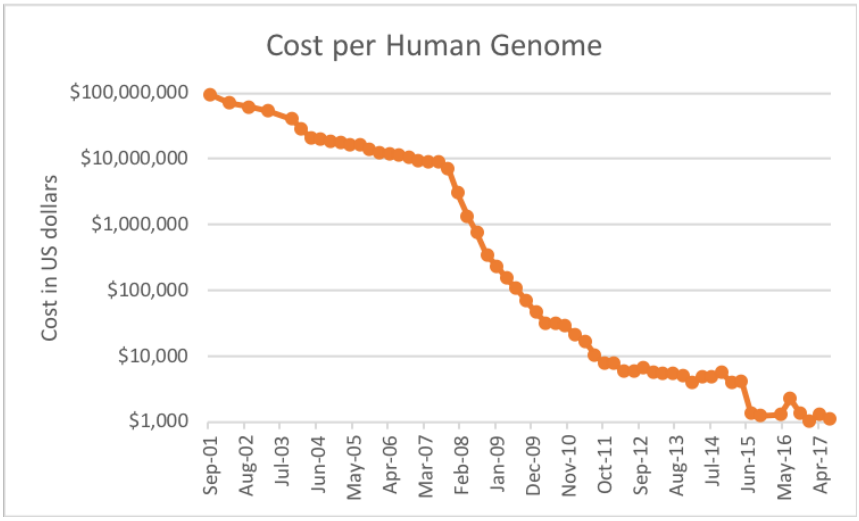


Figure 1.2: Cost of sequencing the human genome over the years.
The cost of sequencing the human genome has dramatically dropped since the introduction of NGS methodologies in 2008. Note the log-scale for the cost in US dollars. Data taken from[41].

1.3.1 Large genome sequencing projects

Since 2008 there have been thousands of cancer genomes and exomes sequenced, many as part of large projects under the International Cancer Genome Consortium (ICGC) or as part of The Cancer Genome Atlas (TCGA) project.

The TCGA was launched by the National Cancer Institute and the National Human Genome Research institute in the US in 2006. It initially started as a pilot project to investigate the genomic and molecular features of ovarian cancer and glioblastoma multiforme but was extended to run for over 10 years and analysed over 11,000 tumours from 33 of the most prevalent forms of cancer. This culminated in the production of the Pan-Cancer Atlas: a collection of 27 papers using cross-cancer analyses to explore overarching themes within cancer (https://www.cell.com/pb-assets/consortium/pancanceratlas/pancani3/index.html). These papers highlight the important discoveries in three key areas: cell-of-origin patterns[43], oncogenic processes[44] and signalling pathways[45]. Key findings include the correlation of aneuploidy with the somatic mutation rate and expression of proliferation genes, the identification of pathogenic germline variants[46] and the characterization of driver genes and mutations[47]. As well as publishing pan-cancer findings, TCGA have also produced detailed investigations into individual cancers through their working groups.

Due to the shear scope of the cancer genomics field, the ICGC was created to coordinate large-scale cancer genome projects with the aim of defining the genomes of 25,000 primary, untreated cancers. The key motives for the formation of the ICGC were to reduce the duplication of effort, generate more complete studies and provide a framework for the standardization of data to enable the merging and comparing of datasets[48]. This has resulted in over 20,000 tumour genomes available world-wide, providing a rich resource of data for cancer researchers. TCGA was a large provider of data to the ICGC and a recent initiative, the Pan Cancer Analysis of Whole Genomes (PCAWG), is a collaboration between the ICGC and TCGA[49]. This aims to analyse over 2,800 whole cancer genomes from the ICGC and explore somatic and germline variants in both coding and non-coding regions focusing on cis-regulatory sites, non-coding RNAs, and large-scale structural alterations.

These large projects have been instrumental in understanding the changes occurring in different cancers and for the identification of pan-cancer patterns. However, both projects have thus far focussed on the genomic characterisation of untreated cancers from the primary site. Whilst this is crucial information, it does not address the continuing development of cancer after the primary tumour. Only by the analysis of

multiple tumours over time and/or space within a patient can we really begin to understand the development and evolution of cancer throughout its lifespan. The ICGC have now developed an initiative named ICGC-ARGO; Accelerating Research in Genomic Oncology. This will use key clinical questions and patient clinical data to drive the interrogation of cancer genomes with the aims of answering questions related to cancer evolution as well as hoping to identify better ways of using treatments.

1.3.2 Heterogeneity and tumour evolution

Deep sequencing within and between tumours has provided an improved understanding of the evolutionary trajectory of cancer. Originally it was thought that cancer was a linear disease with successive rounds of acquisitions of advantageous mutations by a clone, followed by a selective sweep in which this advantageous clone would expand and replace the less fit clones in the tumour[40,50]. However current evidence suggests that this simple view occurs only rarely, such as in AML[51], and the majority of cancers evolve in a branched or punctuated pattern resulting in tumours consisting of multiple different clones[52,53]. Cancers undergoing branched evolution will therefore be a dynamic population of clones resulting in spatial and temporal heterogeneity.

Spatial and temporal heterogeneity has been described in many cancers including breast[54], bladder[55], lung[56], renal[57] and prostate[58] to name a few. A recent pan-cancer analysis of whole genomes investigating intra-tumour heterogeneity identified evidence of recent subclonal expansions in over 95% of the tumours analysed[59]. As this work was carried out on single samples this can be considered a lower limit of intra-tumour heterogeneity as variants found to be clonal in one sample may be subclonal in another[59].

Researchers have investigated both recurrent[51,60,61] and metastatic disease[58,62-64] in a bid to understand clonal dynamics and factors influencing disease recurrence and progression. Analysis of spatially separated areas of the primary tumours and metastasis of renal cell carcinoma uncovered a branched evolutionary tract with one branch evolving into the clones at the metastatic sites and the other diversifying into primary regions[53]. Sequencing of spatially distinct areas of the primary tumour identified extensive intratumour heterogeneity with only 34% of all detected mutations being present in all regions sequenced. This spatial separation of clones has also been noticed in other cancers such as non-small cell lung cancer (NSCLC) where a study identified that if only one region of a tumour was sequenced, the probability of missing a potential driver gene was 83%[56]. These studies indicate that single biopsies from

solid tumours may not be representative of the entire tumour bulk, and this has important implications for therapy.

1.3.3 Mechanisms driving heterogeneity and evolution

1.3.3.1 Genetic drivers of heterogeneity

As described above, studies investigating tumours over time or space have identified a significant amount of heterogeneity both within and between tumours. This heterogeneity is the fuel for tumour evolution as it provides an array of mutations which selective pressures can act upon[65]. Through the use of high throughput sequencing, a deeper understanding of the mechanisms driving heterogeneity has been gained (Figure 1.3).

In order to generate the heterogeneity identified in cancers, many mutations need to accumulate. Using mathematical modelling, Loeb showed that the moderately large numbers of mutations seen in a colorectal cancer could not have arisen within the lifetime of a human if the mutation rate is the same as observed in normal tissues[66]. He suggested that cells need to adopt a mutator phenotype, which increases the rate of genetic mutation. The necessity for this mutator phenotype is contested[67,68] and whilst pan-cancer studies clearly show an elevated mutation frequency for most cancers[69,70], it still remains unclear if this is due to an increased mutation rate or simply due to more cell divisions at a low mutation rate[52].

The mutational diversity required for evolutionary processes can be generated in many ways. Genomic instability is a huge driver of genomic diversity and refers to an increase in the amount and tendency for alterations in the genome[71]. Genomic instability processes can create characteristic mutations leaving a pattern of base changes known as a “mutational signature”. The availability of large datasets has shown that whilst a cancer may have a few driver mutations, it can contain many more passenger mutations[72]. Although these passenger mutations may not contribute to disease, they are a vast resource that can be used to investigate these mutational signatures[70,73,74]. The spectrum of known mutational signatures was greatly expanded in the landmark paper by Alexandrov *et al.*[70] who used data from over 7,000 tumours to identify 22 mutagenic signatures. Ongoing work by several groups has expanded the number of signatures identified and there are now 30 signatures in the Catalogue Of Somatic Mutations In Cancer (COSMIC) database (https://cancer.sanger.ac.uk/cosmic/signatures). Analysis of these signatures has identified the large contribution of exogenous mutagens, such as tobacco smoke and

UV, as well as endogenous processes, such as dysfunctional DNA damage response pathways and mutagenic activity of the apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) family of cytidine deaminases, towards genomic instability.

There are several mutational signatures corresponding to DNA repair deficits[70]. An analysis of the mutational landscape of 12 major cancer types identified a significant correlation between high mutation frequency and DNA repair pathway genes[69], suggesting that deregulation of DNA repair is common across at least a subset of cancers. Defects in mismatch repair can lead to increased instability at the nucleotide level because replication errors cannot be repaired effectively[75]. This results in increased rates of frameshift and/or point mutations[76] and can generate microsatellite instability. Mutations in DNA damage response pathways have also been linked to genomic instability in hereditary cancer[68,77] and alter the cells ability to deal with DNA damage. Alterations in the ability of cells to detect, analyse and repair DNA damage can cause the accumulation of genetic alterations leading to an increased mutation load and genetic instability.

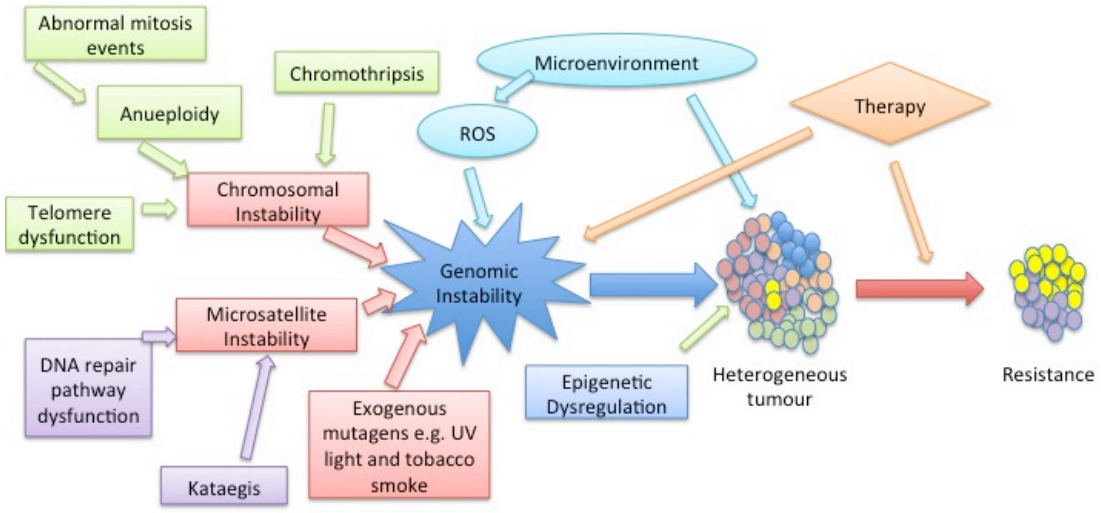


Figure 1.3: Factors driving heterogeneity, tumour evolution and resistance.
Genomic instability is driven by many processes and results in the production of many different tumour subclones resulting in a heterogenous tumour. The microenvironment can select for the fittest clones in the local environment, fuelling tumour evolution, and can also increase genomic instability. Therapy provides a potent selection barrier that can result in the emergence of only a subset of resistant clones and can also contribute to genomic instability.

Other prevalent endogenous mutagens include the activity of APOBEC cytidine deaminases. These enzymes have functions in innate immunity towards viruses and endogenous retroelements as well as in RNA editing[78]. However, they can also cause mutations in DNA through the deamination of cytosine residues in single stranded DNA[79,80]. This leaves a characteristic mutational signature of C > T or C > G mutations at TpC sequences[70,80]. APOBEC mutagenesis has been shown to be highly prevalent in several different cancer types, including bladder cancer[81,82]. In breast cancer, upregulation of APOBEC3B correlated with a doubling of C > T substitutions and overall mutation load[79]. In lung cancer, APOBEC mutations were shown to be present on the branches of the tumour tree rather than the trunk of the tree[56]. Interestingly, mutations in an APOBEC context were identified in driver genes, such as *PIK3CA* and *EP300*, on the branches suggestive of a possible impact of APOBEC mutagenesis on subclone diversity[56]. Indeed, *PIK3CA* is a common target of APOBEC mutagenesis with mutations in the helical domain displaying an APOBEC motif in human papillomavirus driven tumours[83].

Chromosomal instability is another common cause of genomic instability and can be characterised by structural and numerical karyotypic heterogeneity[84]. Aneuploid cells have an unbalanced number of chromosomes and aneuploidy is often attributed to chromosomal partitioning errors during mitosis[85]. In the cancer literature, aneuploidy has recently been further defined as somatic CNAs involving whole chromosomes or complete arms of chromosomes[86]. A recent pan-cancer analysis examining aneuploidy identified that 88% of cancers investigated had some aneuploidy. However the rate varied across cancer types; only 26% of thyroid carcinomas contained an aneuploid event, compared to 99% of all glioblastomas and testicular germ cell tumours[86]. A previous analysis of over 3,000 tumours from 12 tumour types identified an inverse relationship between the number of recurrent CNAs and the number of SNVs[87]. Contrastingly in this study, a positive correlation between aneuploidy score and mutation frequency was observed when hypermutated samples with high levels of microsatellite instability or POLE mutations were excluded[86]. Aneuploidy therefore may not only contribute to genomic instability through changes in large numbers of genes but may be linked to genomic instability at the SNV level.

Not all mutations occur progressively. Investigations into whole cancer genomes have provided evidence that in some cases a huge number of mutations can occur in a short time frame. A process, termed chromothripsis, in which chromosomes shatter and are reassembled resulting in massive genome rearrangement shows how multiple genome rearrangements can occur in one catastrophic event[88,89]. Telomere attrition has also

been linked to massive genomic disruption[90] and Kataegis, where multiple point mutations are acquired in one-off bursts, has been identified in breast cancer[91] and multiple myeloma[92]. These catastrophic events suggest that cancer evolution is not always a gradual process and may provide a clone with a leap, rather than a step, towards malignancy[93].

1.3.3.2 Non-genetic drivers of heterogeneity and tumour evolution

Evolution and heterogeneity do not have to be driven by genetic factors. Epigenetic changes, such as DNA methylation and histone deacetylation, are increasingly being recognised as being important during tumorigenesis[94,95]. Aberrant DNA methylation has been identified in many cancer types[96]. As epigenetic changes are inherited during the cell cycle and alter gene expression, they are also subject to evolutionary forces and can drive clonal evolution. Indeed, intratumour heterogeneity of DNA methylation has been shown to reflect clonal evolution in aggressive prostate cancer[97]. Deep sequencing studies have identified mutations in epigenetic modifiers in many cancers[69,98-100] suggesting that deregulation of this process may be important to cancer progression.

The tumour microenvironment also provides a strong role in the evolution of cancer. It is a complex system including surrounding vasculature, immune cells, fibroblasts and the extracellular matrix[101]. The interaction between cancer cells and their environment is implicated in cancer growth with invading immune cells inducing chronic inflammation and secreting factors that aid in tumour progression, allowing cancer cells to acquire hallmark functions[102]. Tumours may be made up of millions of cells each responding to the environment directly surrounding it. Different microenvironments may have varied degrees of selective pressures including acidity, oxygen and tumour growth factors[103]. Even early in carcinogenesis hypoxia can be present, providing a harsh landscape in which only the well-adapted cells can survive. Not only does hypoxia provide a form of selection but it can also lead to genomic instability through mechanisms such as reactive oxygen species (ROS). ROS are also produced during chronic inflammation and they can induce DNA damage and reduce activity of DNA damage response machinery[77]. Using these mechanisms, the microenvironment not only contributes to genomic instability, thus generating diversity and influencing heterogeneity, but can also act as a selection barrier and has been associated with chemotherapy resistance in serous ovarian cancer[104].

Therapy plays an important role in tumour evolution. It can contribute towards genomic instability, such as seen with temozolomide (TMZ) treatment in malignant melanomas and gliomas[70]. In chronic lymphocytic leukaemia (CLL), therapy correlated with an increase in subclonal mutations[105], and in glioma TMZ treatment induced a hypermutator phenotype, likely through the mutation of the MSH6 mismatch repair gene, leading to genomic instability and resistance[16]. This increase in mutations can lead to an increase in tumour diversity for evolution to act upon. Therapy also acts as a potent selection barrier as only clones resistant to the therapy can survive and propagate.

Genomic instability has been linked to drug resistance, poor prognosis and even progression in certain cancers[106-109]. These elevated levels of DNA variation combined with the selection forces of the local environment lead to inter- and intra-tumour heterogeneity. This heterogeneity within and between the tumour masses is the basis for selection of the fittest clones and thus is a major step in clonal evolution[110].

1.3.4 ITH and treatment

Tumour heterogeneity, both within and between tumour lesions, may have huge implications for drug therapy in cancer. The most critical challenge for oncologists is drug resistance[42]. Targeted therapies aim to eradicate cancer by blocking key signalling proteins and cancers are often treated depending on the presence or absence of these targets. For example, in CML, patients are treated with imatinib, a competitive inhibitor of the BCR-ABL oncogene[111]. However, resistance can occur and examination of patients pre and post treatment (with imatinib) has identified very low frequency subclones carrying resistance mutations in the kinase domain of BCR-ABL prior to treatment in some patients. This emergence of resistant subclones has been identified in many cancers[112-114] and multiple resistance mechanisms can occur in the same patient[115]. This highlights that single targeted therapeutics are likely to be futile and raises serious questions about the need for greater analysis of tumour heterogeneity before targeted treatment in order to determine the best combination therapy strategy.

It is also possible that tumour heterogeneity can fuel *de-novo* drug resistance mechanisms through the contribution of genomic instability and/or mutagenic therapies. Chemotherapy-induced mutagenesis has been identified in AML and glioma[60]. In glioma, the mutagenic TMZ therapy combined with genomic instability led to hypermutation of the genome and malignant progression[114]. This also illustrates that

clonal evolution and heterogeneity can also be implicit in resistance to conventional chemotherapy. One problem with the categorising of resistance to *de-novo* mechanisms it that it is difficult to establish whether resistant subclones are generated during treatment or if they are present at very low levels prior to treatment[116]. Even ultra-deep sequencing may miss a subclone due to sampling bias or the subclone being particularly rare.

Studies have begun to trace clonal evolution in cancer during treatment. Unsurprisingly, many of these studies have been in haematopoietic cancers[60,112,113,117,118] as longitudinal sampling is much easier in these cancers[116]. A study in acute lymphoblastic leukaemia (ALL) identified that the relapsed clone was often ancestral to the diagnostic sample and backtracking analyses also identified the relapsed clone as a minor subpopulation at diagnosis[112]. Different clonal architecture at relapse compared to diagnosis is common[117,118] and relapse can be driven by more than one subclone[117]. The dynamics of subclonal architecture was demonstrated in multiple myeloma in a case sampled at seven time points. This showed alternating clonal dominance between two subclones over the disease course[113]. These studies emphasise the dynamics in clonal evolution and demonstrate that therapy acts upon these dynamics, selecting for resistant clones.

Fewer longitudinal studies have taken place in solid tumours, likely due to difficulties in sampling[116]. In *BRAF*-mutant melanoma, intratumour heterogeneity was seen as a major contributor to BRAF inhibitor resistance with temporally separated resistant clones and multiple resistance mechanisms[115]. Similar to the ALL study, ancestral clones were found in recurrent gliomas at recurrence compared to the dominant clone in the primary tumour[114]. Temporal evolution was identified in a study comparing mutations in a metastatic lobular breast cancer with the primary tumour from 9 years previously. Only 5 of the 32 mutations in the metastatic cancer were present in the primary tumour[119]. This shows that the tumour underwent substantial evolution during this time frame highlighting the implications that time has on heterogeneity.

These studies show that clonal evolution and heterogeneity have an impact upon treatment, yet the understanding of this is still limited. To address this issue a large translational research study TRACERx (TRAcking Cancer Evolution through therapy [Rx]) has been set up with a focus on four cancer types: lung, melanoma, prostate and renal cancers. The aim of the study is to determine the relationship between heterogeneity and disease stage, clinical outcome and treatment response[120]. To date, interim findings have only been published for clear-cell renal cell carcinoma (ccRCC).

These have identified that the early events in ccRCC (chromosome 3p loss and 5q gain) happen extremely early, likely occurring during childhood or adolescence for the majority of the patients, 30-50 years before the kidney cancer was diagnosed[121]. Multi-region sequencing identified a higher frequency of driver mutations compared to studies using single biopsies, emphasizing the importance of intra-tumour heterogeneity[122]. Convergence on the *VHL* pathway was detected in patients with multifocal or synchronous bilateral disease and parallel evolution of mutations in the same genes or pathways within distinct tumour subclones was identified in 13% of untreated primary tumours[122], confirming observations from other studies[57,64]. These results suggest a deterministic nature to clonal evolution[122]. Analysis of metastases identified distinct patterns of metastatic dissemination and the presence of profound evolutionary bottlenecking[123].

Overall, a deterministic nature of the evolution of ccRCC has been described, the understanding of which could stratify patients for surgical intervention or therapeutic drug intervention[123]. Similar analyses are now needed in other cancers in order to define the evolutionary trajectory of each cancer type. If we can understand each cancer's evolutionary process, especially with regards to treatment, then the outcomes of cancer growth and therapy response could be predicated. This would provide a basis on which to design therapeutic interventions that would be best for each individual patient and may identify evolutionary constraints that could be exploited[76,124].

1.4 The genomic landscape of bladder cancer

1.4.1 The pre-genomics era

Prior to the advancement of NGS, several molecular alterations were characterised in bladder cancer using techniques such as LOH analysis, copy number analysis and analysis of mutations in candidate gene studies. Many structural alterations have been identified in the genomes of bladder cancer tumours, including DNA copy number losses and gains, rearrangements, and regions of LOH. Typically, MIBCs exhibit many CNAs and can be highly aneuploid whilst NIMBC, especially stage Ta tumours, exhibit fewer CNAs[125] and are often diploid or near diploid. Stage T1 tumours exhibit a mix of profiles[126], with some stage T1 tumours having few CNAs, resembling stage Ta tumours, whilst others have more unstable genomes and more closely resemble MIBC profiles, suggesting the presence of multiple tumour subgroups[125].

An early karyotyping study in bladder cancer identified high levels of monosomy of chromosome 9[127]. Loss of chromosome 9, or parts thereof, has been confirmed in

subsequent studies[125,126,128-130] and alterations of chromosome 9 by LOH or copy number loss have been identified as the most common structural alterations present in over 50% of all bladder tumours, irrespective of stage and grade[125,131]. Chromosome 9 contains two key tumour suppressor genes in bladder cancer which are affected by chromosome loss. *CDKN2A,* which resides at cytoband location 9p21 and encodes cell-cycle regulators p16 and p14ARF, is commonly inactivated by homozygous deletion (HD) and has been linked to an increased risk of recurrence in NMIBC[132]. On 9q, *TSC1* is the best characterized tumour suppressor gene. This gene is present in many regions of LOH and is mutated in 12-16% of bladder cancer cases[133,134].

Along with loss of chromosome 9, additional alterations have been identified in stage Ta tumours. These include losses of chromosomal regions 10q(20%)[135], 11p(10-24%)[125,135,136], 11q(21%)[125], 17p(15-19%)[125,135], 19p(19%)[125] and 19q(19%)[125], and gains of 20q(13-17%)[135,136]. MIBC have many copy number alterations[125,135] with frequent alterations apparent in almost all chromosomes[125,137].

Key genes involved in bladder cancer, such as *FGFR3, PIK3CA, TP53* and the RAS gene family, were initially identified in candidate gene studies. *FGFR3* (fibroblast growth factor receptor 3) is one of four tyrosine kinase receptors that mediates the intracellular signalling of fibroblast growth factors (FGFs) [138]. Activation of the receptor can lead to the activation of several signalling pathways including the mitogen-activated protein kinase (MAPK) and phosphoinositide 3-kinase (PI3K) pathways[139]. Mutations of *FGFR3* in bladder cancer were identified in the late 1990's[140] and have since been shown to be highly prevalent in the disease[141-143]. Activating point mutations are present in up to 80% of stage Ta tumours[144] whilst mutation rates in MIBC and stage T1 tumours are lower[141,145] with mutations found in 10-45% of stage T1 tumours and around 15% of MIBC[144]. Mutant *FGFR3* has been shown to increase cell proliferation and promote anchorage-independent growth *in vitro*[146] and is thought to contribute to urothelial hyperplasia *in vivo*[144].

The PI3K pathway is an important transducer of signals from tyrosine kinase receptors and is involved in signalling pathways regulating cell growth, differentiation and development[147]. PI3K signalling pathways are often disturbed in cancer and in bladder cancer, activating mutations of the p110α catalytic subunit (PIK3CA) have been identified[148]. Like *FGFR3, PIK3CA* mutations are more common in stage Ta tumours compared to stage T1 NMIBC and MIBC (~40-50% vs ~20%)[134,148] and *PIK3CA* mutations have been shown to commonly occur with *FGFR3* or RAS mutations[143,148].

Mutations in the RAS gene family were the first genetic alterations implicated in cancer and activating mutations in these genes are common in many cancers[149-151]. In bladder cancer, mutations in *HRAS* and *KRAS* are more frequent than mutations in *NRAS*[143]. Unlike *FGFR3* and *PIK3CA*, RAS mutations are not associated with stage or grade with 6-18% of all tumours containing mutations[143,152].

As both *FGFR3* and the RAS family of genes activate the MAPK pathway, Jebar *et al.*[142] investigated the possibility of mutual exclusivity in these genes through the analysis of a mixture of tumour samples and tumour-derived urothelial cell lines. They directly sequenced exons 7, 10 and 15 of the *FGFR3* gene to detect mutations whilst exons 1 and 2 of the *HRAS, KRAS* and *NRAS* genes were investigated for mutations using fluorescent single-strand conformation polymorphism analysis. This was followed by sanger sequencing of samples with potential mutations. This identified a striking mutual exclusivity of these alterations which has since been confirmed in further studies[143,153] including NGS studies[82].

The most common genetic alterations in bladder cancer are mutations of the telomerase reverse transcriptase (*TERT*) promoter which occur in all stages and grades at a very high frequency (60-80%)[153-155]. *TERT* encodes the catalytic sub-unit of the telomerase ribonucleoprotein and is transcriptionally repressed in human non-progenitor or non-stem cells[154]. Hotspot mutations in the *TERT* promoter region create binding motifs for E-twenty-six (ETS) transcription factors which increase transcriptional activity[155]. Due to the high level of telomerase activity in human cancers, it is thought that the activation of telomerase is essential for the immortalization of human cells[154]. Indeed, the identification of *TERT* promoter mutations in both NMIBC and MIBC also suggests that this is a requirement for all pathways of bladder cancer[156].

The mutational spectrum of these well-characterised bladder cancer genes has been confirmed in NGS studies[81,82,157]. These NGS studies have also identified additional recurrent mutations and mutational processes shaping the genomic landscape of bladder cancer. These studies are described in section 1.4.3.

1.4.2 The clonal origins of multifocal and recurrent bladder cancer

Approximately 30% of bladder cancer patients develop multiple synchronous tumours (multifocality)[158]. Two theories have been put forward to explain multifocality; the

monoclonal and the field-cancerization hypotheses[5]. The monoclonal theory suggests that all tumours are descendants of a single malignant cell, which has proliferated, and its descendants have spread throughout the urothelial lining. This could be by intraepithelial spread, where malignant cells migrate throughout the epithelium, or by intraluminal seeding where tumour cells are released from the primary site into the urine where they implant at a secondary site. This spread of monoclonal cells throughout the urothelium is often referred to as a clonal "field change"[4]. The field cancerization theory suggests that carcinogens in the urine affect the urothelium at many sites, allowing several cells to obtain mutations capable of driving tumorigenesis, leading to multiple tumours that have arisen from genetically distinct cells[159].

Many studies have investigated the clonality of multifocal bladder cancer. Louhelainen *et al.*[26] used LOH analysis to demonstrate monoclonality in non-muscle-invasive multifocal bladder cancer. Monoclonality of multifocal tumours was also identified by Sidransky *et al.*[160]. This study used X-chromosome inactivation to assess clonality, but X-chromosome inactivation is limited to analysis of samples from female patients only whilst the majority of bladder cancer patients are male. Additional limitations to this method include the problem of preferential inactivation of a particular allele or, for spatially close tumours, cells may have originated from the same stem cell, and will exhibit the same chromosome inactivation pattern as a consequence[159]. A few patients do show evidence for oligoclonal (i.e. independently forming) tumours. Hafner *et al.*[161] used LOH and *TP53* mutation status to assess tumours and found a mix of monoclonal and oligoclonal tumours.

Studies examining recurrent disease have identified clonal relationships between recurrent tumours[35,162-165]. Xu *et al.*[162] demonstrated monoclonality between the majority of recurrent lesions assessed using *TP53* mutation status. *TP53* mutations, along with LOH in the 17p13 and 9p21 regions, were also used by Trkova *et al.*[165] to determine a clonal relationship between recurrences. A limitation of using mutations in *TP53* to investigate clonality in bladder cancer is their limited applicability to analysis of low-stage low-grade tumours that have very few alterations in this gene[166]. LOH has also been used to look at recurrences spanning periods of up to 17 years[163]. This study identified a clonal relationship in all the patients studied. Interestingly, they observed that the chronological order of tumour presentation did not parallel the genetic evolution of the tumour. Lindgren *et al.*[167] examined both metachronous and synchronous tumours using CGH, LOH and mutation analysis. They noted that although most alterations were clonal, the recurrent tumours were unlikely to have

originated from the preceding tumour, further highlighting the complex genetic relationship between metachronous tumours.

Some recurrent tumours may not arise from a monoclonal origin; CN analysis of recurrent tumours from four patients identified a monoclonal origin in three of the patients[35]. In the fourth patient a clonal relationship between two out of three tumours was identified, however the third tumour only shared a very small proportion of aberrations, making the authors question its clonality. Interestingly, this tumour was an invasive tumour (T3G3) whilst the other two were superficial TaG1 tumours[35]. Further analysis of mutations in tumours from this patient would aid in determining if these tumours were indeed clonally related.

The complex relationships identified in these studies are suggestive of a widespread 'field change' in which a large amount of tissue contains molecular alterations[4]. Indeed, whole organ mapping identified a clonal relationship between geographically distant regions of altered mucosa[168]. These regions often appeared phenotypically normal supporting the idea of the clonal "field change" (Figure 1.4).

With the advent of NGS techniques, studies have been able to take clonality analysis one step further by analysing tumours down to the single cell[169]. This has provided a huge insight into the clonal dynamics of bladder cancer and NGS studies of bladder cancer are discussed in the next section.

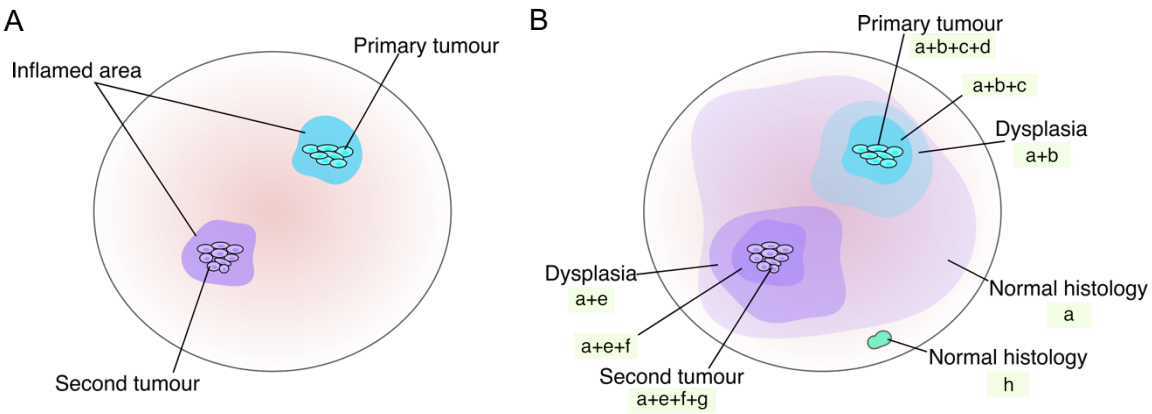


Figure 1.4: Alterations in the tumour-bearing bladder.
Circles represent the interior of the bladder cavity. A) A representation of what the surgeon sees at TURBT. At resection the primary tumour and an area of inflamed urothelium are resected. A recurrence is identified during surveillance and a second tumour and inflamed region are resected. B) The molecular picture. Large areas of the urothelium have been replaced with cells containing alteration "a" from which the primary and secondary tumours have developed. Within this field of change are regions of dysplasia which contain additional alterations to "a". Both tumours contain the ancestral alteration "a" but they differ in their subsequent mutations. A region with normal histology with alteration "h" shows the possibility that independent initiating events could occur resulting in oligoclonal tumours. Figure adapted from Knowles and Hurst[4].

1.4.3 The next generation sequencing era and bladder cancer

Several studies have now employed NGS methodologies in the study of bladder cancer[61,81,82,100,157,164,170-174]. The largest study has been the comprehensive analysis of 412 chemotherapy naïve MIBCs undertaken by the TCGA. In this study, tumours underwent WES, RNA-sequencing, DNA copy number analysis, methylation analysis and protein expression profiling[81]. This identified that MIBC is characterized by a high mutation rate with a mean of 8.2 non-synonymous mutations per megabase (Mb). A high frequency of alterations in genes involved in cell cycle regulation was identified and this has been shown to have an adverse prognostic significance[175]. *TP53* was the most frequently mutated gene, altered in 48% of tumours[81] consistent with other MIBC sequencing studies[176]. Chromatin modifiers and regulators also contained frequent alterations. Interestingly, 67% of all single nucleotide variants (SNVs) were identified within an APOBEC signature sequence context suggesting that APOBEC related mutagenesis is extremely prevalent in MIBC. Unsupervised clustering of samples by molecular signatures produced 4 clusters, one of which was characterized by a high APOBEC-associated mutational load as well as a high overall mutation burden. Patients within this cluster showed a high overall 5-year survival rate of 75% and this contrasted with patients in the cluster that had the lowest mutation rate where 5-year survival was just 22%. This high mutation rate correlated with a high predicted neo-antigen load and it is hypothesized that the improved survival of these patients is likely due to a natural host immune reaction[81].

As MIBC has benefited from inclusion in TCGA studies, many research groups have focused their sequencing efforts on the analysis of NMIBC[61,82,157,164,173]. Hurst *et al.*[82] analyzed 140 primary stage Ta tumours, the majority of which were of a low grade, using a mix of WES, targeted gene panel sequencing, copy number analysis and microarray-based genome-wide mRNA expression analysis whilst Pietzak *et al.*[157] used a targeted cancer gene panel to analyse 105 treatment-naïve NMIBC of mixed stages and grades. These studies have begun to build a detailed picture of the mutational landscape of NMIBC (Table 1.1). These studies have confirmed the high mutation rates seen in *FGFR3* and *PIK3CA*[82,157] and have clearly demonstrated the differential frequency of *TP53* mutations across stages and grades; mutations were identified only very rarely in Ta low-grade tumours[82,157] whilst an increased rate was observed in high-grade and stage T1 tumours[157,173].

Overall, mutation rates are lower in NMIBC with a mean mutation rate of 2.41 mutations per Mb identified in stage Ta low grade tumours[82]. Studies have also identified APOBEC as a large contributor to the overall mutation burden in

NMIBC[61,82,164] with a contribution to 35% of overall mutations identified in Hurst *et al.*[82]. Interestingly, up to 60% of mutations were APOBEC-related in some individual tumours[82]. In high-risk NMIBC, a high mutational burden was linked to an increased response to BCG therapy and a lower progression rate[173] though APOBEC signatures were not assessed.

An interesting finding of these genome sequencing studies is the high frequency of chromatin modifier mutations identified. Frequently mutated chromatin modifier genes include *KDM6A* (a histone demethylase), the histone methyltransferases *KMT2A*, *KMT2C*, and *KMT2D*, the histone acetyltransferases *CREBBP* and *EP300,* and *ARID1A* (part of the SWI/SNF complex) amongst others (Table 1.1). These mutations were initially identified by Gui *et al.*[100] who were the first group to report results from exome sequencing in bladder cancer. Since then, mutations in these genes have been identified in bladder tumours of all stages and grades[81,82,157,164,173] and these mutations are often inactivating with nonsense mutations, small insertions and deletions and mutations in essential splice site regions identified[61,81,82,157].

KDM6A is the most frequently mutated chromatin modifier gene in NMIBC[61,82,157,164] with significantly more mutations than in MIBC[81,164]. *KDM6A* catalyses the demethylation of tri/di-methylated histone H3 lysine 27 (H3K27me2/3) creating a transcription-permissive chromatin state[177]. *KDM6A* forms a complex with *KMT2C/D*[178] and this acts to maintain gene expression. Inactivation of these genes would be predicted to result in reduced transcription. Intriguingly, more *KDM6A* mutations have been identified in non-muscle-invasive tumours from females than males, potentially indicating a gender difference in the epigenetic landscape[82]. However, the relatively small numbers of NMIBCs sequenced combined with a small proportion of those being from females is a major limitation and many more samples need to be analysed. Analysis of MIBC from TCGA did not show the same bias[81].

Table 1.1: Frequently mutated genes in NMIBC and MIBC.
Table from Knowles and Hurst[179]. [a]Genes presented were mutated at ≥10% in the studies of Hurst *et al*.[82] and TCGA 2017[81]. [b]Data from Hurst *et al.*[82] consisting of 79 TaG1/G2 and 3 TaG3 tumours. WES sequencing of 24 tumours and targeted sequencing of 58 tumours was carried out. Mutation frequency is shown in brackets where only WES sequencing data is available. [c]Data from Nordentoft *et al.*[164] WES sequencing was carried out on 20 TaG1/G2 tumours. [d]Data from Pietzak *et al.*[157] Targeted sequencing of 55 Ta (23 grade1/2; 32 grade 3) and 38 T1 tumours was carried out. "NA" indicates that genes were not covered by the targeted panel used. [e]Data from Guo *et al.*[172] WES sequencing was carried out on 32 T1 tumours. [f]Data from TCGA 2017[81]. WES sequencing was carried out on 412 MIBC.

Gene[a]	Hurst *et al*. 2017[b] Ta (%)	Nordentoft *et al*. 2004[c] Ta (%)	Pietzak *et al*. 2017[d] Ta (%)	Pietzak *et al*. 2017[d] T1 (%)	Guo *et al*. 2013[e] T1 (%)	TCGA 2017[f] MIBC (%)
FGFR3	79	40	66	30	25	14
PIK3CA	54	25	36	22	6	22
KDM6A	52	65	50	43	50	26
STAG2	37	25	24	22	25	14
KMT2D	30	15	31	26	0	28
ARID1A	18	35	25	27	6	25
EP300	18	25	20	8	16	15
CREBBP	15	20	23	19	12	12
KMT2C	15	20	16	5	3	18
RHOB	13	0	NA	NA	0	11
HRAS	12	10	2	8	16	9
KMT2A	11	0	9	11	9	11
TSC1	11	5	5	22	12	8
BRCA2	10	0	11	11	0	7
COL11A1	10	0	NA	NA	0	5
RBM10	10	20	22	5	0	9
TP53	4	5	11	35	25	48
FAT1	(2)	10	13	17	0	12
KRAS	2	0	11	8	6	4
ATM	(1)	5	13	19	3	14
CDKN1A	(1)	0	11	13	0	9
ELF3	(1)	25	NA	NA	12	12
ERCC2	(1)	0	21	13	6	9
ERBB2	(0)	0	11	19	3	12
ERBB3	(0)	0	9	19	3	10
RB1	(0)	5	0	5	9	17

A few studies have sequenced multiple bladder tumours or tumour regions from individual patients[55,61,63,164,174,180]. NGS has been employed to study the clonal relationships between multifocal tumours[174,180]. Analysis of physically separated, synchronous, multifocal tumours from 3 patients identified a clonal relationship in all three cases[180]. Interestingly APBOEC signature mutations tended to be shared across tumours suggesting that APOBEC mutagenesis was an early event in these tumours. A major limitation of this study was the use of "normal mucosa" as the matched normal. As previously discussed, phenotypically normal mucosa may contain some of the variants shared by the tumours. By using this as the matched normal some of the early events shared by all tumours may have been filtered out as germline variants when they were actually somatic, limiting the ability of the researchers to identify early shared events. Another study used NGS to investigate the relationship between six spatially distinct tumours from a single patient, one of which had an underlying invasive component[174]. This study identified that all tumours were related and the invasive component arose from a shared progenitor prior to tumour outgrowth.

NGS has also been used to investigate metachronous tumours[61,164]. Analysis of paired samples from patients with NMIBC that had stage progression (two patients progressed from stage Ta to stage T1, two patients progressed from stage Ta to stage T2-4) identified that all paired tumours shared a set of mutations at a high cellular presence indicating a common ancestral clone[164]. The ancestral clones contained mutations in well-known cancer driver genes such as *FGFR3, PIK3CA* and *TP53,* suggestive of the ancestral clone being a major driver of the malignant process. Between 14% and 25% of SNVs were shared between paired tumours, and all tumours contained some private mutations. A general increase in the number of SNVs detected in the invasive tumours was identified. Subclone reconstruction identified between 3-4 subclones per tumour. Some subclones present in the Ta tumours were lost in the progressed tumours and some subclonal populations that expanded in the progressed tumours were absent or a very minor subclone in the Ta tumours, demonstrating dynamic clonal evolution in these tumours[164].

Larger numbers of metachronous samples were analysed in the study of Lamy *et al.*[61] who sequenced two or more metachronous tumours from 29 patients. Comparisons between patients with progressive disease and patients with non-progressive disease identified no difference in mutation rates between the two groups but a higher intra-patient variation of the mutation spectrum was observed in patients with progressive disease. To identify if this intra-patient mutational heterogeneity was reflective of intra-tumour heterogeneity that may have been missed by sampling bias, the researchers

sequenced eight separate regions that had been laser microdissected from a single muscle invasive tumour[61]. All regions were evolutionary similar with few private mutations identified. Additionally, a similar intermix of the two main subclones was present in all regions, suggestive of low spatial heterogeneity. Two-thirds of patients with progressive disease had a high level of APOBEC mutagenesis compared to only one-third of non-progressive patients. Subclone analysis identified an ancestral clone present in all tumours from each patient, confirming a monoclonal origin of recurrent tumours. Few subclones were identified, with each tumour containing only 1 or 2 private subclones[61].

Low spatial heterogeneity has been observed in primary tumours in an additional study[63]. Analysis of primary and metastatic lesions from three patients with metastatic bladder cancer showed that whilst heterogeneity in the primary tumours was low, a much higher level of intratumour heterogeneity was observed in the metastatic lesions[63]. Patients with multifocal disease have also been shown to have higher spatial heterogeneity than patients with unifocal disease using multi-region analysis[55]. Multifocal tumours were shown to be clonal in origin and analysis of the surrounding normal urothelium identified the presence of mutations shared by multifocal tumours suggestive of intraepithelial spread or seeding from tumours. Few mutations were identified in the surrounding normal tissue in the two patients with unifocal disease suggesting that the presence of “field disease” is likely linked with multifocality and recurrent tumours[55]. Sequencing of adjacent urothelium may provide targetable mutations in the field disease which could reduce recurrences[55]. Additionally, the mutational burden of the adjacent urothelium could provide a way to predict patients that are likely to recur. More studies are needed to assess this.

Overall, NGS has driven progress in the molecular characterization of bladder cancers, yet the numbers of sequenced NMIBCs are still low. As this disease can be highly diverse in terms of recurrence and progression, many more studies are needed in order to elucidate relationships between genomic characteristics and outcome. The monitoring of recurrences from these patients throughout the disease course is required to identify potential markers of recurrence, whilst studies investigating pre- and post-treatment tumours can provide information on the differences between responders and non-responders and may identify markers of resistance that can help guide treatment options.

1.5 Therapy and bladder cancer

Studies investigating the effect of chemotherapy on tumour evolution have been undertaken in MIBC[62,176] and these have revealed important insights into the evolution of these cancers. Faltas *et al.*[62] performed WES and clonality analysis on 16 matched sets of tumours collected before and after chemotherapy. They identified no significant difference in the number of SNVs between pre- and post-chemotherapy tumours but analysis of the number of shared and unique mutations showed a substantial level of mutational heterogeneity between the samples. On average, only 28.4% of mutations were shared across pre- and post-chemotherapy tumours and this heterogeneity was consistent across both primary-primary and primary-metastatic tumour pairs. Mutations in driver genes were not always shared and instances of convergent evolution were identified, exemplified by the presence of a *TP53* mutation in a lung metastasis which was different to the *TP53* mutation shared by other tumours from the patient. Phylogenetic reconstruction identified early branching evolution in all patients allowing for early metastatic spread. Interestingly, post-chemotherapy tumours demonstrated an increase in clonality which could be a reflection of the selective bottleneck caused by chemotherapy[62,181]. Analysis of CNAs between pre- and post-chemotherapy tumours showed very little intra-patient heterogeneity and tumours from the same patient tended to cluster together suggesting a relatively stable cancer at the CN level[62].

Another study investigating pre- and post-chemotherapy tumours in MIBC focused on characterizing the genetic alterations associated with cisplatin-based chemotherapy[176]. They identified a novel mutational signature in their post-chemotherapy tumours that was enriched in T > A and C > A substitutions. This signature shared features with an experimentally derived signature of cisplatin-induced mutagenesis[176]. Transcription strand bias, consistent with cisplatin crosslinking was also identified. As reported by Faltas *et al.*[62], no significant change in mutation load was identified post-chemotherapy and the CNA landscape between matched pre- and post-chemotherapy tumours was very similar for the majority of tumours. Significant intratumour heterogeneity was identified and the level of heterogeneity, especially post-chemotherapy treatment, was found to predict overall survival in the cohort.

As seen in other cancers[60], these two studies show that chemotherapy plays an important role in the evolution of the genomic landscape of MIBC. Chemotherapy has been demonstrated to be a large driver of tumour evolution through its ability to generate new mutations[176] as well as forming a selective pressure that only the fittest clones can pass. Chemotherapy is therefore likely to play a role in the evolution of

most cancers, including NMIBC. To date, there have been no studies investigating the effect of chemotherapy on tumour evolution in NMIBC.

1.5.1 Treatment of NMIBC

The treatment regime for NMIBC depends on the risk of recurrence and progression for that patient. Following the European Association of Urology (EAU) guidelines[182], tumours considered low-risk are primary, solitary, stage Ta of low grade (G1) tumours less than 3cm in size and with no concurrent carcinoma *in situ* (CIS). High-risk tumours are tumours with any of the following: stage T1, high-grade/G3, CIS or stage Ta G1/2 tumours that are multiple and recurrent and large (must be all 3). Intermediate risk tumours are any tumours that fall in between these two groups. Treatment of low-risk tumours consists of TURBT followed by an immediate single intravesical instillation of chemotherapy and follow up cystoscopy after 3 months[182]. Intermediate risk patients are also treated with a single instillation of chemotherapy, but this is considered incomplete treatment. For intermediate risk patients, an induction cycle of mitomycin C (MMC) is recommended in several guidelines[10,182]. Alternatively, intermediate risk patients can undergo BCG treatment for a year[183]. For high risk tumours BCG therapy should be carried out for 1-3 years whilst for the patients deemed at the highest risk (e.g. multiple and/or large HG/G3 stage T1 tumours, or, recurrent T1 HG/G3 tumours) radical cystectomy may be advised[6]. All patients with NMIBC should be followed up with frequent cystoscopy, the duration of which changes with the different risk groups[182].

1.5.2 Mitomycin C chemotherapy

At our institution the chemotherapeutic agent MMC is used for the treatment of NMIBC patients (Figure 1.5A). As per the guidelines outlined above[182], patients at low-risk will have an immediate instillation of MMC after TURBT whilst patients at an intermediate risk may go on to have an induction cycle, consisting of once-weekly instillations of MMC for 6 weeks.

MMC is a type of anti-tumour antibiotic widely used as a cancer chemotherapy agent. MMC's anti-tumour capacity is believed to come from its ability to crosslink DNA strands[184], inhibiting both DNA replication and translation[185]. It is considered a bio-reductive alkylating agent as it requires enzymatic reduction before it can react with DNA. The initially reduced molecule can alkylate DNA resulting in a monoadduct[186] (Figure 1.5B) or it can undergo a second alkylation step with DNA, resulting in intrastrand (Figure 1.5C) or interstrand (Figure 1.5D) crosslinks[184,187,188].

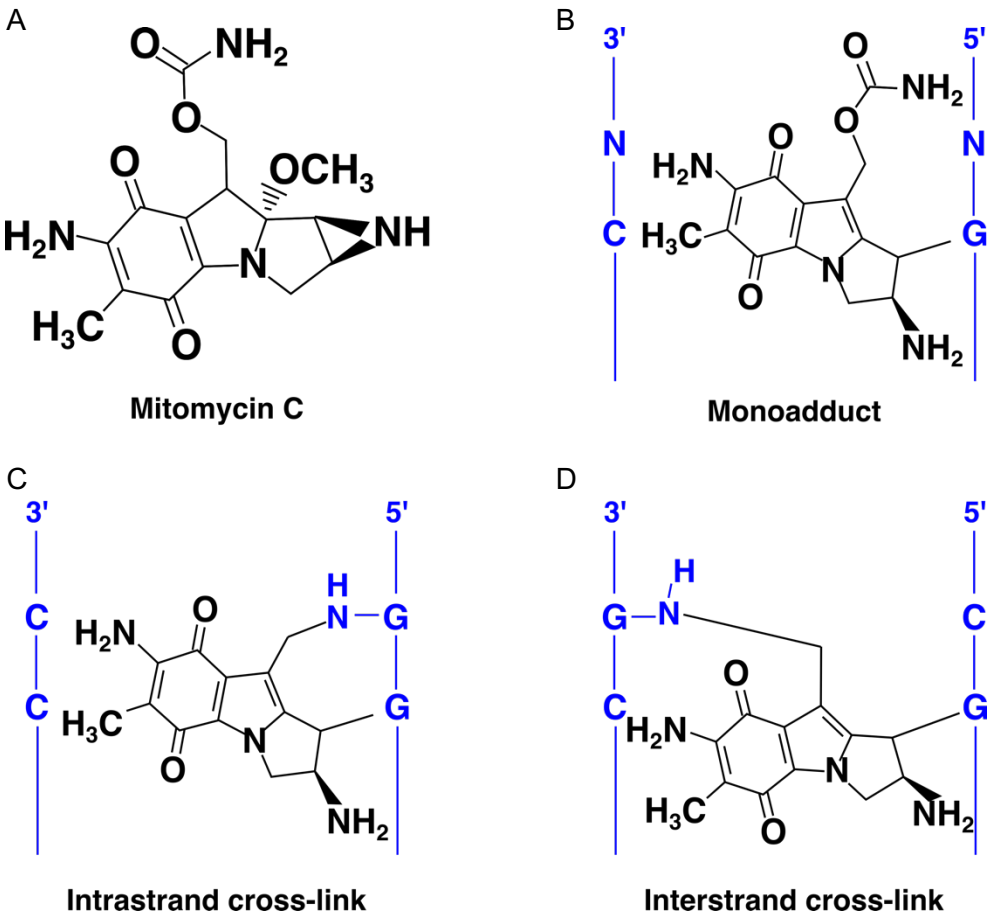


Figure 1.5: The structure and adducts of mitomycin C.
MMC can form a range of DNA adducts. MMC (A) can undergo monofunctional activation and form monoadducts on guanine residues (B), or it can undergo bifunctional activation forming intrastrand (C) or interstrand (D) crosslinks. Figure adapted from Tomasz[189] and Avenaño and Menéndez[190].

MMC has been shown to have sequence specificity[189]. Monoadduct formation can occur at NpG sequences where N is any base. However alkylation is enhanced at 5'-CpG-3' (CpG) sequences by 5-10 times[191,192]. Formation of interstrand crosslinks is absolutely specific for CpG sequences as MMC requires a second guanine base in the correct position for the second round of alkylation, and this can only occur at the guanine base opposite the cytosine of the CpG[192]. Intrastrand crosslinks occur at GpG dinucleotides and occur less frequently than interstrand crosslinks, likely due to the preferential binding of MMC to CpGs for monoadduct formation[187].

It is thought that the cytotoxicity, and subsequent anti-tumour activity, of MMC comes from its ability to crosslink DNA strands[184], which can inhibit both DNA replication and translation. Monoadducts have been shown to inhibit DNA synthesis in a cell-free

system and this could potentially be a cytotoxic event. However, the lesion created by the monoadduct was bypassed at a low rate allowing replication to continue[185], thus representing a potential source of mutation. Despite this, a study using a cell-line system provided evidence to suggest that the interstrand cross-link is the critical cytotoxic lesion[193]. They compared the cytotoxicity of MMC to that of 10-decarbamoyl mitomycin C (DMC), an artificial derivative of MMC that was thought to be monofunctional yet was reported to be more cytotoxic than MMC to certain cell lines. In their analysis they identified that DMC could, surprisingly, induce the formation of interstrand cross-links, albeit at a much-reduced rate to the formation of monoadducts, and this was at a similar rate of cross-link formation to that seen with MMC. Both drugs showed similar cytotoxicity and levels of interstrand crosslinking but varied on the number of monoadducts. Therefore, it was postulated that this correlation between cytotoxicity and cross-links was evidence for the cross-links being the critical cause of cell death[193]. This study focused only on the induction of cell death and did not assess any surviving cells for mutations. It is therefore possible that whilst monoadducts are unlikely to be cytotoxic, they could possibly be mutagenic.

The genotoxicity and cytotoxicity of MMC has been investigated frequently over the years. Early studies in the 1960s used simple cell observations and growth curves to identify a dose-dependent inhibition of proliferation when cells were incubated with MMC. This inhibition persisted after removal of the drug[194,195] and it was noted that the cells continued to grow but did not divide after treatment, resulting in the formation of giant cells. Moving on from simple cell growth observations, Cohen and Shaw[196] used karyotyping to assess the effects of MMC over a range of concentrations. They identified that higher concentrations of MMC destroyed all cells whilst lower concentrations reduced the mitotic rate and induced numerous chromosome breaks and exchanges. Analysis of these breaks and exchanges was seen to be non-random with chromosomes 1, 9 and 16 most affected, a feature also identified by Morad *et al.*[197]. A drawback to the use of karyotyping in these studies was that at this time G-banding was not available and it was impossible to identify many individual chromosomes so they were analysed as groups. Fortunately, chromosomes 1, 9 and 16 were able to be individually identified at this time due to their regions of secondary constriction.

The introduction of banding techniques paved the way for a more accurate assessment of MMC damage using karyotypic techniques[198,199]. Despite the advancement in technology these studies drew similar conclusions to those described previously; the

distribution of break points was seen to be non-random with chromosomes 1, 9 and 16 most commonly affected, particularly in the heterochromatic, peri-centromeric regions.

Induction of micronuclei in cells treated with MMC has been observed[33,200]. Analysis of leukocytes from a single donor using FISH identified an 18-fold higher induction of micronuclei by MMC compared to controls. Chromosome 9 was identified as the biggest contributor of material to the micronuclei with chromosome 1 being the next highest contributor[33]. Hovhannisyan *et al.*[200] also used multi-coloured FISH, using both centromeric and whole-chromosome painting to identify the involvement of chromosomes 3, 4, 6, 7, 9, 16,17,18 and X in MMC-induced micronuclei formation. Material from chromosomes 9 and 16 were contained most often in micronuclei and this occurred more often than expected based on DNA content. However, chromosome 9 was only involved in around one third of all micronuclei yet Fauth *et al.*[33] found that around two thirds of micronuclei involved chromosome 9. This difference could be down to the difference in MMC concentration; Hovhannisyan *et al.* used 0.1μg/ml whilst Fauth *et al.* used 0.5μg/ml, a 5x larger dose.

Whilst these cytogenetic studies gave valuable information on large events occurring in the DNA, they may have missed smaller changes in copy number where insertions or deletions are limited to a small part of the chromosome. They also do not provide any information on the molecular nature of MMC-induced events.

There are a limited number of studies that have specifically investigated the mutational spectrum of MMC. Srikanth *et al.*[201] were amongst the first to use sequencing to identify mutations caused by MMC. They used a shuttle vector system in which plasmid DNA was incubated in reduced MMC, replicated in cells and then used to transform bacteria to increase the material available for assay and sequencing. Mutational hotspots in the target region of the plasmid were identified in GC rich areas, correlating with the specificity of MMC for guanine bases. The most common types of mutations identified were substitutions, accounting for 76% of all mutations. Interestingly, 30% of the base substitutions were tandem substitutions, all of which were at GpG sites. Single base deletions were the second most common event identified.

Maccubbin *et al.*[202] also used a shuttle vector system to assess the mutagenic potential of MMC. This group focused on the mutations caused specifically by MMC-induced monoadduct formation by incubating the shuttle vector with MMC under conditions favouring monofunctional activation. Under these conditions, they frequently observed

single base substitutions corresponding to almost 60% of all mutations. Single base deletions accounted for 24.5% of all mutations and nearly 80% of mutations involving a single base occurred at G:C base pairs, confirming the specificity of the monoadduct. No tandem substitutions were identified, consistent with the monofunctional activation of MMC.

Despite *in vitro* evidence suggesting that MMC can indeed induce substitutions, this was not reflected in an *in vivo* study. Takeiri *et al.*[203] investigated the mutation spectrum of MMC in the bone marrow of mice. They identified many large deletions between 110bp and 8kb long, several of which contained 2-6bp short homologous sequences in their junctions that were eliminated when the DNA joined back together. Contrary to the findings *in vitro, s*ingle base substitutions or deletions were not found at a higher level than in the control. They did, however, identify that tandem base substitutions were only found in the MMC-treated samples. When the group established a cell line from the mice and treated these with MMC they identified large deletions and tandem substitutions consistent with the *in vivo* data. Additionally, they identified single base substitutions, most commonly G:C > T:A transversions, which had not been identified in the *in vivo* data but correlated with the *in vitro* results seen previously[204].

The number of studies investigating MMC action on DNA declined in the early 2000s. Despite the vast improvements in technology in the interim, it was not until recently that next-generation sequencing was employed to look at the MMC-induced DNA mutation spectrum. Tam *et al.*[205] used next-generation whole-genome sequencing to characterise the mutational spectrum of MMC in the model organism *Caenorhabditis elegans*. By using the genetic balancer *hT2* they were able to capture and maintain mutations that would otherwise be lethal, such as mutations in essential genes. As in the study of Takeiri *et al.*[203], they reported that MMC primarily induced deletions, with few SNVs or insertions identified. Analysis of the sequence context of the deletions showed a preference for 5'-CpG-3' dinucleotides, consistent with previous data[185,191,192].

The contrast in findings between the *in vitro* and *in vivo* studies is interesting. It is possible that the large deletions caused by MMC may be missed in the *in vitro* systems of Srikanth and Maccubbin as such deletions could interfere with the packaging of the lambda phage. The large deletions were identified *in vitro* using a different method[204]. Maccubbin *et al.*[202] restricted their analysis to monofunctional activation of MMC. This monofunctional activation could be rare *in vivo* where several enzymes have the

capacity to reduce MMC, thus reducing the observation of mutations caused by this adduct. Interestingly Tam *et al.*[205] restricted their analysis of SNVs to only homozygous mutations which may have limited the discovery of MMC induced substitutions.

Overall, the literature clearly shows that MMC can cause alterations to DNA that lead to mutations that are maintained. It may therefore be possible to identify CNVs, SNVs and tandem substitutions reflective of MMC-induced damage. The cytogenetic studies described above identified chromosomes 1, 9 and 16 as the most affected by MMC treatment, particularly in their areas of secondary constriction. These areas of secondary constriction are actually large areas of gene-poor peri-centromeric heterochromatin found on these 3 chromosomes. These regions have yet to be annotated with sequence and therefore will not have coverage following analysis of next generation sequencing data. This will mean that small changes in these regions, small deletions, SNVs etc., may not be identified. However, CN changes that span beyond these regions could occur, and these have the potential to be identified. Molecular data from the *in vitro* studies suggest that single base substitutions, particularly C > A transversions may be detected, though the *in vivo* data contradicts this. Reduction of the drug may occur differently within a tumour environment compared to within model organisms meaning that induction of SNVs cannot be ruled out. It will be interesting to see if any of these changes are detected in tumours after treatment. Additionally, the sequence specificity of MMC for CpG dinucleotides could have interesting consequences. The abundance of CpG dinucleotides in the human genome is less than expected considering the overall GC content. However, CpG dinucleotides are enriched in exons and promoter regions relative to the rest of the genome, likely due to coding restraints[206]. By specifically targeting these regions MMC has the potential to disrupt gene function or alter gene expression.

1.5.3 Clinical studies of MMC

MMC was first shown to be effective as a bladder cancer treatment in the 1970's[207]. In the late 80's to early 90's, large scale randomised trials were published investigating the efficacy of intravesical chemotherapy for the treatment of bladder cancer[208,209]. Tolley *et al.*[210] compared two chemotherapy regimens after complete transurethral resection: a single immediate instillation or an immediate instillation followed by four further instillations at check-ups within the first year. The control group underwent TURBT only. At their interim study with two years of follow up, all patients who received MMC (single or multiple instillations) experienced reduced rates of recurrence compared to TURBT alone, with those who received multiple installations having a considerably lower recurrence rate compared to a single instillation[210]. After seven

years of follow up a single instillation of MMC at TURBT was shown to have a long-term advantage over TURBT alone, with a significantly decreased recurrence rate. There was a suggestion that 5 installations may offer an advantage over a single installation, but this was not significant[209]. This treatment effect may have been diluted as patients who recurred at 3 months were counted despite not having started the additional instillations[211].

Even with adjuvant treatment schedules, immediate instillation reduces the rate of recurrence compared to delayed instillation[212]. However results are conflicted as a meta-analysis suggested that this was only true if patients were treated for the shorter term of 6 months (9 instillations) rather than a year (11 instillations)[211] whilst Bosschieter *et al.*[212] identified a benefit even for patients receiving up to 15 instillations of MMC. Several meta-analyses have now confirmed the benefit of a single immediate instillation of chemotherapy post TURBT[213-215]. A large meta-analysis by Sylvester *et al.* looked at seven trials that compared TURBT alone with TURBT and an immediate instillation of chemotherapy[214]. Two of these studies used MMC chemotherapy. Overall, an immediate instillation of chemotherapy reduced the odds of recurrence by 39%. Patients with one single tumour gained the most benefit from the instillation as patients with multiple tumours still had a high recurrence rate and further treatment was suggested. Despite the benefits of reduced recurrence rates, MMC treatment has not been shown to affect progression[213].

The use of multiple instillations of chemotherapy has been less investigated with fewer clinical studies available. The study by Tolley *et al.*[209] suggested that maintenance treatment with an additional 4 instillations of MMC over the year may reduce the recurrence rate compared to a single instillation. Bouffioux *et al.*[208] compared a total of four treatment schedules with patients receiving early (on the day of resection) or delayed instillation (7-15 days after resection) and then each of these groups having short-term treatment (instillations given every week for 4 weeks followed by every month for 5 months) or long term treatment (where the monthly treatment was extended by a further 6 months). This study identified that patients having delayed and short-term treatment did the worst. Unfortunately this study did not compare their extended schedule to TURBT or TURBT plus a single instillation of MMC. A study comparing one instillation of MMC a week for 6 weeks compared to 3 instillations a week for two weeks showed a higher response in the two week schedule[216]. However, this was neoadjuvant treatment with chemotherapy given prior to tumour resection after the detection of a recurrence.

BCG therapy has long been known to reduce recurrence in stage Ta and T1 NMIBC[217] and BCG treatment is the standard of care for patients with high risk disease[218]. Many trials investigating the use of a course of MMC treatment for intermediate or high risk tumours were done in comparison to BCG[219-221]. An early study comparing MMC and BCG identified a significant increase in the number of patients disease-free in the BCG arm compared to the MMC arm (49% vs. 34%) however BCG treatment gave more frequent side effects[219]. A limitation of this study was the short follow-up time; the median follow-up was only 39 months. A long-term study with 20 years of follow up identified a significantly reduced recurrence rate for patients treated with BCG compared to MMC[221]. The majority of these patients were of intermediate risk, however the study was limited by a small patient number (89 patients) and a low concentration of MMC (0.2 mg/ml). Another study comparing a 6-week cycle of either BCG or MMC treatment identified no significant difference in the recurrence rates between the two treatments. Three-year recurrence free rates were 65.5% for BCG therapy compared to 68.6% for MMC chemotherapy[220].

Due to the contradictory results from clinical studies, several meta-analyses have been performed. An early meta-analysis by the Cochrane group found that BCG was superior to MMC in reducing recurrences only in high risk patients[222]. No overall significant difference in progression rates was identified between BCG and MMC treated groups in a meta-analysis by Böhle and Bock[223]. However, additional treatment with BCG maintenance therapy was shown to be favorable with a significant reduction in progression compared to MMC treatment. Malmström *et al.*[224] performed an individual patient data meta-analysis looking at the long-term outcome of studies comparing BCG and MMC. Overall, no difference in the risk of recurrence was identified between BCG and MMC. Separating the patients into patients who had BCG maintenance and patients who did not identified that BCG maintenance reduced the risk of recurrence by 32% compared to MMC treatment. However, a lack of BCG maintenance increased the risk of recurrence by 28% compared to MMC treatment. There were no significant differences identified regarding progression or overall survival. However, BCG therapy was associated with higher toxicity, leading to the suggestion from the authors that BCG with maintenance should be the standard of care for high-risk patients whilst MMC, being less toxic, could be considered for intermediate risk patients with failures switched to BCG. As there are still contradictory results about the beneficial effect of BCG over MMC regarding tumour recurrence and progression, an updated Cochrane systematic review comparing the use of BCG and MMC for the treatment of stage Ta and T1 bladder cancer is underway[225].

Overall, BCG therapy with maintenance has become the standard of care for high-risk tumours whilst for intermediate-risk tumours the increased toxicity associated with BCG treatment[219,226] alongside the evidence that MMC treatment can be as efficacious as BCG therapy has resulted in either BCG therapy or chemotherapy being advised for these tumours[182]. A lack of consistency with respect to the duration of treatment within trials has meant that there is debate concerning the optimal schedule for chemotherapeutic intervention[211,220,227]. At our institution patients at an intermediate risk are offered an induction cycle of MMC treatment consisting of 6 weekly instillations. This schedule has been widely used over the years, but there is an awareness that empirically defined habit has defined this schedule rather than biological rationale[216]. Additionally, treatment options for intermediate risk patients may have been affected by the recent worldwide shortage of BCG[228]. This has resulted in an increase in the use of chemotherapeutic regimes with BCG treatment being reserved for managing high risk tumours or for patients that have failed chemotherapy treatment[229].

Interestingly, a recent systematic review by Sylvester *et al*.[213] identified that a single immediate instillation of chemotherapy after resection resulted in an increased risk of death in NMIBC patients. Separation of the cause of death by treatment group (according to the European Organisation for Research and Treatment of Cancer (EORTC) recurrence risk score) suggested that for patients with a recurrence risk score of 5 or greater, more of those who received a single instillation died, and a higher percentage of these deaths were due to malignant disease (bladder cancer or other) compared to TURBT alone. This subgroup (EORTC recurrence risk score of 5 or greater) consisted of patients with multiple tumours, tumours 3cm or greater, and T1 tumours. This data could suggest that chemotherapy is influencing tumour evolution, perhaps selecting for more aggressive disease, in this patient subgroup.

NMIBC is an interesting disease as it is treated with harsh, DNA damaging chemotherapies despite having a favourable prognosis[4]. It is likely that treatment with genotoxic agents, such as MMC, may induce more mutations and this could alter the evolutionary trajectory in certain patient subgroups. Whilst this is accepted in treating later-stage disease when life-span is trying to be increased, should these agents be used in a cancer which has an otherwise good prognosis? To the best of our knowledge, there have been no studies investigating chemotherapy-related genetic changes in NMIBC to date. Therefore, the aim of this project was to identify any possible genetic alterations associated with MMC treatment.

1.6 Project Aims

- To characterise the copy number landscape of tumours from patients with recurrent NMIBC.
- To use copy number alterations and the mutation status of key bladder cancer genes to assess the clonal origins of tumours from the same patient.
- To identify any features selected by mitomycin C chemotherapy and assess the effects of therapy on overall genomic complexity by comparing tumours from patients who did or did-not receive intravesical mitomycin C.
- To investigate the subclonal composition of pre- and post-MMC treatment tumours and infer the temporal sequence of events during tumour evolution.

Chapter 2
Methods

2.1 Sample processing

2.1.1 Ethics and sample collection

Ethical approval for this study was granted by the Local Research Ethics Committee (Leeds East 99/156 and Leeds Multidisciplinary Research Tissue Bank (RTB) (10/H1306/7)). Informed consent was obtained from all patients. Could cup biopsies were collected, embedded in optimal cutting temperature (OCT) compound (VWR International) snap-frozen and stored in liquid nitrogen. The remainder of the sample was embedded in paraffin for diagnostic assessment. Sample grading and staging was carried out by a consultant urological pathologist (JR) using the 1973 WHO[8] and TNM criteria, respectively.

2.1.2 Sample selection and study participants

Sixty-seven tumours from 23 patients with recurrent non-muscle-invasive bladder cancer (NMIBC) were included in the study. These were retrospectively selected from the Leeds RTB using the following criteria:

- Initial presentation of low stage (<T2) non-muscle-invasive disease

- Fresh frozen material available from at least two temporally separated tumours

A search of the Leeds Multidisciplinary Research Tissue Bank (RTB) database for patients with recurrent bladder tumours was carried out in November 2014 and this identified 207 patients with more than one tumour in the tissue bank. Initially the focus of the study was on patients with multiple recurrent tumours of low stage and grade, therefore at this time patients were restricted to those with primarily stage Ta disease. Patients with synchronous events were not excluded from the study. This identified 57 suitable patients with 2-5 tumour events (30 x 2 events, 17 x 3 events and 10 x 4-5 events). The RTB database was checked periodically to identify new potential patients and check for new recurrences from patients already included in the study. Patients with the most recurrences were initially prioritised but due to difficulties extracting sufficient DNA from all tumours from such patients, numbers were bulked by the inclusion of patients with only two tumour events. Tumours from 20 patients were

sectioned for DNA extraction and sufficient DNA was obtained for 11 of these patients (P0198, P0536, P0712, P0717, P0933, P0990, P1175, P1326, P2065, P2104 and P2291) by May 2016. These samples were sufficient to fill a single lane for copy number analysis on the HiSeq3000.

After sequencing of the first batch of patients for copy number analysis was completed the focus of the project changed to investigating genomic alterations associated with a course of MMC chemotherapy. A list of patients who had received a course of MMC chemotherapy treatment between June 2012 and December 2015 at St James's University Hospital was provided by Urologist Mr Sanjay Jain. This was cross-referenced with the list of patients who had provided informed consent for the RTB and 54 patients were identified, of which 22 patients had recurrent tumours. Of those 22 patients, 10 had tumour material available in the tissue bank from both pre- and post-MMC treatment and these were included in the cohort (P0468, P0533, P0926, P1485, P1777, P1870, P2161, P2218, P2329 and P2440). Patient P0960 had previously been sectioned for copy number analysis however this sample was not ready for sequencing in the first batch. Analysis of clinical information for the samples sectioned for the first copy number analysis run identified that this patient had received a course of MMC treatment (before 2012). Patient P0418 was also identified as having received a course of MMC pre 2012 therefore both P0960 and P0418 were included in the second batch of samples. Patients P0536, P0717, P0990 and P1175 had also received a 6 week course of MMC treatment and had tumours from both pre- and post-treatment available. A summary of patient information is given in Table 2.1 and clinical timelines for each patient can be found in Appendix A.

Overall there were 15 males (65%) and 8 females (35%) with a median age at diagnosis of 72 years (range 46-82). The number of tumours ranged from 2 to 5 metachronous tumours per patient. A total of 67 tumours were included in the study: 8 stage Ta grade 1 (G1), 42 stage Ta grade 2 (G2), 7 stage Ta grade 3 (G3), 4 stage T1 G2, 2 stage T1 G3, 1 stage Ta(x) G3, 2 stage T1(x) G3 and 1 possible low-grade urothelial carcinoma in which small, heavily diathermied fragments of tissue made grading difficult. (x) refers to tumours where insufficient sampling of the muscle layer occurred and therefore invasion could not be ruled out. All (x) tumours were from a single patient and the disease history of that patient (P0418) suggests that these tumours were not invasive therefore the patient was included in the study.

Table 2.1: Patient summary.
Basic information about each patient and the histopathological stages and grades of all patients' tumours used in the study are shown. The sample ID reflects the time point at which the sample was taken. Multifocal samples have an additional letter to identify tumours from the same time point. [a]Mtr = months to recurrence. This is the number of months between the first tumour in the study and each subsequent tumour studied. [b]Denotes the cases where the first tumour available for the study was not the primary tumour for that patient. For more information on the timeline of events for each patient see the patient timelines in Appendix A.

Patient ID	Sex	Age at Diagnosis	Sample ID	Stage and Grade	Mtr[a]
P0198	M	70	P0198-S01[b]	TaG2	0
			P0198-S02	T1G2	36
			P0198-S03	TaG2	43
			P0198-S04	TaG2	46
P0418	F	73	P0418-S01	T1xG3	0
			P0418-S02	TaxG3	4
			P0418-S03	T1xG3	7
P0468	M	53	P0468-S01	T1G2	0
			P0468-S03	TaG2	143
			P0468-S05	TaG2	153
P0533	M	72	P0533-S01[b]	TaG2	0
			P0533-S02	TaG2	130
			P0533-S3A	TaG2	137
			P0533-S3B	TaG2	137
			P0533-S04	TaG3	146
P0536	M	79	P0536-S01	TaG2	0
			P0536-S02	TaG2	4
			P0536-S03	TaG2	46
P0712	M	80	P0712-S01	TaG2	0
			P0712-S02	TaG2	29
			P0712-S03	TaG3	46
			P0712-S04	TaG3	120
			P0712-S05	TaG2	129
P0717	F	79	P0717-S01[b]	TaG2	0
			P0717-S02	TaG2	2
			P0717-S03	T1G2	6
			P0717-S04	TaG2	15
			P0717-S05	T1G3	26
P0926	F	62	P0926-S01[b]	TaG1	0
			P0926-S03	TaG3	36
P0933	F	74	P0933-S01	TaG2	0
			P0933-S02	TaG2	12
P0960	F	73	P0960-S01	TaG2	0
			P0960-S03	TaG1	11
			P0960-S04	TaG2	28
P0990	F	72	P0990-S01	TaG2	0
			P0990-S04	TaG2	75
			P0990-S05	T1G2	78
P1175	M	69	P1175-S01	TaG2	0
			P1175-S02	TaG2	13

Patient ID	Sex	Age at Diagnosis	Sample ID	Stage and Grade	Mtr[a]
P1326	M	46	P1326-S01	TaG2	0
			P1326-S02	TaG2	101
P1485	F	73	P1485-S01	TaG2	0
			P1485-S02	TaG1	91
			P1485-S03	TaG2	98
P1777	M	82	P1777-S01	T1cG3	0
			P1777-S02	TaG3	56
P1870	M	70	P1870-S01	TaG3	0
			P1870-S2A	TaG2	43
			P1870-S03	TaG2	50
			P1870-S05	TaG1	75
P2065	M	62	P2065-S01	TaG2	0
			P2065-S02	TaG2	36
P2104	M	71	P2104-S01	TaG2	0
			P2104-S02	TaG2	28
P2161	M	67	P2161-S01	TaG2	0
			P2161-S02	TaG2	26
P2218	M	76	P2218-S1A[b]	TaG2	0
			P2218-S1B[b]	TaG2	0
			P2218-S02	TaG3	10
P2291	M	64	P2291-S01	TaG2	0
			P2291-S02	TaG2	7
P2329	M	74	P2329-S01[b]	TaG2	0
			P2329-S02	TaG1	8
			P2329-S03	TaG2	21
P2440	F	68	P2440-S01[b]	TaG2	0
			P2440-S02	possible low grade UCC	5

2.1.3 Tissue sectioning

Tissue sectioning of samples embedded in OCT compound was performed at -20°C in a Leica cryostat (Leica Microsystems Ltd). A 5 μm section was taken and stained in haematoxylin (Sigma-Aldrich) for 1 min to estimate tumour purity. If purity was estimated to be 70% or greater then tumours were sectioned into sets of 10 slices of 20 μm thickness for nucleic acid extraction and stored at -80°C in micro-centrifuge tubes or at -180°C in cryo-tubes. A further 5 μm section on a glass slide was taken before and after each set of 10 slices for full haematoxylin and eosin (H&E) staining to monitor the purity of the sample throughout. If purity was estimated to be less than 70% then samples were sectioned for laser-capture microdissection (LCM). For LCM 10 μm sections were captured on polyethylene napthalate (PEN) coated slides (Applied Biosystems, part of Thermo Fisher Scientific). Depending on size, 4-16 sections were captured per slide and 4-8 slides filled per sample. A 5 μm section was taken before and after every 1-2 slides for H&E staining.

2.1.4 H&E Staining

2.1.4.1 Full H&E staining

Slides were removed from -80°C storage and allowed to come to room temperature for 5-10 min before being submerged in acetone for 1 min. Slides were then air-dried for 5-10 min before submerging in haematoxylin for 3 min. Slides were then washed under running water for 1 min, submerged in Scotts tap water (Leica Biosystems, part of Leica Microsystems Ltd) for 1 min, washed in running water for a further 1 min, submerged in eosin (Leica Biosystems) for 1.5 min and washed for a further 30 seconds in running water. Slides were then drained onto tissue before being submerged in a series of 100% ethanol washes (1 x 1 min, 3 x 3 min). Slides were drained onto tissue again before a series of 3 x 5 min xylene washes. Slides were left in xylene until they were mounted onto coverslips using DPX Mountant (Sigma-Aldrich). Sections were examined under light microscopy.

2.1.4.2 Staining of PEN slides for LCM

Slides were stained as per the protocol described for full H&E staining (2.1.4.1) with a few modifications; submersion times for PEN slides in haematoxylin and eosin were reduced to 1.5 min and 45 seconds, respectively, and the xylene steps were removed. Instead, after the ethanol washes, slides were allowed to air dry before being stored at room temperature in a slide box containing desiccant. Tumour cells were then captured using LCM within 4 days.

2.1.5 Macrodissection and laser-capture microdissection

Tumours that contained less than 70% tumour cells were sectioned onto slides for dissection. If areas of tumour cells were spatially distinct from the non-tumour cells then samples were macrodissected by cutting and capturing the area of the PEN membrane under the cells of interest using a dissecting microscope. If cells were more intermixed, LCM was performed using the ArcturusXT^{TM} LCM system with CapSure Macro LCM (0211) caps (Applied Biosystems). This utilises a dual laser system to capture the tissue required; the area of interest is drawn around on the screen and this line is used as the UV cutting line. To ensure the tissue of interest is captured onto the caps a gentle infra-red laser is used to melt the film on the cap to the cells of interest. This was set at 30-100 mV. Tumour cells were isolated by one of two methods: drawing directly around the tumour cells and excising them onto the cap or by drawing around the stroma and contaminating normal cells and removing them from the slide. The film containing neoplastic cells was removed from the cap, or the PEN membrane containing the neoplastic cells if the stroma was removed from the slide, and placed in a 1.5 ml micro-centrifuge tube and stored at room temperature in a box containing desiccant for a maximum of three days before DNA extraction.

2.2 DNA Extraction, Quantification and amplification

2.2.1 Tumour DNA Extraction

DNA was extracted using the Gentra PureGene Tissue kit (QIAGEN) according to the manufacturer's instructions. Briefly tissue was lysed overnight by incubation at 55°C in lysis buffer containing proteinase K. Protein was then removed via a modified salt-precipitation method and the DNA precipitated using isopropanol with glycogen added to help recovery. Precipitated DNA was pelleted via centrifugation and washed with 70% ethanol before allowing the pellet to air dry. The pellet was re-suspended in 50 μl of DNA hydration solution.

2.2.2 Blood DNA extraction

DNA was extracted from venous blood using an Illustra Nucleon BACC DNA extraction kit (GE Healthcare Life Sciences) according to the manufacturer's instructions.

2.2.3 DNA quantification using PicoGreen assay

DNA was quantified using the Quant-iT PicoGreen dsDNA assay kit (Invitrogen, part of Thermo Fisher Scientific) with the high-range standard curve protocol. The provided

λDNA standard was diluted in 1 x Tris-EDTA (TE) buffer (10mM Tris-HCl, 1mM EDTA, pH 7.5) to concentrations of 2 ng/ml, 20 ng/ml, 200 ng/ml and 2000 ng/ml whilst samples were prepared by adding 1 μl of sample to 499 μl of 1 x TE. PicoGreen reagent was diluted 200-fold in 1 x TE and 100 μl was added to 100 μl of sample/standard in a black 96 well optical plate (BMG Labtech Ltd) and incubated in the dark for 2-5 min. Fluorescence intensity was measured using a FLUOstar Galaxy fluorescence plate reader (BMG Labtech Ltd) at an excitation wavelength of 480 nm and emission of 520 nm. DNA concentration was determined using the standard curve.

2.2.4 Whole Genome Amplification

To preserve the maximum amount of genomic DNA (gDNA) available for NGS techniques, Whole Genome Amplification (WGA) was performed using the REPLI-g Mini Kit (QIAGEN) according to the manufacturer's instructions. This kit uses the Phi29 polymerase which is highly processive and has strong strand displacement activity[230]. This activity allows the polymerase to copy over the same material several times by extending new primers and displacing other amplified strands in an isothermal reaction[231]. This is known as multiple displacement amplification and generates micrograms of DNA from only 25 nanograms of input DNA. This WGA-DNA was used for the SNaPshot assays detailed below. A brief methodology for WGA follows.

Briefly; samples were diluted to 10 ng/μl (where possible) and 2.5 μl of this was added to 2.5 μl of denaturation buffer and incubated at room temperature for 3 min. Denaturation was stopped by the addition of 5 μl of neutralisation buffer before addition of 40 μl of a master mix consisting of 1 μl REPLI-g Mini DNA polymerase, 29 μl of Mini Reaction Buffer and 10 μl of Nuclease Free Water (NFW). Samples were incubated at 30°C for 16 h before heat inactivation at 65°C for 3 min. A WGA negative control consisting of 2.5 μl of NFW was included in each WGA run to check for contamination. After amplification, samples were aliquoted into 2 x 24 μl concentrated stocks, stored at -20°C and a dilute working aliquot, produced by diluting the WGA DNA 1 in 30 with NFW. The diluted working aliquot was stored at 4°C. The DNA produced was quality checked by PCR amplification using the PCR step of one of the SNaPshot assays (see section 2.4.1).

2.3 Short Tandem Repeat profiling

Short tandem repeat (STR) analysis uses the length of STRs from 16 locations in the genome to make comparisons between samples to see if they come from the same

individual. This was used to: 1) confirm that tumours and bloods were from the same person and 2) where monoclonality could not be determined, confirm that all tumours are from the same individual.

The Promega PowerPlex 16 system (cat# DC6531) was used for STR analysis following the manufacturers protocol with some modifications. Briefly, genomic DNA was diluted to 0.5 ng/μl and 1 μl of this was combined with 1 μl of the PowerPlex 10x primer mix, 5 μl of QIAGEN multiplex PCR master mix (cat# 206143) and 3 μl of NFW to a total volume of 10 μl. Thermal cycling conditions were: an initial denaturation at 95°C for 15 min then 96°C for 1 min, then 10 cycles of: denaturation at 94°C for 30 seconds, annealing at 60°C for 2 min and extension at 70°C for 45 seconds, followed by 22 cycles of: denaturation at 90°C for 30 seconds, annealing at 60°C for 2 min and extension at 70°C for 45 seconds with a final extension at 60°C for 10 min before reactions were held at 4°C. After PCR amplification samples were prepared for sequencing: 0.6 μl of PCR products were combined with 0.5 μl internal lane standard (ILS) 600 and 10 μl HiDi™ formamide (Applied Biosystems). Immediately prior to running, samples were denatured at 95°C for 3 min then snap cooled on ice water. Samples were run on an ABI PRISM 3130xl Genetic Analyser with a 36cm length capillary and POP-7 polymer™ and analysed using the GeneMapper 3.7 software. Repeat lengths at each location for each allele were compared between samples.

2.4 Single gene mutation analysis

2.4.1 SNaPshot assay

The SNaPshot assay combines a PCR amplification step and a primer-extension step for the rapid identification of known hotspot mutations. In bladder cancer hotspot mutations occur in *FGFR3*, *PIK3CA* and the RAS family of genes and also in the *TERT* promoter. SNaPshot assays have been designed to identify 11 hotspot mutations in *FGFR3*[232], 4 hotspot mutations in *PIK3CA*[233], 3 mutations in the *TERT* promoter[156], 7 hotspot mutations in *HRAS* and *KRAS* and 8 in *NRAS*[143]. These four SNaPshot assays were run on WGA-DNA from each of the tumours. The assay workflow and a brief description of the methodology is shown in Figure 2.1.

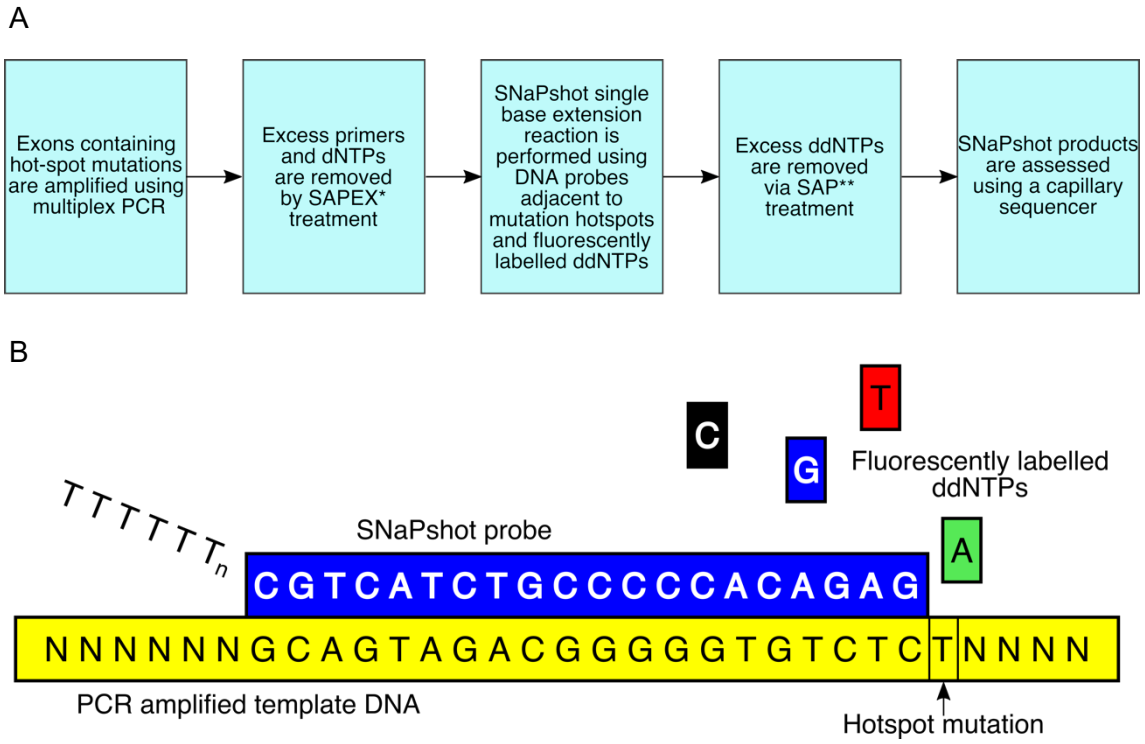


Figure 2.1: Workflow for the SNaPshot assay.
A) Basic outline of the SNaPshot assay workflow. *SAPEX treatment refers to treatment with Shrimp Alkaline Phosphatase and Exonuclease I. ** SAP treatment is treatment with Shrimp Alkaline Phosphatase . B) The SNaPshot single base extension reaction. The SNaPshot probe anneals next to the hotspot mutation and is extended by the incorporation of a single fluorescently labelled dideoxynucleotide triphosphate(ddNTP). Each probe has a tail of T bases at the 5' end that differs in size. This allows for separation of the fragments during sequencing.

2.4.1.1 PCR amplification, agarose gel electrophoresis and SAP/ExoI treatment

For each gene the regions of interest were amplified in a multiplex PCR reaction, except for the *TERT* promoter in which two separate singleplex PCR reactions were run (for a list of regions targeted in each assay and the primers used see Table 2.2). 20 ng of gDNA or 2 μl of diluted WGA product was used for amplification in a 15 μl reaction containing 1 x PCR buffer, 1.5 mM $MgCl_2$, 0.17 mM dNTPs, 5% glycerol and 1 unit of GoTaq DNA polymerase (Promega) with primer concentrations as detailed in Table 2.2. 2 μl water was used as a PCR negative control. Thermal cycling conditions were: an initial denaturation at 95°C for 5 min, followed by 35 cycles of: denaturation at 95°C for 45 seconds, annealing for 45 seconds (60°C for *FGFR3* and *PIK3CA,* 65°C for *TERT* and 55°C for RAS), and extension at 72°C for 45 seconds, followed by a final extension at 72°C for 10 min. The number of cycles was increased to 40 for WGA material.

PCR products were checked for yield and quality using agarose gel electrophoresis. 3 μl of PCR product was combined with 6 μl of water and 1 μl of 10 x DNA loading dye (40% w/v sucrose, 0.25% w/v bromophenol blue) and loaded onto a 2% agarose gel cast in 1 x Tris-Borate-EDTA (TBE) buffer (0.089M Tris, 0.089M Boric Acid, 0.002M EDTA, pH 8.3) (Severn Biotech Ltd) and containing 0.7 μg/ml ethidium bromide, alongside a 100 bp DNA ladder (New England Biolabs (NEB)). Samples were electrophoresed in 1 x TBE for 1 h at 90V. Gels were visualised using a ChemiDoc XRS system (Bio-Rad Laboratories Ltd) and imaged using Image Lab software (Bio-Rad Laboratories Ltd., Version 5.2.1). PCR product sizes were estimated by comparison to the 100 bp DNA ladder.

The remaining 12 μl of PCR product was incubated at 37°C for 60 min with 3 units of shrimp alkaline phosphatase (SAP) (GE Healthcare Life Sciences) and 2 units of exonuclease I (ExoI) (Applied Biosystems) to remove excess dNTP's and primers, respectively. SAP and ExoI were subsequently heat inactivated by incubation at 72°C for 15 min.

Table 2.2: Regions of interest and primer sequences for SNaPshot assays.

Gene	Regions of interest	Strand	Sequence (5'->3')	Concentration		PCR product sizes (bp)
				In primer mix (μM)	In PCR reaction (μM)	
FGFR3	Exon 7	Forward	AGTGGCGGTGGTGGTGAGGGAG	18	1.2	115
		Reverse	GCACCGCCGTCTGGTTGG	18	1.2	
	Exon 10	Forward	CAACGCCCATGTCTTTGCAG	7.5	0.5	135
		Reverse	AGGCGGCAGAGCGTCACAG	7.5	0.5	
	Exon 15	Forward	GACCGAGGACAACGTGATG	10	0.67	161
		Reverse	GTGTGGGAAGGCGGTGTT G	10	0.67	
PIK3CA	Exon 9	Forward	AGTAACAGACTAGCTAGAGA	10	0.67	138
		Reverse	ATTTTAGCACTTACCTGTGAC	10	0.67	
	Exon 20	Forward	GACCCTAGCCTTAGATAAAAC	10	0.67	109
		Reverse	GTGGAAGATCCAATCCATTT	10	0.67	
RAS	*HRAS* exon 1	Forward	CAGGAGACCCTGTAGGAGG	9	0.6	139
		Reverse	TCGTCCACAAAATGGTTCTG	9	0.6	
	HRAS exon 2	Forward	GGAGACGTGCCTGTTGGA	5	0.33	140
		Reverse	GGTGGATGTCCTCAAAAGAC	5	0.33	
	KRAS exon 1	Forward	GGCCTGCTGAAAATGACTG	5	0.33	163
		Reverse	GGTCCTGCACCAGTAATATG	5	0.33	
	KRAS exon 2	Forward	CCAGACTGTGTTTCTCCCTT	5	0.33	155
		Reverse	CACAAAGAAAGCCCTCCCCA	5	0.33	
	NRAS exon 1	Forward	GGTGTGAAATGACTGAGTAC	5	0.33	128
		Reverse	GGGCCTCACCTCTATGGTG	5	0.33	
	NRAS exon 2	Forward	GGTGAAACCTGTTTGTTGGA	5	0.33	103
		Reverse	ATACACAGAGGAAGCCTTCG	5	0.33	
TERT promoter	Position -57bp	Forward	AGCACCTCGCGGTAGTGG	10	0.67	175
		Reverse	AGCCCCTCCCCTTCCTTT	10	0.67	
	Position -124/-146bp	Forward	CAGCGCTGCCTGAAACTC	10	0.67	163
		Reverse	GTCCTGCCCCTTCACCTT	10	0.67	

2.4.1.2 SNaPshot single base extension reactions

SNaPshot single base extension reactions were performed using a SNaPshot Multiplex Kit (Applied Biosystems) and probes designed as described previously[156,232,233]. Reactions were performed in a total volume of 9 μl and contained 2.5 μl of SNaPshot Ready Mix (Applied Biosystems), 2 μl of 5x BigDye sequencing buffer v1.1 (Applied Biosystems), 1 μl of SNaPshot probe mix (Table 2.3), 2.5 μl NFW and 1 μl of SAP/ExoI treated PCR product. Thermal cycling conditions for extension reactions were 35 cycles of denaturation at 96°C for 10 seconds and annealing/extension at 58.5°C for 40 s. Extension products were treated with 1 unit SAP and incubated at 37°C for 60 min then 72°C for 15 min. Products were diluted 1/10 with NFW. 1 μl of diluted product was added to 10 μl formamide mix (9.8 μl HiDi™ formamide, 0.2 μl of 1/10 diluted GeneScan-120LIZ size standard (Applied Biosystems)) and denatured at 100°C for 5 min before snap-cooling on ice. Products were run on an ABI PRISM 3130xl Genetic Analyser with a 36 cm length capillary and POP-7 polymer™ and analysed using GeneMapper 3.7 software (Applied Biosystems). Genotypes were scored manually based on peak colour and position.

Table 2.3: SNaPshot probes for the detection of *FGFR3*, *PIK3CA*, RAS gene and *TERT* promoter mutations.

For the RAS probe sets position refers to the position of the nucleotide in the cDNA with numbering starting from the first base of the ATG start codon. For the *TERT* promoter, the number refers to the distance from the ATG translation start site in bp.

Gene	Probe	Sequence (5'->3')	Size (bp)	Mutation	Concentration in mix (μM)
FGFR3	R248C	T46 CGTCATCTGCCCCCACAGAG	66	C>T	2
	S249C	T36 TCTGCCCCCACAGAGCGCT	55	C>G	1.2
	G372C	T29 GGTGGAGGCTGACGAGGCG	48	G>T	0.4
	Y375C	T43 ACGAGGCGGGCAGTGTGT	61	A>G	1.2
	A393E	T34 CCTGTTCATCCTGGTGGTGG	54	C>A	2.4
	K652M/T	T20 CACAACCTCGACTACTACAAGA	42	A>T/C	0.8
	K652Q/E	T50 GCACAACCTCGACTACTACAAG	72	A>G/C	0.6
	S373C	T19 GAGGATGCCTGCATACACAC	39	T>A	0.5
	G382R	T56 GAACAGGAAGAAGCCCACCC	76	C>T	1.2
PIK3CA	E542K	T19 TACACGAGATCCTCTCTCT	38	G>A	0.8
	E545G	T29 TCCTCTCTCTGAAATCACTG	49	A>G	2.3
	E545K	T34 ATCCTCTCTCTGAAATCACT	54	G>A	1.5
	H1047R/L	T46 TGAAACAAATGAATGATGCAC	67	A>G/T	1.5
RAS set 1	HRAS pos.34	T17 CTGGTGGTGGTGGGCGCC	35	G>C/T/A	5
	HRAS pos.182	T18 GCATGGCGCTGTACTCCTCC	38	T>G/C/A	1.5
	KRAS pos.34	T25 GGCACTCTTGCCTACGCCAC	45	C>G/A/T	5
	HRAS pos.35	T31 CGCACTCTTGCCCACACCG	50	C>G/A/T	7
	NRAS pos.182	T33 GACATACTGGATACAGCTGGAC	55	A>G/C/T	5
	KRAS pos.181	T41 CTCATTGCACTGTACTCCTCTT	63	G>T/C	2
	HRAS pos.181	T46 CATCCTGGATACCGCCGGC	65	C>A/G	7
	KRAS pos.35	T49 AACTTGTGGTAGTTGGAGCTG	70	G>C/T/A	2
	HRAS pos 37	T55 CAGCGCACTCTTGCCCACAC	75	C>G/A/T	7
	NRAS pos.34	T62 CTGGTGGTGGTTGGAGCA	80	G>C/T/A	2
RAS set 2	KRAS pos.37	T15 CAAGGCACTCTTGCCTACGC	35	C>G/A/T	7
	NRAS pos.181	T18 CTCATGGCACTGTACTCTTCTT	40	G>T/C	2
	NRAS pos.37	T26 GGTGGTGGTTGGAGCAGGT	45	G>C/T/A	2
	KRAS pos.183	T29 CCTCATTGCACTGTACTCCTC	50	T>A/G	7
	KRAS pos.38	T33 CTTGTGGTAGTTGGAGCTGGTG	55	G>C/T/A	2
	NRAS pos.183	T38 CTCTCATGGCACTGTACTCTTC	60	T>G/C/A	5
	NRAS pos.38	T44 GTCAGTGCGCTTTTCCCAACA	65	C>G/A/T	5
	NRAS pos.180	T49 GGACATACTGGATACAGCTGG	70	A>T	3
	KRAS pos.182	T56 ATTCTCGACACAGCAGGTC	75	A>T/C/G	5
	HRAS pos.183	T62 CCTGGATACCGCCGGCCA	80	G>C/T/A	3
	HRAS pos.38	T64 GTCAGCGCACTCTTGCCCACA	85	C>A/T	5
	NRAS pos.35	T71 CTGGTGGTGGTTGGAGCAG	90	G>C/A/T	5
TERT Promoter	-57 probe	T29 TCCTCGCGGCGCGAGTTTC	48	A>C	2
	-124 probe	T19 GGGGCTGGGAGGGCCCGGA	38	G>A/T	1.5
	-146 probe	T34 GGCTGGGCCGGGGACCCGG	53	G>A	3

2.4.2 PCR/Sanger sequencing analysis of RAS gene mutations

The RAS SNaPshot is heavily multiplexed and can produce less clear SNaPshot profiles than observed with the other genes. Therefore, RAS gene mutations identified by SNaPshot analysis were confirmed by Sanger sequencing. Briefly the exon of the RAS gene with a suspected mutation was PCR amplified using 10-20 ng of gDNA in a 25 μl reaction containing 1 x PCR Gold buffer (Applied Biosystems), 2.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 μM of each forward and reverse primer for the relevant exon, and 1 unit of AmpliTaq Gold DNA polymerase (for primer details see Table 2.4). Thermal cycling conditions were: 95°C for 5 min, followed by 35 or 40 cycles of 95°C for 30 seconds annealing for 30 seconds (at 55°C for *HRAS* exon 2, *NRAS*, *KRAS* or 60°C for *HRAS* exon 1), and 72°C for 30 seconds followed by a final extension at 72°C for 10 min. 5 μl of PCR product was analysed on a 2% agarose-TBE gel to check PCR product sizes and yield. 2.5 μl of PCR product was then treated with 1 μl of ExoProStar (GE Healthcare Life Sciences) and incubated at 37°C for 15 min to remove excess dNTPs and primers. Reactions were heat inactivated at 80°C for 15 min.

Sanger sequencing of PCR products was performed using the same primers used for PCR. Reactions contained 0.16 μM primer (forward or reverse), 1 μl of ExoProStar treated template, 1 x BigDye sequencing buffer and 0.25 μl BigDye terminator ready reaction mix v1.1 (Life Technologies, part of Thermo Fisher Scientific) in a final volume made up to 10 μl with NFW. The sequencing thermal cycler profile consisted of: 96°C for 1 min followed by 25 cycles of: 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 min. Sequencing reactions were then precipitated by addition of 1 μl of 3 M sodium acetate and 25 μl of ice cold 95% ethanol followed by incubation at room temperature for 30 min. 96 well plates containing precipitated DNAs were then centrifuged at 2250 x g for 30 min to pellet the DNA. The supernatant was removed by inverting the plate onto absorbent paper and centrifuging at 180 x g for 1 min. The pellet was washed by addition of 70 μl of 70% ethanol and the plate was centrifuged at 1650 x g at 8°C for 15 min. The supernatant was removed by inverting the plate onto absorbent paper and centrifuging at 180 x g for 1 min. The pellet was then dried by heating the plate to 95°C for 1 min in a PCR machine and resuspended in 15 μl of HiDi™ formamide. Immediately before running, samples were heated to 95°C for 1 min then snap-cooled on ice. Samples were run on an ABI PRISM 3130xl Genetic Analyser with a 36cm length capillary and POP-7 polymer™. Sequence traces were evaluated manually for mutations.

Table 2.4: Primers used for PCR amplification and sequencing of RAS genes.

Target Exon	Primer	Sequence 5'-> 3'	PCR product size (bp)
HRAS exon 1	H1_F	CAGGAGACCCTGTAGGAGGA	139bp
	H1_R	TCGTCCACAAAATGGTTCTG	
HRAS exon 2	H2_F	TCCTGCAGGATTCCTACCGG	194bp
	H2_R	GGTTCACCTGTACTGGTGGA	
KRAS exon 1	K1 _Brose F	GTGTGACATGTTCTAATATAGTCA	214bp
	K1_R	GTCCTGCACCAGTAA	
KRAS exon 2	K2_2F	GTGCACTGTAATAATCCAGAC	220bp
	K2 _2R	CCTATAATGGTGAATATCTTC	
NRAS exon 1	N1_2F	TAAAGTACTGTAGATGTGGC	240bp
	N1_2R	AAGATGATCCGACAAGTGAG	
NRAS exon 2	N2_F	GGTGAAACCTGTTTGTTGGA	197bp
	N2_2R	TTCAGAACACAAAGATCATC	

2.5 Copy number analysis

2.5.1 Library preparation

The NEBNext® DNA Library Prep Master Mix Set for Illumina® (NEB) was used for library preparation procedures.

2.5.1.1 Shearing

500 ng of gDNA was volume-adjusted to 250 μl with 1 x TE buffer, pH8 low EDTA (0.1 mM) and transferred into a Covaris tube compatible with the Covaris S2 system (Covaris Ltd). Samples were sheared by sonication using the Covaris S2 system by running 38 cycles with batch settings as detailed in Table 2.5.

Table 2.5: Covaris settings
The 500 cycles per burst (cpb) and 1000 cpb settings were combined in a batch which was run for 38 cycles to shear the DNA.

Cycles per burst	Duty Cycle	Intensity
1000	19.9%	9.9
500	15%	8

After shearing the DNA was cleaned up with a MinElute kit (QIAGEN) and eluted in 11 μl of Buffer EB. 2 μl was analysed using the Agilent TapeStation High Sensitivity D1000 kit (Agilent) with an Agilent 2200 TapeStation System (Agilent) to check that the DNA had sheared appropriately, with the peak size being between 175 and 205 bp.

2.5.1.2 End-repair of fragmented DNA

After shearing, the ends of the fragmented DNA were repaired. In a 96 well optical plate the remaining 9 μl of sheared DNA was end-repaired in a 50 μl reaction containing 1 x NEBNext end repair reaction buffer and 2.5 μl of NEBNext end repair enzyme mix. The reaction was incubated at 20°C for 30 min in a PCR machine without the use of a heated lid. The end-repaired sample was purified using a QIAquick PCR Purification kit (QIAGEN) and eluted in 21 μl of Buffer EB.

2.5.1.3 dA-tailing of end-repaired DNA

End-repaired DNA was dA-tailed to enable ligation of the adaptor. For each sample, the end-repaired DNA (21 μl) was added to a well of a 96 well plate containing 2.5 μl of NEBNext dA-tailing reaction buffer (10x) and 1.5 μl of Klenow (3'>5' exo-) and incubated at 37°C for 30 min in a PCR machine without the use of a heated lid. dA-tailed DNA was purified using a MinElute kit and eluted in 12.5 μl of Buffer EB.

2.5.1.4 Adaptor ligation of dA-tailed DNA

The NEBNext Adaptor was ligated to dA-tailed DNA in a 25 μl reaction containing the dA-tailed DNA (12.5 μl), 1x Quick ligation reaction buffer, 2.5 μl of the NEBNext Adaptor and 2.5 μl of Quick T4 ligase. This was incubated at 20°C for 15 min in a PCR machine without the use of a heated lid then 3 μl of USER enzyme was added and the reaction mix further incubated in a PCR machine at 37°C for 15 min without the use of a heated lid. After incubation, 25 μl of Buffer EB was added to give a total volume of 50 μl and size selection was performed immediately.

2.5.1.5 Size selection of adaptor ligated DNA

Adaptor ligated DNA was size selected to remove any free adaptors and select fragments of the correct size for the PCR enrichment step. 40 μl of suspended AMPure XP beads (Beckman Coulter) was added to 50 μl of DNA solution from the adapter ligation step and mixing performed by pipetting. Samples were incubated at room temperature (r.t.) for 5 min then placed in a magnetic stand for 5 min to concentrate the magnetic beads at the back of the tube. The supernatant was transferred to a new tube

and the beads, containing the larger fragments, were discarded. 10 μl of beads were added to the supernatant, mixed by pipetting, and incubated for 10 min at r.t.. Samples were placed in the magnetic stand for 5 min to concentrate the beads and the supernatant discarded. Beads were then washed with 2 x 200 μl of 80% ethanol (whilst in the magnetic stand) and the supernatant discarded before allowing the beads to air-dry for 10 min. Tubes were removed from the rack, the beads were resuspended in 22 μl of Buffer EB and incubated at r.t. for 3 min to elute the DNA. Beads were concentrated for 3 min using the magnetic stand and the supernatant transferred to a new tube and stored for PCR enrichment.

2.5.1.6 PCR enrichment of size-selected adaptor ligated DNA

Size-selected DNA was tagged with an index primer and enriched using PCR in 25 μl reactions containing either 10 μl or 8.15 μl of size-selected adaptor ligated DNA (depending on the kit version), 1x NEB High Fidelity PCR master mix, 1.25 μl or 3.1 μl of Universal PCR primer (at 25 μM or 10 μM respectively, depending on the kit version) and 1.25 μl of indexing primer (25 μM) and carried out in 96 well optical plates. Indexing primers used for each sample were recorded. The library was amplified using the following thermal cycler conditions: 1 cycle at 98°C for 30 s, then 12 cycles of 10 s at 98°C, 30 s at 65°C and 30 s at 72°C, followed by a final extension for 5 min at 72°C.

2.5.1.7 Purification of PCR products

PCR products were purified using AMPure XP beads. PCR products (25 μl) were transferred into a 1.5 ml microcentrifuge tube containing 25 μl of NFW. 62.5 μl of suspended AMPure XP beads were added, mixed by pipetting and incubated at r.t. for 5 min with intermittent mixing. Beads were separated from the supernatant using a magnetic stand for 5 min. Once clear, the supernatant was discarded and the beads were washed with 2 x 200 μl of 70% ethanol (whilst in the magnetic stand) before allowing the beads to air-dry for 10 min. Once dry, the tubes were removed from the rack and the beads were resuspended in 42 μl of Buffer EB and incubated at r.t. for 3 min. Beads were then concentrated using a magnetic stand for 3 min and the supernatant transferred to a new tube for storage and QC analysis.

The indexed library was checked for quality using the Agilent TapeStation High Sensitivity D1000 kit. 2 μl of library was analysed to ensure that the fragment size was within the range of 220 and 260 bp and that there were no significant primer dimer peaks (significant peaks were designated as those that were >10% of the total

integrated area). Libraries passing these quality criteria were stored at -20°C ready for pooling. For sequencing, 30 ng of each library were pooled in batches of 50 samples.

2.5.2 Data processing

Pooled libraries were run on the Illumina HiSeq3000 at the University of Leeds with 150 bp paired end sequencing. Samples were sequenced to an average raw coverage depth of 0.7x. Quality control (QC) checks were performed on the raw FASTQ files using FastQC 0.10.0[234]. Reads were quality trimmed with a threshold of 20 to remove low quality read tails, adaptor sequences were removed and reads shorter than 20 bp were discarded with cutadapt 1.14. These trimmed reads were then quality checked again by FastQC before alignment with BWA 0.7.15[235] to hg38 to create the sequence alignment map (SAM) file. Samtools 1.5[236] was used to convert the SAM to a binary alignment map (BAM.) Aligned reads were sorted and PCR duplicates were marked using Picard 2.10.2 prior to indexing with samtools. Indels were identified and local realignment performed using the GATK v3.7 RealignerTargetCreator and IndelRealigner. Post-processing was performed with samtools to remove low-quality reads such as those with a mapping quality <20, those marked as PCR duplicates or reads that were unmapped.

A pseudo-CGH algorithm, ngCGH (https://github.com/seandavi/ngCGH), was used to compare read counts in the tumour and matched normal with a window size of 1000 reads. This algorithm requires two BAM files, one from the tumour and one from the matched normal. The normal sample is segmented into blocks of 1000 reads - this is known as the genomic window. The number of reads within each of these genomic windows in the matched tumour is then quantified. A ratio is made between the number of reads in the normal and tumour for each genomic window and this is log2 transformed. The resulting dataset is then median centred to get the final values and the data is converted to a format that can be read by the Nexus Copy Number software (BioDiscovery).

The Nexus Copy Number software package was used for GC correction and segmentation using the FASST2 Segmentation algorithm, which is a hidden Markov model (HMM) based approach. The segmentation significance threshold was set at 1.0E-5, with a requirement for a minimum of 2 probes per segment and a maximum probe spacing of 1000 Kbp between adjacent probes. Single copy gain and single copy loss log2 ratio thresholds were set at 0.25 and -0.25 respectively. Log2 ratio thresholds for two or more copy gains and homozygous losses were set at 1.2 and -1.2 respectively. For some samples, the log2 ratio thresholds for single copy gains and

losses were modified to 0.3 and -0.3 respectively in order to compensate for a higher level of “noise” in the data from these samples and to reduce the risk of erroneous segmentation. The patients for whom the boundaries were modified were: P1175, P1485, P1777, P2161, P2218, P2329 and P2440. Initially no filtering or blacklisting was applied to the data and data was manually curated to remove centromeric events or to insert any shared missed breakpoints. A list of all the regions removed from the tumours due to them being centromeric or telomeric and unreliable was compiled into a blacklist (**Error! Reference source not found.**) that was later applied to all samples for image acquisition.

2.5.3 Data analysis

2.5.3.1 Genome-wide frequency plots and statistical comparisons

The Nexus Copy Number software package was used for generating genome-wide frequency plots and statistical comparisons. The comparisons function was used to identify significant differences in the CN data for specified sample subgroups. This takes the two groups and subtracts the profile of one from the other then uses a Fishers exact test with a *p*-value cut-off of 0.05 and a differential threshold of 25% to identify regions with significant differences. Comparisons were made with respect to stage, grade, tumour resection timepoint and mitomycin C course treatment.

2.5.3.2 Construction of phylogenetic trees from CN data

Analysis of shared and unique copy number alterations (CNAs) in tumours from the same patients can be used to reconstruct the relationship between individual tumours, infer clonality and also order specific events. To analyse the relationship between tumours from the same patient in this study the TuMult algorithim was applied[35]. TuMult uses the simple reasoning that two tumours descended from the same cancerous cell will have a number of genetic events in common. Clones that separated early on will have fewer shared genetic events than clones that separated later. By analysing common and unique chromosome breakpoints, TuMult can reconstruct the sequence of chromosomal events that gave rise to each tumour.

The TuMult algorithm is run through the R programming language and was designed for use on arrayGCH and SNP array data. These methods use a series of probes to assess the CN status of a sample and this probe level data is required as an input to the TuMult algorithm. The NGS data had to undergo a series of manipulations to create the probe-style input required by the TuMult algorithm. This was done within R using basic R manipulations and the package “bedr” which provides a R-wraparound for the

command line tool, bedtools. Briefly, the output of the ngCGH script was used to create the probe backbone required for analysis (see Appendix C for an example of this file format). As described above, ngCGH works using a windowing methodology set to 1000 reads. Each window can be used as the equivalent of a probe making it compatible for use with TuMult. Each window (probe) was given a unique identifier and labelled with the cytoband region of the probe. Any probes that corresponded to blacklisted regions were removed to create the final probe table with column headers of: Name (unique identifier), Chromosome, StartPosition, EndPosition, StartCytoband, EndCytoband (see Appendix C for an example of this file format). The segmented CN profiles for each tumour from a patient were then joined to the probe table and probes with no CN changes were set to a value of 0. Segmented data was used as the raw output from the ngCGH contains noise and GC bias that is filtered and corrected for during segmentation. A discretized CN value was then given to every probe encoded as follows: 0 = no CN alteration (log_2 ratio between -0.25 and +0.25), +1 = gain (log_2 value >0.25, <1.2), +2 = high level gain (log_2 value ≥1.2), -1 = loss (log_2 value >-1.2, <-0.25) and -2 = high level loss (log_2 value ≤ -1.2). A “profiles file” was then created by sub-setting the table to include only the exact log ratio (Sample1.value) and the discretized CN status (Sample1.status) for each tumour from the patient (see Appendix C for an example of this file format). The lines in the profiles file and probe table are kept in the same order. A reference data set was created which contains all the probes with a discretized reference value. For the NGS data this value was set to 0 because germ-line CNVs are effectively normalized using the ngCGH script (see Appendix C for an example of this file format).

2.5.3.3 Clustering

Clustering was performed using discretized CN data for the samples in this study along with data for 141 low-stage low-grade NMIBC tumours from Hurst *et al.*[82]. One-way unsupervised hierarchical cluster analysis was performed using Partek Genomics Suite 6.6 (Partek Inc.). This requires the CNAs to be aligned to bacterial artificial chromosome (BAC) clone IDs and locations. The BAC clone locations were converted from hg19 to hg38 using the Lift-Over tool within the Galaxy web environment on the public server at usegalaxy.org[237]. The segmented CN profiles for each tumour were joined to the BAC clones using the bedr package within the R software environment. A discretized CN value was given to each probe as per section 2.5.3.2. All discretized data was joined together and the resulting file was used for clustering. Samples were clustered using one-way unsupervised hierarchical cluster analysis with Euclidean distance and the Ward method of linkage. RJA generated the input files and clustering was performed by Dr Carolyn Hurst.

2.6 Whole-exome and targeted sequencing

Whole exome sequencing (WES) was performed on matched pre- and post-MMC tumours from 8 patients with peripheral blood used as a germline control. Patients were selected for WES from the copy number cohort using the following criteria:

- No BCG treatment between pre- and post-MMC tumours
- No more than 1.5 years between pre- and post-MMC tumours
- More than 200ng of gDNA available for sequencing

Patients who had previously undergone BCG therapy prior to the pre-MMC tumour or proceeded to BCG therapy after the post chemotherapy tumour were not excluded.

Targeted sequencing of a panel of 140 genes that have previously been identified to be altered in bladder cancer[82,100,164,170-172] was performed on any additional tumours from the 8 patients in the WES cohort (for genes in targeted sequencing panel see Appendix D). Tumours from patients for whom a monoclonal origin could not be identified using CNAs and hotspot mutation analysis also underwent targeted sequencing.

2.6.1 Library preparation

Libraries were generated using 200 ng of DNA and enriched for exonic regions using either the SureSelectXT Human All Exon V6 (WES) or the SureSelectXT targeted sequencing panel described above, according to the manufacturer's instructions (see Figure 2.2 for an outline of the library preparation process). Six sample libraries were pooled per lane, with tumours and matched bloods run in the same library pool.

2.6.2 Next generation sequencing and variant calling

FASTQ processing and variant calling on WES samples was performed by our in-house bioinformatician Dr Guo Cheng. An outline of the process can be seen in Figure 2.3. Pooled libraries were run on the Illumina HiSeq3000 at the University of Leeds with 150 bp paired end sequencing. Quality control checks were performed on the raw FASTQ files using FastQC. Reads were quality trimmed with a threshold of 20 to remove low quality read tails, adaptor sequences were removed and reads shorter than 20 bp were discarded using cutadapt 1.14. These trimmed reads were then quality checked again by FastQC before alignment with BWA 0.7.15 to hg38 to create the SAM file. Samtools 1.5 was used to convert the SAM to a BAM. Aligned reads were sorted and PCR duplicates were marked using Picard 2.10.2 prior to indexing with samtools. Indels were identified and local realignment performed using the GATK v3.7

RealignerTargetCreator and IndelRealigner. Post-processing was performed with samtools to remove low-quality reads such as those with a mapping quality <20 or reads that were unmapped.

SureSelect targets are provided in hg19. These were converted to hg38 coordinates using UCSC liftOver and used as the target files for variant calling. Variant calling was performed using Mutect2 from GATK, Strelka2[238] and EBCall[239] with additional calling performed by VarScan2[240] and Muse[241] for exome samples from P418 and P2161 initially to identify the best combination of callers. Variant calling was performed using the default parameters for all callers.

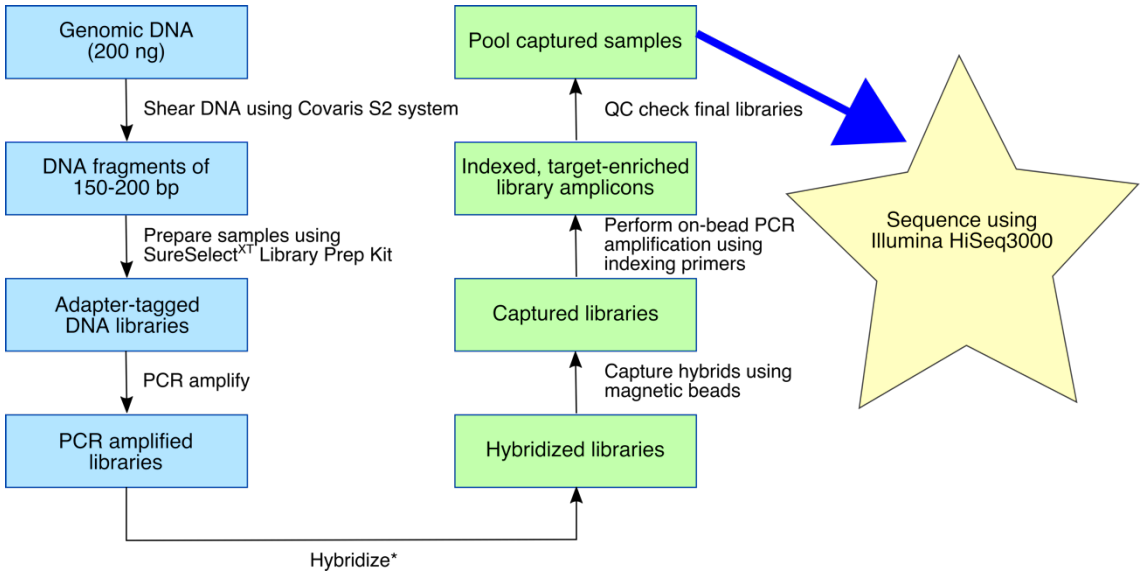


Figure 2.2: Workflow for WES and targeted sequencing.
Practical steps in the generation of NGS libraries are presented. Blue boxes represent steps prior to hybridization, steps post hybridization are in green and the final sequencing step is in yellow. *Hybridization step is performed using either the whole-exome target capture baits or the bladder cancer gene panel targeted capture baits.

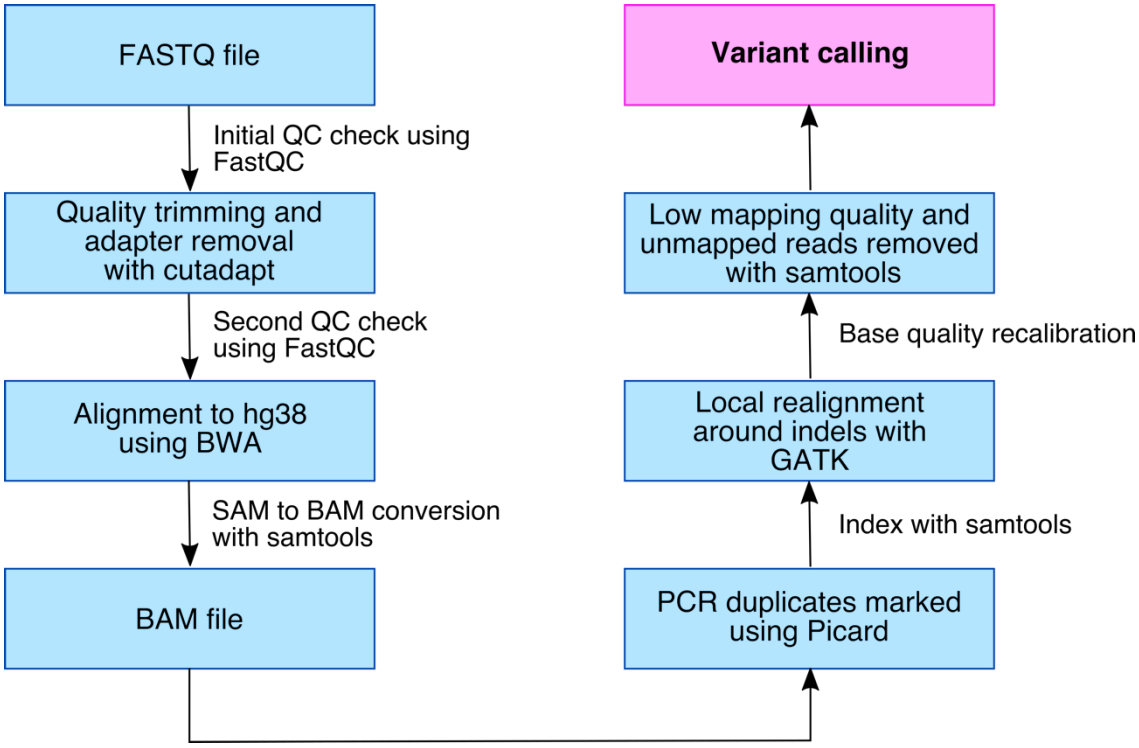


Figure 2.3: FASTQ processing for variant calling.
An overview of the steps undertaken to convert the raw read data (FASTQ) into a format ready for variant calling.

2.6.3 Downstream variant analysis

2.6.3.1 Identification of shared variants and consensus variant calling

Development of a two-pronged variant calling pipeline is detailed in Chapter 4 and was performed by RJA. Briefly: to reduce the number of false positive calls, the outputs from the five variant callers were used to identify the best combination of three variant callers for consensus calling. Consensus calls were generated for all 10 possible combinations of callers using the vcftools function "vcf-isec" with "-n +2" specified so that only variants present in two or more input VCFs were kept and the best combination of callers was identified. Variants were taken forward if they were identified by at least two of these three variant callers. Additionally, variants were included if they were identified in both tumours from a patient by any one caller of the final three. This is based on the methodology proposed by Droop *et al.*[242] who reason that variants are internally validated if they are identified in paired tumours, regardless of the strength of evidence in any one tumour. Once all shared and consensus called variants had been identified, a variant call format (VCF) file was generated for each tumour. An additional VCF file was generated containing the post-MMC unique variants for each patient for further analysis. The bcftools (https://github.com/samtools/bcftools)

package was used to generate stats for each VCF such as the number of transitions, transversions and variants for each base change.

2.6.3.2 MAF generation and variant annotation

All VCF files were converted into a single mutation annotation file (MAF) using "vcf2maf" from MSKCC (https://github.com/mskcc/vcf2maf/blob/master/README.md). This annotates the variants using the Variant Effect Predictor (VEP) from Ensembl[243]. MAF manipulations and visualizations were performed using maftools[244] within the R software environment. This included the generation of oncoplots.

2.6.3.3 Identification of tandem substitutions

Each patient's VCF files were analysed in R to find any consecutive variants. These variants were subset out from the large MAF file to create a tandem substitution MAF file. Any tandem substitutions with a classification of "frameshift" were removed as these were obviously erroneous. Two tandem substitutions were filtered at this point. Each tandem substitution was then investigated in the interactive genome viewer (IGV) version 2.4.6 to check for miss-labelling due to a near-by indel, to ensure that both variants were present on the same read and, where a tandem substitution is unique, ensure there were no reads in the other tumour(s). Two tandem substitutions were filtered at this point.

2.6.3.4 Calculation of mutations/Mb

Coverage files for the SureSelect Human All Exon Version 6 whole exome capture kit were downloaded from Agilent. The total number of bases covered were calculated for each interval using (end position – start position) +1. This was divided by 1000000 to get the number of megabases (Mb). The number of variants was then divided by the number of Mb to obtain the total number of variants per Mb.

2.6.4 FACETS CN estimation

Allele-specific copy number (ASCN) estimations were generated from the WES data using the FACETS package as per the authors' instructions[245]. Briefly; sequence readcount information is parsed from tumour-normal BAM files and readcounts for SNPs in both the tumour and the normal are generated using "snp-pileup". Positions with a total readcount below 35 or above 1000 in the matched normal are removed. To reduce hyper-segmentation in SNP-dense regions SNPs are subsampled in 150—250 bp regions. The readcount in the tumour vs normal is used to calculate log ratio (logR) which provides information on the total CN ratio. A log-odds-ratio (logOR) for each

position is also calculated and is defined by the log-odds ratio of the variant allele count in the tumour vs the normal. Data is normalised for the library size and GC corrected using a loess regression model along 1 kb windows. Segmentation is performed using an extension to the circular binary segmentation (CBS) algorithm to include both logR and logOR. Allele-specific CN is then estimated and further refined using the "fit$cncf" command and estimates of tumour purity and ploidy are generated.

2.6.5 Kernel density plots

Kernel density plots were generated to give a preliminary indication of the presence or absence of intratumour heterogeneity. Analysis was restricted to potentially functional variants (missense, nonsense, frameshift or mutations of the invariant dinucleotides at splice junctions) in regions that did not exhibit altered CN. These were obtained by sub-setting for variants within regions where the FACETS prediction of the total CN was 2 and the minor CN was 1 thereby excluding any variants with gains, losses or CN neutral LOH. Density plots were then generated using the "ggplot2" package within the R software environment.

2.6.6 PyClone

PyClone version 0.13.0 was used to infer the prevalence of point mutations within the tumours according to the authors' instructions[246]. To gain a better understanding of events that may influence protein structure or function, analysis was restricted to potentially functional variants. Briefly; input files were generated per sample by combining the CN analysis data from FACETS with the read count data for each variant. The PyClone "run_analysis_pipeline" command was then run using tumour purity estimates generated by visual analysis of the H&E stained tissue samples (Table 2.6) unless otherwise stated and "parental_copy_number" as formerly specified. This command runs the entire pipeline. Plots were restricted to a minimum cluster size of 2 mutations unless otherwise stated.

2.6.7 ClonEvol

ClonEvol was used for clonal ordering and clonal evolution visualisation as per the authors instructions[247]. ClonEvol infers consensus clonal evolution trees using the clustering of variants identified using other tools, such as PyClone, as an input. It estimates the cancer cell fraction of the clones (CCF) via a bootstrap resampling approach. Driver gene status was extracted from Bailey *et al.*[47] using the list of 299 genes that have been identified as drivers in some form of cancer. Fishplot models containing both tumours were created using the fishplot package in R[248].

Table 2.6: Tumour purity estimates from H&E stained tissue samples.
Tumour purity was estimated using the H&E stained tissue samples taken throughout the cutting process. *indicates sample underwent laser-capture microdissection prior to DNA extraction. In these cases the purity was estimated to be 95% to arbitrarily account for impurities in the sample that may have been missed. $ this sample underwent LCM however kernel density analysis suggested a lower purity for this sample due to a shift away from a VAF of 0.5 for heterozygous mutations. Therefore, the purity estimation from FACETS was used.

Tumour ID	H&E purity estimate	Tumour ID	H&E purity estimate
P0418-S02-BX	0.90	P1870-S2A-BX	0.8
P0418-S03-PX	0.98	P1870-S03-PX	0.87
P0533-S3B-BX	0.85	P2161-S01-BX	0.8
P0533-S04-PX	0.95*	P2161-S02-PX	0.8
P0960-S01-BX	0.8	P2218-S1A-BX	0.85
P0960-S03-PX	0.95*	P2218-S1B-BX	0.7
P0960-S04-PX	0.92	P2218-S02-PX	0.7
P1175-S01-BX	0.8	P2329-S02-BX	0.66$
P1175-S02-PX	0.95*	P2329-S03-PX	0.8

Chapter 3
Analysis of copy number alterations and hotspot mutations in tumours from patients with recurrent NMIBC

3.1 Introduction

An objective of this study was to characterise the genomic landscape of somatic copy number alterations (CNAs) in tumours from patients with recurrent non-muscle-invasive bladder cancer (NMIBC). A number of studies have profiled CNAs in bladder tumours. This has identified the most common alteration to be loss of chromosome 9 which is present in over 50% of all bladder tumours irrespective of stage and grade[125,131]. On chromosome 9, loss of 9p21.3, the region containing the tumour suppressor gene *CDKN2A,* is the most common event occurring in 50-60% of all tumours[4] and homozygous deletion (HD) of this region is associated with high stage and grade[125,249] and has been linked to recurrence in NMIBC[132]

MIBC and NMIBC genomes are very different; MIBC are chromosomally unstable with frequent alterations identified in almost all chromosomes whilst NMIBC are more stable with fewer CNAs identified[125]. Profiling of stage Ta tumours has identified recurrent CNAs including losses of 11p[125,135,136] and 17p[125,135], and gain of 20q[135,136]. A recent study investigating CNAs in patients with multiple tumours identified that homozygous deletions occur more frequently in patients with multiple tumours[250]. Homozygous deletion on chromosomes 10q, 11q, 18q and 21q were identified exclusively in recurrent or multifocal tumours and these regions were demonstrated to contain cell adhesion genes. Deletion of these genes was postulated to contribute to the spread of tumour cells in the bladder and could contribute the multiplicity of bladder tumours within patients[250].

The majority of these studies used array comparative genomic hybridization (aCGH) for the generation of copy number (CN) profiles[125,135,250]. The advent of next generation sequencing (NGS) has enabled the study of cancer genomes in much greater detail than ever before. CN analysis using NGS-based approaches has an improved resolution and faster turnaround time compared to array-based methods[251,252]. There are several different methods that can be used to extract CNAs from NGS data. Read-depth methods use the depth of coverage to determine CN. This method assumes that the number of reads mapping to a particular region is proportional to the CN of the region. Therefore, regions of gain or amplification will have a higher sequencing

coverage whilst deletions will have lower coverage[253]. The advantage of this method is that very low depth of coverage, as low as 0.1x, can be used to generate CN profiles[254,255] making it a very cost-effective method for the screening of multiple samples. Mate-pair sequencing uses the knowledge of the insert size between paired ends to identify read-pairs that map discordantly. Read pairs that map too far apart are indicative of deletions whilst pairs mapping too closely together indicate insertions[256]. Sequencing depth can be as low as 1-5x although greater sensitivity is found at 5x[252]. However, analysis is more challenging than read-depth methods and insertions larger than the insert size cannot be identified without additionally using read depth analysis[251,257]. Split-read methods use unmapped or partially mapped reads for the identification of breakpoints, however this can only be applied to unique regions of the genome[251].

Due to its low cost and ease of analysis, we employed shallow-pass whole genome sequencing and the read depth approach for the detection of CNAs in our cohort of 67 tumours from 23 patients. This method has been used extensively in our lab[82] and previous work carried out by the group has shown that the results generated are compatible with results generated by aCGH (see Appendix E for a comparison of CNAs in tumour P0468-S01 identified by aCGH and NGS). CNAs were used to identify recurrent regions of alteration and investigate differences in tumours according to stage and grade. We also investigated potential changes in CNAs between pre- and post-treatment tumours from a subset of patients who underwent a course of MMC chemotherapy.

Copy number profiles have also been used to define genomic subgroups of tumours. Hurst *et al.*[82] recently described two distinct genomic subgroups of stage Ta bladder tumours termed Genomic Subtype 1 (GS1) and Genomic subtype 2 (GS2). GS1 was characterized by no or few CNAs whilst GS2 was more genomically unstable with characteristic loss of 9q and a higher mutation rate. Genomic subgroups have also been identified for stage T1 grade 3 bladder tumours with hierarchical clustering identifying 3 separate clusters that differed with respect to CN profile[125]. We carried out hierarchical clustering using the CN data for tumours profiled in the study of Hurst *et al.*[82] alongside our CN data with the aim of identifying which subgroup the recurrent tumours cluster into. Hierarchical clustering was also performed on tumours from our cohort alone to examine if tumours from the same patient tend to cluster together.

Patients with bladder cancer often have multiple tumours and these can be synchronous tumours resected at the same time (multifocal disease) or metachronous,

where recurrent tumours are resected over a period of time. Historically there has been much debate on the clonal origins of such tumours[159]. Evidence from previous studies suggests that the majority of bladder tumours are monoclonal in origin[26,29,160,163,165,167,174] but there is also some evidence for oligoclonality[35,161]. We used our CN data to identify if tumors from our cohort were monoclonal or oligoclonal in origin. Additionally, CNAs were used to order the predicted genomic evolution of each tumour and create a phylogenetic tree for each patient.

As detailed in the introduction, key genes that are altered in bladder cancer include *FGFR3, PIK3CA, HRAS, KRAS* and *NRAS,* and the promoter region of the *TERT* gene. These genes are frequently mutated in bladder cancer and contain hotspot mutations that can be targeted using simple SNaPshot assays[143,156,232,233]. To improve the assessment of clonality, tumours from our cohort were analysed for hotspot mutations in these genes and associations between mutation status and CNAs were investigated.

3.2 Results

3.2.1 Copy number alterations

To assess copy number alterations (CNAs), shallow-pass whole genome sequencing was performed on 67 tumours from 23 patients. Samples were sequenced to an average raw coverage depth of 0.7x and analysed for CN changes using a pseudo-CGH algorithm followed by GC correction, segmentation and visualisation with the Nexus Copy Number software. Single copy gains were defined as regions with a $\log_2$ ratio greater than or equal to 0.25, and regions were classed as amplifications if the $\log_2$ ratio was greater than or equal to 1.2. Single copy losses were defined as regions with a $\log_2$ ratio less than or equal to -0.25, and regions were classified as homozygous deletions if the $\log_2$ ratio was less than or equal to -1.2. Overall, there was heterogeneity in the frequency of CNAs; some tumours from patients had no or few CNAs whilst others exhibited multiple regions of loss or gain.

To give an overview of CNAs within the cohort, a genome-wide frequency plot (GWFP) of CNAs identified in all 67 tumours was generated using Nexus (Figure 3.1A). Including all of the tumours in the analysis may artificially inflate the frequency of CNAs in certain regions if they are present in multiple tumours from the same person. To investigate this, GWFPs were generated for the first and last tumour from each patient

only (Figure 3.1B). These plots are very similar to the GWFP for all tumours, however, on visual inspection some CNAs appear to differ between the two groups; regions of CN gain on chromosomes 1, 3 and 12 appear to be a feature of the later tumours whilst the early tumours appear to have a higher frequency of chromosome 7 gain.

To investigate whether these apparent differences between early and late tumours were statistically significant, the genome wide frequencies of copy number alterations in the first and last tumours were compared using the Nexus software. The analysis subtracts the genome wide frequencies of CNAs in one group from the other then uses a Fisher's exact test to check for a significant difference in the frequency of CNAs between the two groups ($p < 0.05$, differential threshold of 25%). No significant differences in the frequencies of copy number events were identified.

The fraction of genome altered (FGA) is a measure of how much of the genome is affected by CNAs and provides a measure of chromosomal instability. Overall, the median FGA of the cohort of 67 tumours was 6.08% (range 0-25%). To investigate if there was a difference in chromosomal stability of the first and last tumours, the FGA was compared. No statistical difference between these two groups was identified and both groups showed very similar median FGA values (first = 5.165%, last = 5.549%, p=0.1231, two-tailed Wilcoxon matched-pairs signed rank test).

3.2.2 Copy number alterations and stage and grade

Tumour stage and grade are key components in the assessment of risk of recurrence and progression. Previous studies have identified an increased frequency of alterations in tumours of a higher stage or grade[125]. To assess if this is also the case in our cohort, GWFPs for stage Ta and T1 tumours were created (Figure 3.2A). Stage Ta tumours (n=57) demonstrate a stable genome overall, with deletion of chromosome 9 the only recurrent event. In contrast, the stage T1 tumours (n=8) exhibited a higher frequency of CNAs.

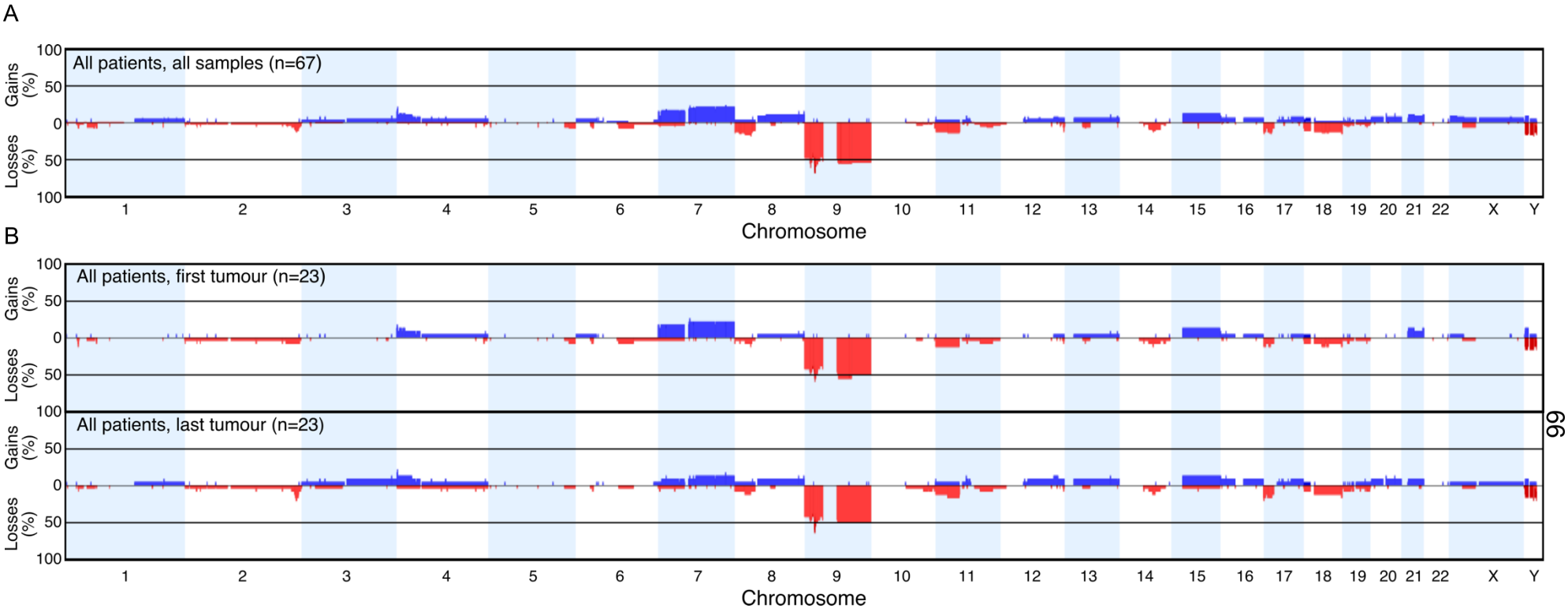


Figure 3.1: Genome-wide frequency plots of CNAs identified in 67 tumours from 23 patients.
A) Regions of CNA in all 67 tumours. B) Regions of CNA in the first tumour from each patient (top) and the last tumour from each patient (bottom). The *x*-axis corresponds to chromosome number and the *y*-axis corresponds to the percentage of gains and losses. Copy number gains are shown in blue and losses in red.

We were interested to identify if there were any statistically significant regions of CNA associated with stage. Using the Nexus comparisons function 70 regions distributed across the genome were identified to be differentially altered (25% threshold; $p < 0.05$) between stage Ta and T1 tumours (Figure 3.2B). All 70 events were regions of gain or loss that were significantly more frequent in the stage T1 tumours. Most notably, an increase in chromosome arm level gains of 7p, 7q and 8q as well as gain of chromosome region 17q22-q25.3 were identified in the stage T1 tumours.

Separation of tumours by grade (as defined by the 1973 WHO guidelines[8]) revealed a GWFP profile similar to that generated for tumours separated according to stage (Figure 3.2C). Of the 12 grade 3 (G3) tumours, 4 were stage T1 and 8 stage Ta. Chromosome 9 deletion remained the most common event in both groups whilst the chromosome 7 gains seen in stage T1 tumours were equally divided between G1/2 and G3 tumours. The comparisons function within Nexus was used to identify if any regions were significantly different according to grade. This identified 56 regions that differed significantly in frequency between G1/2 and G3 tumours (Figure 3.2D). All 56 regions were statistically more frequent in grade 3 tumours and included gain of chromosome region 6p25.3 – p21.1 (p = 0.005), loss of 11p15.5 - p14.3 (p = 0.007), loss of 18q12.2 - q12.3 and loss of 18q21.1 - q21.2 (p = 0.012 for both) (25% differential threshold, Fishers exact test).

Quantification of the level of chromosomal instability using FGA identified a significantly higher median FGA in the stage T1 tumours compared to the stage Ta tumours (Figure 3.2E, Mann-Whitney U test, p=0.0012, Ta median=5.55, T1 median = 17.08), and in grade 3 tumours compared to grade 1&2 tumours (Figure 3.2F, Mann-Whitney U test, p=0.0114, G1&G2 median=5.55, G3 median = 14.11), confirming the higher level of chromosomal instability in higher stage and/or grade tumours.

3.2.3 Alterations involving chromosome 9

The most common deletion identified in bladder cancer is deletion of all or part of chromosome 9[125,131]. Analysis of GWFPs identified deletion of chromosome 9 to be the most common event in this cohort, and this remained true when analysing only the first or last tumours from each patient (Figure 3.1). Ten patients demonstrated a loss of the whole of chromosome 9. Loss of chromosome 9 was identified in all tumours from a given patient, suggesting that it is an early event in these patients.

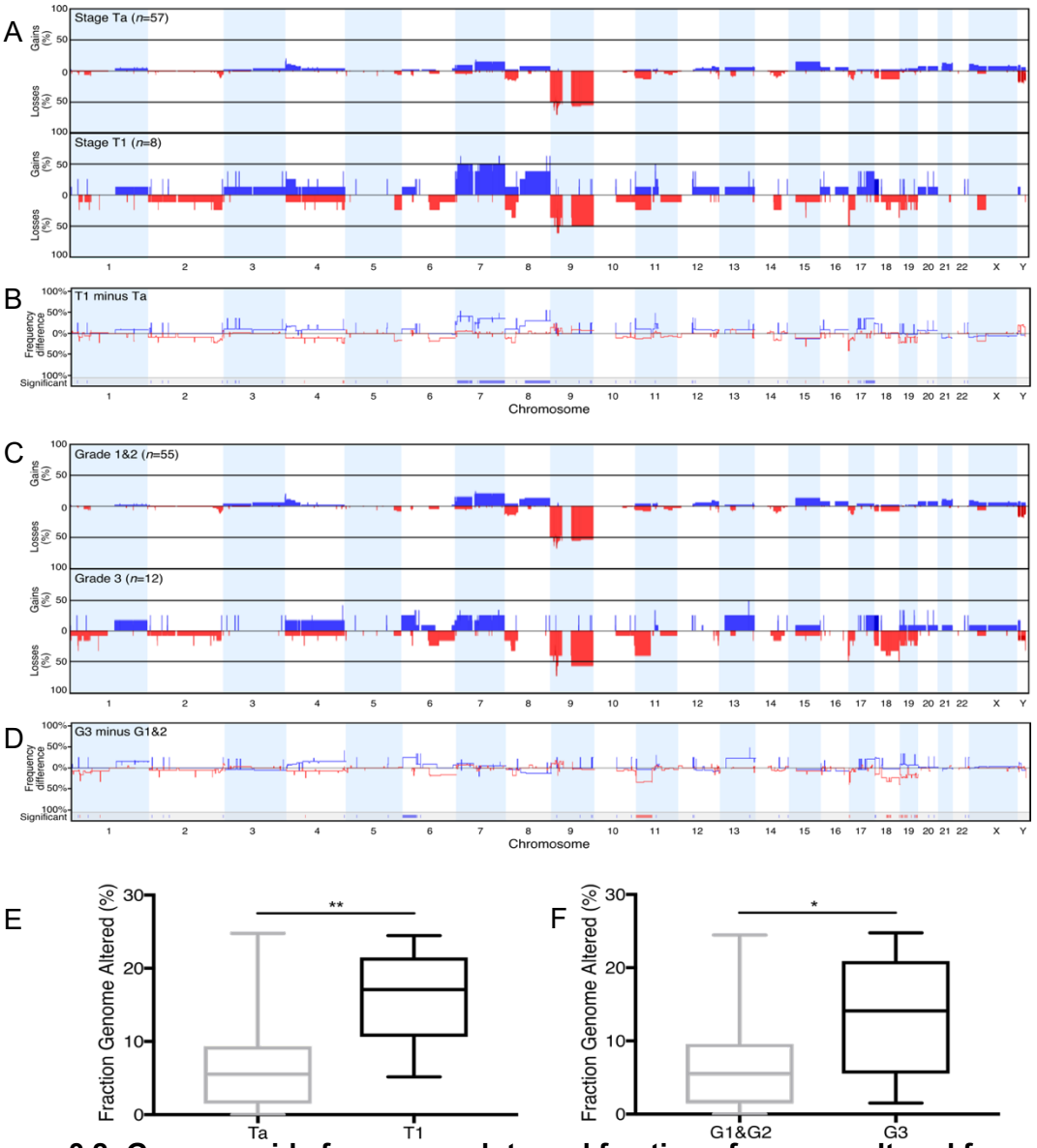


Figure 3.2: Genome-wide frequency plots and fraction of genome altered for tumours according to stage and grade.

A) Genome-wide frequency plots of CNAs in tumours according to stage. B) Frequency difference plot for Ta versus T1 tumours. The difference is obtained by subtracting the CNA frequencies of the stage Ta tumours from the stage T1 tumours. Gains are plotted in blue, losses in red. If the magnitude of gain or loss is greater in the stage T1 tumours then the gain is plotted in blue above the 0 baseline and the loss is plotted in red below the 0 baseline. If the magnitude of gain or loss is greater in the stage Ta tumours then the gain is plotted in blue below the 0 baseline and the loss is plotted in red above the 0 baseline. Regions with a significant difference ($p < 0.05$, 25% differential threshold, Fisher's exact test) are indicated at the bottom of the plot by coloured blocks. C) Genome-wide frequency plots in tumours according to grade. D) Frequency difference plot for grade 1&2 versus grade 3 tumours. The difference is obtained by subtracting the CNA frequencies of the grade 1&2 tumours from the grade 3 tumours. Gains are plotted in blue, losses in red. If the magnitude of gain or loss is greater in the grade 3 tumours then the gain is plotted in blue above the 0 baseline and the loss is plotted in red below the 0 baseline. If the magnitude of gain or loss is greater in the grade 1&2 tumours then the gain is plotted in blue below the 0 baseline and the loss is plotted in red above the 0 baseline. E&F) FGA was compared according to stage and grade and represented as boxplots. A significant difference in FGA was detected between; E) stage Ta and T1 tumours (p=0.0012; Ta median=5.55, T1 median=17.08) and F) grade 1&2 (G1&G2) and grade 3 (G3) tumours (p=0.0114; G1&G2 median=5.55, G3 median=14.11). Boxes show the interquartile range and median with the whiskers showing the range of the data. Significance was tested with the Mann-Whitney U test.

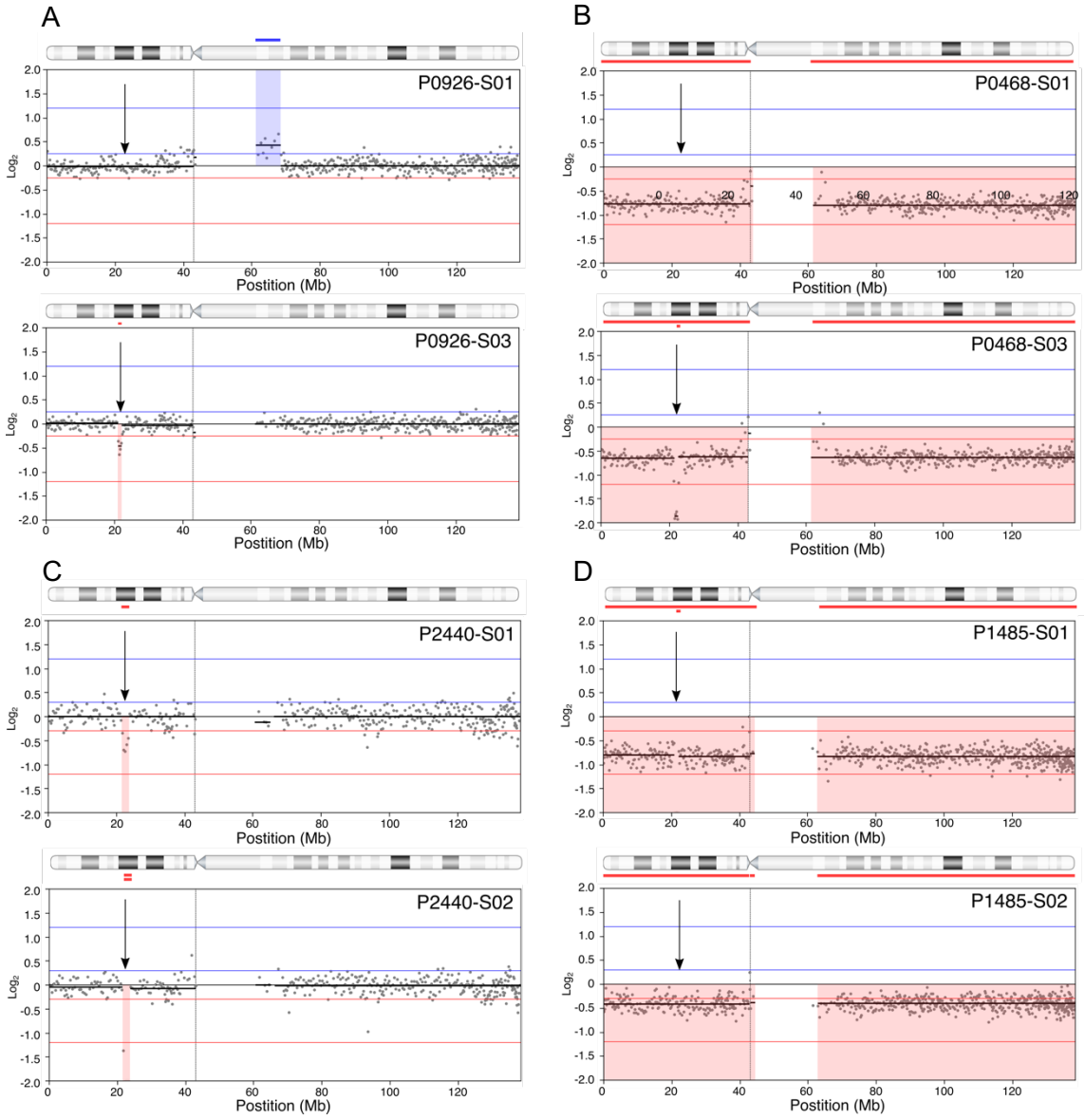


Figure 3.3: Copy number analysis reveals differences in *CDKN2A* loss between tumours from the same patient.

Analysis of chromosome 9 identified some patients where the tumours differed with respect to *CDKN2A* loss. The images show individual chromosome plots of CN data for chromosome 9. The *CDKN2A* locus is highlighted by the black arrows. Regions of CN loss are shaded in red below the zero line whilst regions of CN gain are shaded in blue above the zero line. Chromosome 9 with cytobands is depicted above each plot. Regions of CNA are highlighted by blue (gains) or red (loss) lines above and below the chromosome respectively. A single red line represents loss of the region (log_2 ≤-0.25 >-1.2), whilst a double red line represents homozygous deletion (log_2 ≤-1.2). A) P0923-S03 shows loss of the *CDKN2A* locus whereas P0926-S01 does not. B) P0468-S03 has HD of the *CDKN2A* locus whereas P4068-S01 has only loss of this region due to loss of the whole of chromosome 9. C) Loss of *CDKN2A* was identified in P2440-S01 whilst HD of this region was identified in P2440-S02. D) The HD of *CDKN2A* seen in P1485-S01 is not present in P1485-S02. The *x*-axis is the chromosome position (Mb) whilst the *y*-axis is the log_2 value.

The most common event involving chromosome 9 was loss of copy number at 9p21.3. A total of 47 tumours from 15 patients had loss in this region. In one patient, P0926, loss of this region was identified only in the recurrent tumour (Figure 3.3A). Of the 47 tumours with loss of 9p21.3, 31 tumours from 10 patients had HD at this locus. Three patients had tumours that exhibited different levels of loss and HD at 9p21.3. In patient P0468, the initial tumour (P0468-S01) had loss of the whole of chromosome 9 only, whilst in the recurrent tumour (P0468-S03), an additional focal loss of 9p21.3 resulted in HD of this region (Figure 3.3B). In patient P2440, tumour P2440-S01 contained a focal loss of 9p21.3 whilst in tumour P2440-S02 this was a HD (Figure 3.3C). In patient P1485, tumour P1485-S01 demonstrated HD of 9p21.3 whilst tumour P1485-S02 did not (Figure 3.3D), suggesting that tumour P1485-S02 may be genomically ancestral to tumour P1485-S01, despite being resected at a later date.

The size of the region of homozygous deletion at 9p21.3 varied between patients. Patient P0533 had the largest region of deletion, containing 33 genes including the *IFNα* cluster, *MLLT3* and the *CDKN2A/B* genes. This extended deletion was also seen in tumour P2218-S1B. The region deleted in tumours from patient P1777 and tumour P1485-S01 also included the *IFN1α* cluster but did not include *MLLT3*. The minimum deletion region in all tumours included 5 genes (*MTAP, C9orf53, CDKN2A, CDKN2B* and *CDKN2B-AS1*). To give a more accurate representation of the frequency of deletion in these regions, the *CDKN2A* minimum region was listed separately (Table 3.1).

Five other homozygous deletions were identified on chromosome 9; all five tumours from patient P0533 contained a HD of 9p23 (*PTPRD)*, all five tumours from patient P0712 contained a HD of 9p22.3-p22.2 (*CCDC171, BNC2, C9orf92, CNTLN, SH3GL2*), all four tumours from patient P0198 contained a HD of 9p21.2 (*CAAP1, PLAA, IFT74, TEK*, *MOB3B*), tumour P0717-S02 contained a HD of 9q21.13 (*RORB, TRPM6, OSF1, CARNMT1)* and tumour P0960-S01 contained a HD of 9q22.31-q22.32 (*PHF2, BARX1)* (Table 3.1).

3.2.4 Focal regions of copy number change

It is difficult to identify the target gene(s) in large regions of copy number alteration where many genes may be affected by the alteration. In contrast, potential candidate genes are more easily identifiable in focal regions of CNA. Focal CNAs tend to be less than 3 Mb in size[258] and these are often regions of amplification or HD. Amplifications are genetic alterations that produce high copy numbers of a small section of the

genome whilst HD are deletions of both copies of a genomic region. These regions often contain oncogenes or tumour suppressor genes, respectively, making their identification important in the analysis of cancer development and progression.

3.2.4.1 Amplifications

Overall there were 7 regions of amplification detected in 5 tumours (Table 3.1). Two amplification events were each unique to an individual tumour; 11q13.3 (P1485-S01) and 13q33.3-q34 (P0468-S01). The remaining five amplification events (3p25.2, 7q34, 7p21.1, 17q21.33 and 18p11.32-p11.21) were present in the three tumours from patient P0418.

Figure 3.4 illustrates the high-level amplifications at 11q13.3 and 13q13.3-q34, detected in patients P1485 and P0468, respectively. The amplification detected at 11q13.3 contained 16 candidate genes including *CCND1* and *FGF3, FGF4* and *FGF19.* This high-level amplification ($\log_2 = 4.03$) was present in tumour P1485-S01 but not in any additional tumours from this patient (Figure 3.4A). The high-level amplification at 13q33.3-q34 was identified in tumour P0468-S01. This region contains 12 genes including *IRS2, RAB20, COL4A1* and *COL4A2*. In P0468-S01 this 2Mb region of high-level amplification was surrounded by focal losses (Figure 3.4B). A smaller part of this region, containing *IRS2* and *COL4A1* and a long intergenic non-protein coding RNA (LINC00676), was gained in P0468-S03 but not in tumour P0468-S05 (Figure 3.4B).

Patient P0418 proved to be an interesting case. The three tumours from this patient, shared over 49 regions of focal gain/amplification across the genome (Figure 3.5A). Only the sex chromosomes and autosomes 15 and 21 did not contain any focal gains. Despite the large number of gains, there were only five regions of amplification: 3 regions shared by all three tumours and 2 regions shared by two tumours only. A 1.2 Mb amplification on chromosome 3p25.2 was shared by all 3 tumours (Figure 3.5B). This region contains 13 genes including *RAF1* and *PPARG.* All 3 tumours also shared a region of amplification on chromosome 17q21.33. In this case, the size of this region differed for tumour P0418-S02 which had a larger region of amplification compared to the other two tumours. As the other two tumours had a smaller amplification it is likely that this is the minimal region and this contains 17 genes including *SPAG9* (Table 3.1).

Table 3.1: Regions of amplification and homozygous deletion identified within the cohort of 67 tumours.
Regions of amplification (log_2 ≥1.2) and homozygous deletion (log_2 ≤-1.2) are listed along with information about each region. Position information is for genome build hg38.

Chr	Position (Mb)	Size of region (Mb)	Cytoband	No. of genes	Candidate Genes	No. of tumours	Tumour IDs	Tumour stage and grade
Amplifications								
chr3	11,668,795-12,871,104	1.2	p25.2	13	*RAF1, PPARG*	3	P0418-S01, P0418-S02, P0418-S03	2 T1(x)G3, 1 Ta(x)G3
chr7	140,533,114-141,337,049	0.8	q34	7	*BRAF*	2	P0418-S01, P0418-S02	T1(x)G3, Ta(x)G3
chr7	16,797,571-17,661,296	0.9	p21.1	5	*AGR2, AGR3, AHR*	2	P0418-S01, P0418-S02	T1(x)G3, Ta(x)G3
chr11	68,965,695-70,123,475	1.2	q13.3	16	*CCND1,FGF3, FGF4, FGF19*	1	P1485-S01	TaG2
chr13	108,804,496-110,772,166	2.0	q33.3 - q34	12	*IRS2, RAB20, COL4A1, COL4A2*	1	P0468-S01	T1G2
chr17	50,409,038-51,152,545	0.7	q21.33	17	*SPAG9*	3	P0418-S01, P0418-S02, P0418-S03	2 T1(x)G3, 1 Ta(x)G3
chr18	194,371-14,397,773	14.2	p11.32 - p11.21	98	*TYMS, YES1*	3	P0418-S01, P0418-S02, P0418-S03	2 T1(x)G3, 1 Ta(x)G3
Homozygous Deletions								
chr9	10,477,846-13,155,435	2.7	p23	6	*PTPRD*	5	P0533-S01, P0533-S02, P0533-S3A, P0533-S3B, P0533-S04	4 TaG2, 1 TaG3
chr9	15,807,283-18,441,002	2.6	p22.3 - p22.2	5	*CCDC171, BNC2, C9orf92, CNTLN, SH3GL2*	5	P0712-S01, P0712-S02, P0712-S03, P0712-S04, P0712-S05	3 TaG2, 2 TaG3

Chr	Position (Mb)	Size of region (Mb)	Cytoband	No. of genes	Candidate Genes	No. of tumours	Tumour IDs	Tumour stage and grade
Homozygous Deletions								
chr9	20,908,413-21,459,428	0.6	p21.3	22	*IFN1a* locus, IFNε	9	P0533-S01, P0533-S02, P0533-S3A, P0533-S3B, P0533-S04, P2218-S1B, P1485-S01, P1777-S01, P1777-S02	5 TaG2, 3 TaG3, 1 T1G3
chr9	21,874,069-22,477,467	0.6	p21.3	5	*CDKN2A, CDKN2B, MTAP, C9orf53, CDKN2B-AS1*	31	P0198-S01, P0198-S02, P0198-S03, P0198-S04, P0468-S03, P0468-S05, P0533-S01, P0533-S02, P0533-S3A, P0533-S3B, P0533-S04, P0712-S01, P0712-S02, P0712-S03, P0712-S04, P0712-S05, P0717-S01, P0717-S02, P0717-S03, P0717-S04, P0717-S05, P1175-S01, P1175-S02, P1485-S01, P1485-S03, P1777-S01, P1777-S02, P2218-S1A, P2218-S1B, P2218-S02, P2440-S02	22 TaG2, 5TaG3, 2 T1G2, 2 T1G3
chr9	26,779,074-27,328,234	0.5	p21.2	9	*CAAP1, PLAA, IFT74, TEK, MOB3B*	4	P0198-S01, P0198-S02, P0198-S03, P0198-S04	3 TaG2, 1 T1G2
chr9	74,202,597-75,624,522	1.4	q21.13	9	*RORB, TRPM6, OSF1, CARNMT1*	1	P0717-S02	TaG2
chr9	93,366,129-93,999,641	0.6	q22.31 - q22.32	6	*PHF2, BARX1*	1	P0960-S01	TaG2
chrX	123,761,766-125,065,906	1.3	q25	4	*XIAP, STAG2, SH2D1A, TENM1*	1	P0717-S03	T1G2

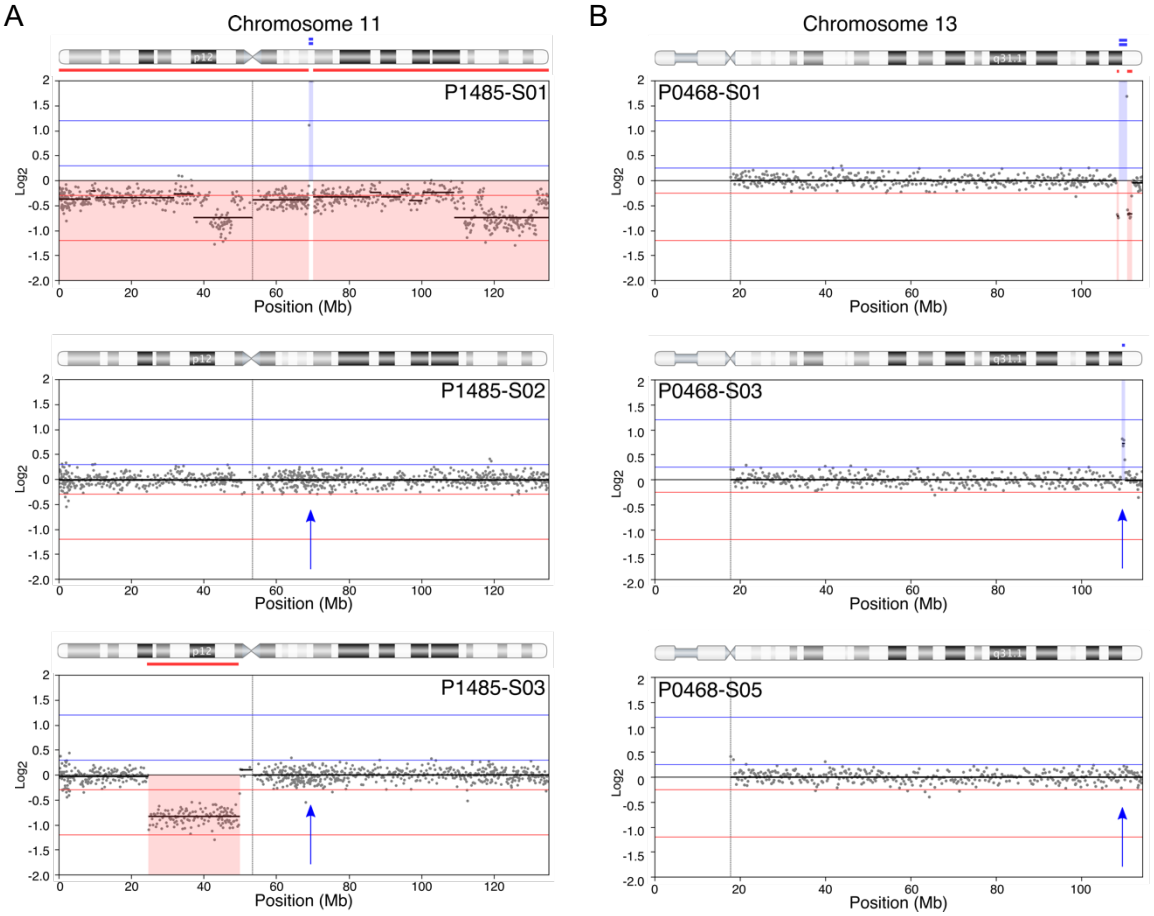


Figure 3.4: Examples of regions of focal amplification identified in two patients. Amplifications identified on chromosome 11 and chromosome 13 in tumours P1485-S01 and P0468-S01 respectively, are depicted on whole chromosome CN plots. Regions of CN loss are shaded in red below the zero line whilst regions of CN gain are shaded in blue above the zero line. The chromosome with cytobands is depicted above the plot. Regions of CNA are highlighted by blue (gains) or red (loss) lines above and below the chromosome respectively. A single line represents gain/loss whilst a double line represents amplification or HD (log_2 ≥1.2 or ≤-1.2 respectively). The *x*-axis is the chromosome position (Mb) whilst the *y*-axis is the log_2 value. A) Individual chromosome 11 CN plots for tumours from patient P1485. P1485-S01 has an amplification at 11q13.3 highlighted in blue. This amplification is not present in the other tumours from this patient (blue arrows depict where the amplification would be). B) Individual chromosome 13 CN plots for tumours from patient P0468. P0468-S01 has an amplification of 13q33.3-q34 highlighted in blue. In P0468-S03 gain of this region is detected but there is no amplification and in P0468-S05 there are no CNAs in this region. Blue arrows depict where the gain would be seen in the other tumours.

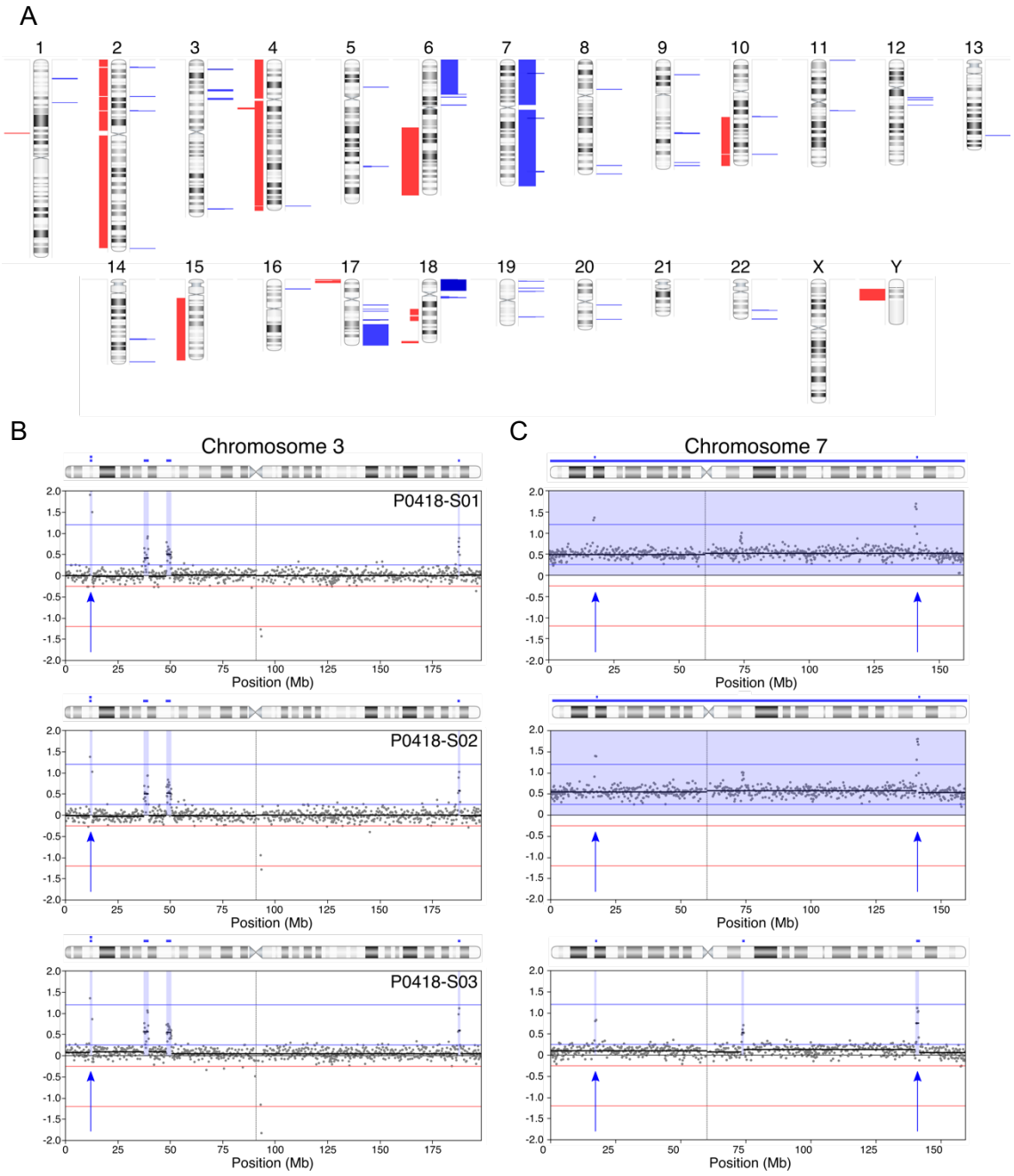


Figure 3.5: Focal gains and amplifications identified in patient P0418
The genome of patient P0418 is punctuated with several regions of focal CN gain and amplification. A) A summary of all events detected within the three tumours from patient P0418. Each chromosome is depicted with losses in red to the left and gains in blue to the right. These are stacked so that the thickness is related to the number of tumours with that alteration. B-C) Individual chromosome CN plots of chromosomes 3 and 7 for each tumour. Regions of CN gain are shaded in blue above the zero line. The chromosome with cytobands is depicted above each plot. Regions of gain/amplification are highlighted by blue above the chromosome. A single line represents a gain whilst a double line represents an amplification ($\log_2$ ≥1.2). The *x*-axis is the chromosome position (Mb) whilst the *y*-axis is the $\log_2$ value. B) A region of amplification at 3p25.2 is present in all 3 tumours. This region contains the genes *RAF1* and *PPARG*. C) All three tumours have 3 focal regions of gain on chromosome 7. Tumours P0418-S01 and P0418-S02 also contain a whole chromosome gain. This likely occurred after the focal gains.

All 3 tumours from patient P0418 also contained a high-level amplification of a region of 18p11.32 - p11.21 (log_2 = 1.76-1.90) which contains 98 genes including *YES1* and *TYMS* (Table 3.1). On chromosome 7, all three tumours from patient P0418 shared 3 focal alterations; 7p21.1, 7q11.23 and 7q34. In tumours P0418-S01 and P0418-S02, the alterations at 7p21.1 and 7q34 were amplifications whilst in tumour P0418-S03 these were focal gains (Figure 3.5C). Tumours P0418-S01 and P0418-S02 also contained a gain of the whole of chromosome 7. This could suggest that the focal gains occurred first followed by gain of the whole of chromosome 7. The region at 7q34 contains the proto-oncogene *BRAF* whilst the region at 7p21.1 contains no genes currently implicated in cancer.

3.2.4.2 Homozygous deletions

Deletions of chromosome Y were common, with 22 tumours from 5 patients showing complete loss of all or part of chromosome Y. Two tumours from P0712 had loss of Xq25, a region containing 4 genes (*XIAP, STAG2, SH2D1A* and *TENM*) (Table 3.1). This was a HD event in tumour P0712-S03 with a log_2 value of -1.52 whilst in tumour P0712-S02 this was a focal loss with a log_2 value of -0.99.

3.2.4.3 The effect of MMC chemotherapy course on CNAs

A subset of 16 patients underwent at least one 6-week course of MMC chemotherapy during their disease span. *In vitro* and *in vivo* studies have identified structural alterations and deletions associated with MMC treatment. Therefore we hypothesized that post-MMC treatment tumours may contain more CNAs. Although some patients had multiple tumours available, only a single tumour pre-MMC and post-MMC treatment was analysed per patient. These were the tumours resected closest to the treatment time. GWFP were generated for pre-MMC and post-MMC treatment tumours (Figure 3.6A). Visualisation of these profiles suggested that they were very similar, although the post-MMC samples appeared to have slightly more gains than the pre-MMC group. Statistical analysis using the Nexus comparisons function identified no significantly different regions of CNA (Fishers exact test, differential threshold of 25%). The relationship between FGA and treatment status was also compared and no significant difference was identified between the pre- and post-chemotherapy samples (Figure 3.6B, paired t-test p=0.4715).

3.2.5 Whole genome plots of individual tumours

Recurrent and multifocal bladder tumours from an individual patient have been shown to be clonally related with only a few instances of oligoclonality identified within the literature. We investigated whether the recurrent tumours were clonally related. Whole genome plots (WGP) were created for each tumour to enable visualisation of genome-wide changes on a global scale. Comparison of these plots allows easy identification of shared and unique CNAs amongst tumours from the same patient as well as assessment of overall chromosome stability. Some patients had chromosomally stable tumours with few CNAs, whilst other tumours contained many CNAs. An example of WGP from two patients (P1175; P2161) with stable genomes can be seen in Figure 3.7 whilst an example of tumours from a patient (P0712) with more CNAs is shown in Figure 3.8. WGPs for all other patients are presented in Appendix F.

Overall, shared CNAs suggestive of a clear monoclonal origin were found for 15 patients (P0198, P0418, P0533, P0712, P0717, P0926, P0960, P1175, P1326, P1485, P1777, P1870, P2104, P2329 and P2440). For 3 patients, P0933, P2065 and P2291, monoclonality could not be established as tumours did not share any CNAs. For a further 4 patients (P468, P536, P2161 and P2218) there was a suggestion of monoclonality due to shared loss of chromosome 9 (P468 and P2218) or alterations of chromosome Y (P0536 and P2161). However, these are common alterations in bladder cancer and could have occurred independently. For the final patient, P0990, a monoclonal origin between tumours P0990-S04 and P0990-S05 could be established as these shared all their CNAs, but tumour P0990-S01 did not contain any CNAs, therefore a monoclonal origin for this tumour could not be inferred.

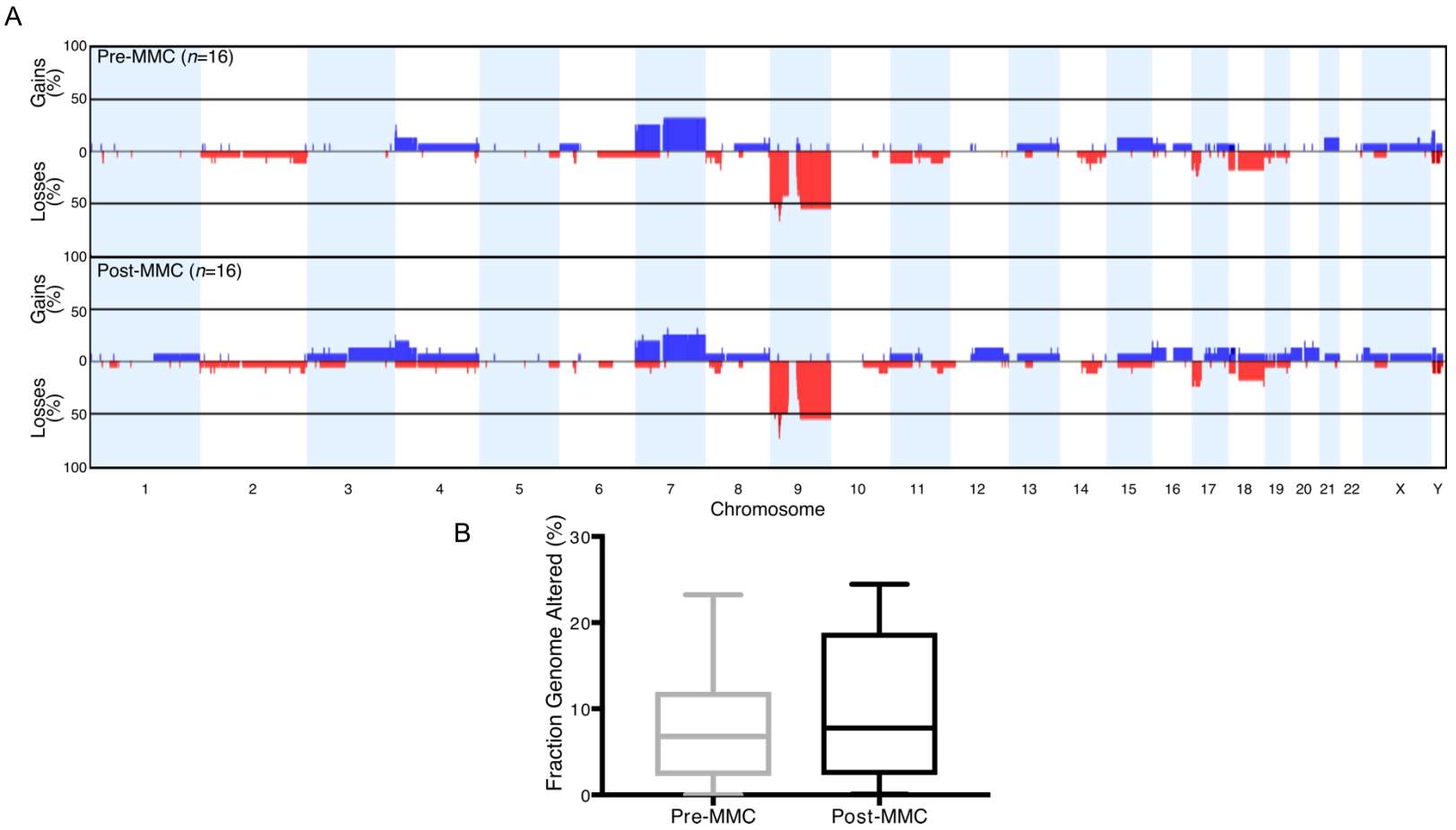


Figure 3.6: Genome-wide frequency plots of CNAs and FGA in matched pre- and post-MMC treatment tumours.
A) Genome-wide frequency plots of CNAs in tumours pre- and post-MMC chemotherapy treatment. The *x*-axis corresponds to chromosome number and the *y*-axis corresponds to the percentage of gains and losses seen. Copy number gains are shown in blue and losses in red. B) Comparison of FGA in tumours pre- and post-MMC treatment. No significant difference in FGA was identified. Boxes show the interquartile range and median with the whiskers showing the range of the data (Pre-MMC median = 6.81%, post-MMC median = 7.76, p = 0.4715, paired t-test).

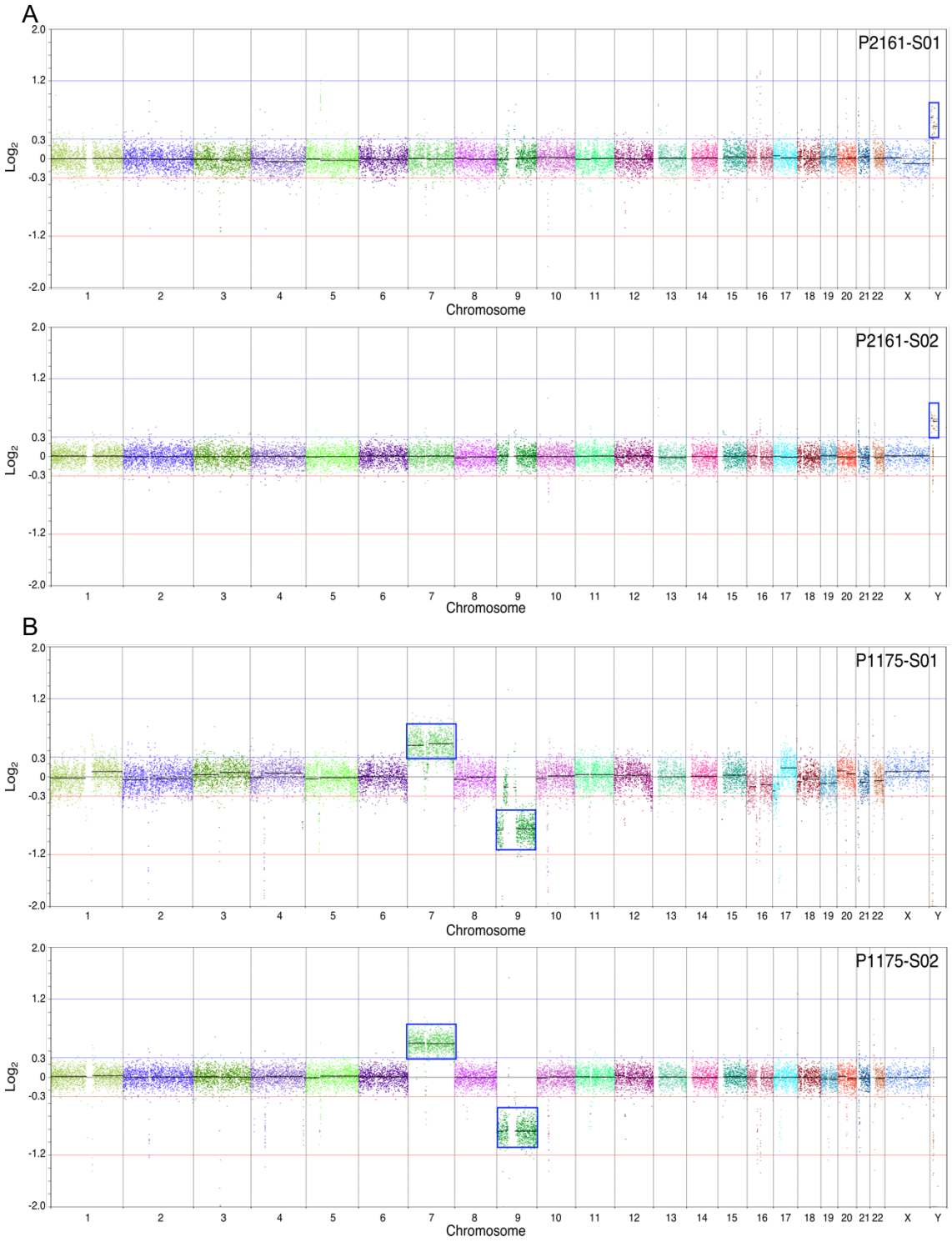


Figure 3.7: Examples of whole genome copy number plots for two patients.
A) WGPs of tumours from patient P2161 which exhibit very stable genomes with the exception of a gain of chromosome Y seen in both tumours. B) WGPs of tumours from P1175. These two tumours share two regions of CNA (a gain of chromosome 7 and a loss of chromosome 9) highlighted in the blue boxes. The x-axis represents the chromosome number and the y-axis the log2 values.

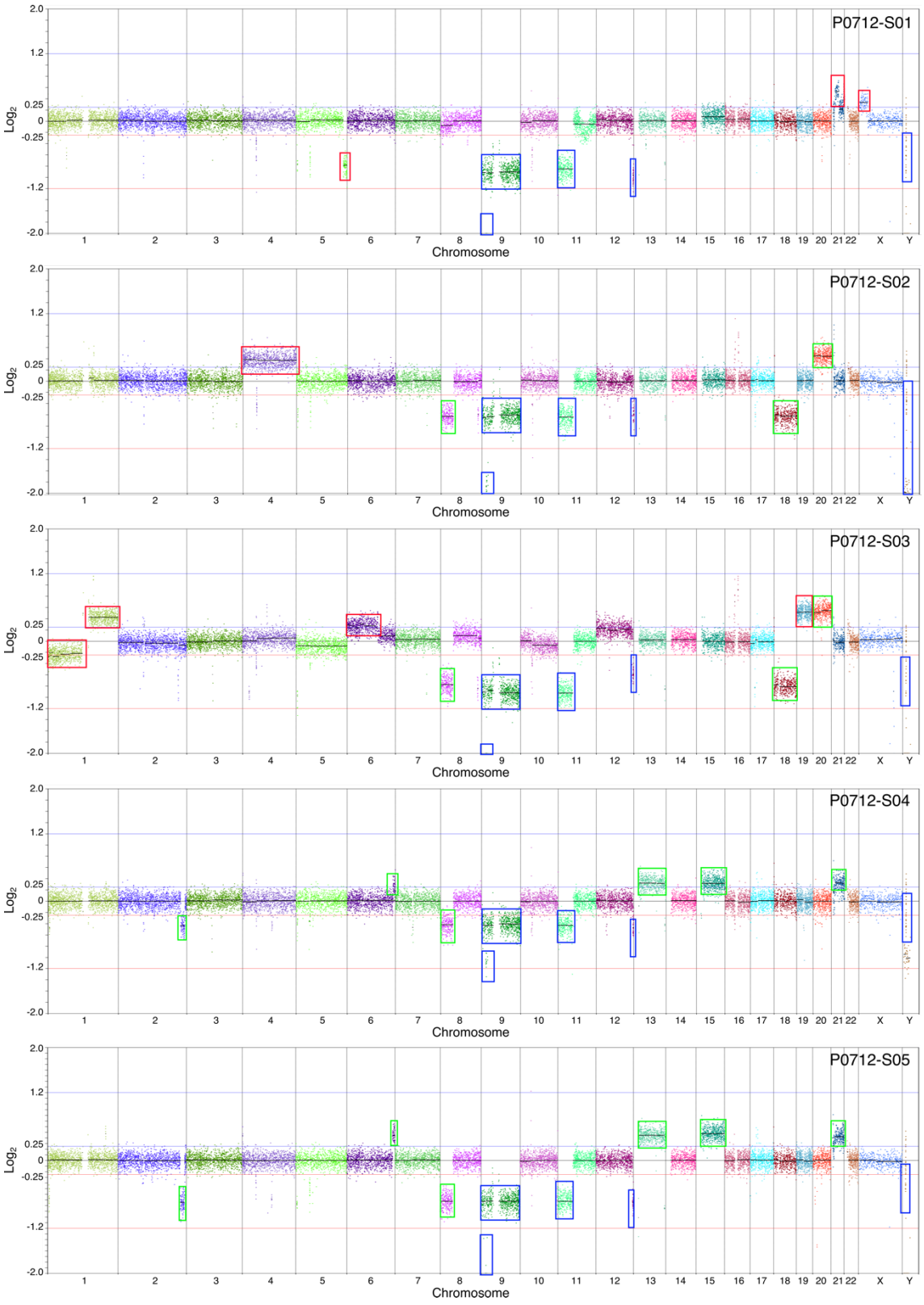


Figure 3.8: Whole genome copy number plots for five tumours from patient P0712.

The genomes of tumours from patient P0712 exhibit several regions of CNA. Some of these are shared between all tumours (highlighted by blue boxes) allowing a monoclonal origin to be inferred. These tumours also exhibit diversity in copy number alterations, with some events being shared by more than one tumour but not all tumours (green box). Some tumours also have unique events (red box).

3.2.6 Hotspot mutation analysis of key genes

Bladder cancer has mutational hotspots in the coding regions of five key genes: *FGFR3, PIK3CA, HRAS, KRAS* and *NRAS.* Additionally, recurrent mutations in the promoter region of the *TERT* gene have been identified. These genes play an important role in cellular processes and have been heavily implicated in carcinogenesis where they can act as oncogenes. Analysis of mutations in these genes may provide insight into factors driving tumour growth. Additionally, shared and unique mutations can provide more information for the assessment of clonality. This may be useful in the eight cases where clonality could not be established from CNAs alone.

The mutation status of hotspot mutations was interrogated using SNaPshot assays (Table 3.2). Mutations in the *TERT* promoter were most frequent, with mutations detected in at least one tumour for 21 patients (61/67 tumours). A total of 18 patients demonstrated *FGFR3* mutations in at least one of their tumours with a total of 48/67 tumours containing *FGFR3* mutations (43/59 stage Ta (72.9%), 4/8 stage T1 (50%) and 1 possible low grade UCC where a highly fragmented sample did not allow for accurate staging and grading). The most common *FGFR3* mutation was S249C seen in 32 tumours from 12 patients, followed by Y375C which was detected in 11 tumours from 4 patients. G372C and R248C mutations were identified in one patient each. Only 2 patients had RAS mutations (P0198 and P0533) and these were mutually exclusive with *FGFR3* mutations.

Analysis of *FGFR3* mutations in P1175 identified that tumour P1175-S01 was homozygous for the G372C mutation whilst tumour P1175-S02 was heterozygous (Figure 3.9A). To see if this was due to changes in CN, the genomic position at which the *FGFR3* gene is located on chromosome 4 was examined. No changes in CN were identified (Figure 3.9B), suggesting that tumour P1175-S01 has CN neutral loss of heterozygosity (LOH) in this region. We considered the possibility that LOH may not have been detected in tumour P1175-S02 due to the presence of contaminating normal cells. Examination of H&E sections for the two tumours showed that tumour P1175-S01 was large with a purity of >90% tumour cells whilst S02 was small and mostly impure. However, P1175-S02 had good spatial separation of the tumour and non-tumour cells, enabling the tumour cells to be isolated using macrodissection prior to DNA extraction. This would have resulted in a tumour purity of close to 100%, therefore it is unlikely that contamination of normal DNA obscured detection of LOH in this tumour. This suggests that the evolutionary path of the tumours diverged prior to the LOH event.

PIK3CA was mutated in 26.9% of tumours (18/67) with the majority of these tumours containing concurrent *FGFR3* mutations (13/18). Patient P0533 represented an interesting case as only 2 of the 5 tumours contained a *PIK3CA* mutation and these were both different (Figure 3.10). Tumour P0533-S02 contained an E542K mutation and tumour P0533-S04 contained an E545K mutation, neither of which were detectable in the other tumours from this patient. As this analysis was completed using whole-genome amplified DNA we decided to assess the mutation status in unamplified genomic DNA. This confirmed that both mutations were present and different.

Mutations in the RAS family of genes were identified in two patients only (P0198 and P0533). All tumours from patient P0198 contained a *HRAS* G13R mutation whilst all tumours from patient P0533 contained a *KRAS* G12V. Mutations in the RAS genes and *FGFR3* have been shown to be mutually exclusive. Compatible with this, the two patients with RAS mutations were wild-type for *FGFR3*.

All patients had at least one hotspot mutation in one of their tumours (Table 3.2). Two tumours (P2065-S02 and P2291-S02) did not contain any hotspot mutations even though other tumours from the same patients carried mutations. Patient P2065 had two tumours, one of which contained a *FGFR3* S249C mutation and a *PIK3CA* H1047R mutation (P2065-S01) whilst tumour P0265-S02 did not carry either of these mutations. A similar situation was seen in P2291 where tumour P2291-S01 contained a *FGFR3* S249C mutation whilst tumour P2291-S02 did not (Figure 3.11).

Table 3.2: SNaPshot mutation analysis.
SNaPshot assays were used to scan all whole-genome amplified DNA samples for mutations in known hotspots. Mutations are referred to by codon location and amino acid change (*FGFR3, PIK3CA* and RAS genes) or by distance from the transcription start site and base change (*TERT*).

Patient ID	Sample ID	Stage and Grade	*FGFR3* Mutation Status	*PIK3CA* Mutation Status	RAS Mutation Status	*TERT* Promoter Mutation Status
P0198	P0198-S01	TaG2	WT	WT	*HRAS* G13R	-124 G>A
	P0198-S02	T1G2	WT	WT	*HRAS G13R*	-124 G>A
	P0198-S03	TaG2	WT	WT	*HRAS* G13R	-124 G>A
	P0198-S04	TaG2	WT	WT	*HRAS* G13R	-124 G>A
P0418	P0418-S01	T1xG3	WT	E545K	WT	-124 G>A
	P0418-S02	TaxG3	WT	E545K	WT	-124 G>A
	P0418-S03	T1xG3	WT	E545K	WT	-124 G>A
P0468	P0468-S01	T1G2	WT	WT	WT	-124 G>A
	P0468-S03	TaG2	WT	WT	WT	-124 G>A
	P0468-S05	TaG2	WT	WT	WT	-124 G>A
P0533	P0533-S01	TaG2	WT	WT	*KRAS* G12V	-124 G>A
	P0533-S02	TaG2	WT	E542K	*KRAS* G12V	-124 G>A
	P0533-S3A	TaG2	WT	WT	*KRAS* G12V	-124 G>A
	P0533-S3B	TaG2	WT	WT	*KRAS* G12V	-124 G>A
	P0533-S04	TaG3	WT	E545K	*KRAS* G12V	-124 G>A
P0536	P0536-S01	TaG2	S249C	E545K	WT	WT
	P0536-S02	TaG2	S249C	E545K	WT	WT
	P0536-S03	TaG2	S249C	E545K	WT	-146 G>A
P0712	P0712-S01	TaG2	S249C	WT	WT	-124 G>A
	P0712-S02	TaG2	S249C	WT	WT	-124 G>A
	P0712-S03	TaG3	S249C	WT	WT	-124 G>A
	P0712-S04	TaG3	S249C	WT	WT	-124 G>A
	P0712-S05	TaG2	S249C	WT	WT	-124 G>A

Patient ID	Sample ID	Stage and Grade	*FGFR3* Mutation Status	*PIK3CA* Mutation Status	RAS Mutation Status	*TERT* Promoter Mutation Status
P0717	P0717-S01	TaG2	S249C	WT	WT	-124 G>A
	P0717-S02	TaG2	S249C	WT	WT	-124 G>A
	P0717-S03	T1G2	S249C	WT	WT	-124 G>A
	P0717-S04	TaG3	S249C	WT	WT	-124 G>A
	P0717-S05	T1G3	S249C	WT	WT	-124 G>A
P0926	P0926-S01	TaG1	WT	WT	WT	-124 G>A
	P0926-S03	TaG3	WT	WT	WT	-124 G>A
P0933	P0933-S01	TaG2	S249C	WT	WT	-124 G>T
	P0933-S02	TaG2	S249C	E542K	WT	-124 G>T
P0960	P0960-S01	TaG2	Y375C	WT	WT	-146 G>A
	P0960-S03	TaG1	Y375C	WT	WT	-146 G>A
	P0960-S04	TaG2	Y375C	WT	WT	-146 G>A
P0990	P0990-S01	TaG2	Y375C	E542K	WT	-124 G>A
	P0990-S04	TaG2	Y375C	E542K	WT	-124 G>A
	P0990-S05	T1G2	Y375C	E542K	WT	-124 G>A
P1175	P1175-S01	TaG2	G372C*	WT	WT	-124 G>A
	P1175-S02	TaG2	G372C	WT	WT	-124 G>A
P1326	P1326-S01	TaG2	S249C	WT	WT	-124 G>A
	P1326-S02	TaG2	S249C	WT	WT	-124 G>A
P1485	P1485-S01	TaG2	Y375C	WT	WT	-124 G>A
	P1485-S02	TaG1	Y375C	WT	WT	-124 G>A
	P1485-S03	TaG2	Y375C	WT	WT	-124 G>A
P1777	P1777-S01	T1G3	Y375C	H1047L	WT	-146 G>A
	P1777-S02	TaG3	Y375C	H1047L	WT	-146 G>A
P1870	P1870-S01	TaG3	S249C	WT	WT	-124 G>A
	P1870-S2A	TaG2	S249C	WT	WT	-124 G>A
	P1870-S03	TaG2	S249C	WT	WT	-124 G>A
	P1870-S05	TaG1	S249C	WT	WT	-124 G>A

Patient ID	Sample ID	Stage and Grade	*FGFR3* Mutation Status	*PIK3CA* Mutation Status	RAS Mutation Status	*TERT* Promoter Mutation Status
P2065	P2065-S01	TaG2	S249C	H1047R	WT	WT
	P2065-S02	TaG2	WT	WT	WT	WT
P2104	P2104-S01	TaG2	S249C	WT	WT	-124 G>A
	P2104-S02	TaG2	S249C	WT	WT	-124 G>A
P2161	P2161-S01	TaG2	S249C	WT	WT	-124 G>A
	P2161-S02	TaG2	S249C	WT	WT	-124 G>A
P2218	P2218-S1A	TaG2	R248C	WT	WT	-124 G>A
	P2218-S1B	TaG2	R248C	WT	WT	-124 G>A
	P2218-S02	TaG3	R248C	WT	WT	-124 G>A
P2291	P2291-S01	TaG2	S249C	WT	WT	WT
	P2291-S02	TaG2	WT	WT	WT	WT
P2329	P2329-S01	TaG2	S249C	E545K	WT	-124 G>A
	P2329-S02	TaG1	S249C	E545K	WT	-124 G>A
	P2329-S03	TaG2	S249C	E545K	WT	-124 G>A
P2440	P2440-S01	TaG2	S249C	WT	WT	-124 G>A
	P2440-S02	possible low grade UCC	S249C	WT	WT	-124 G>A

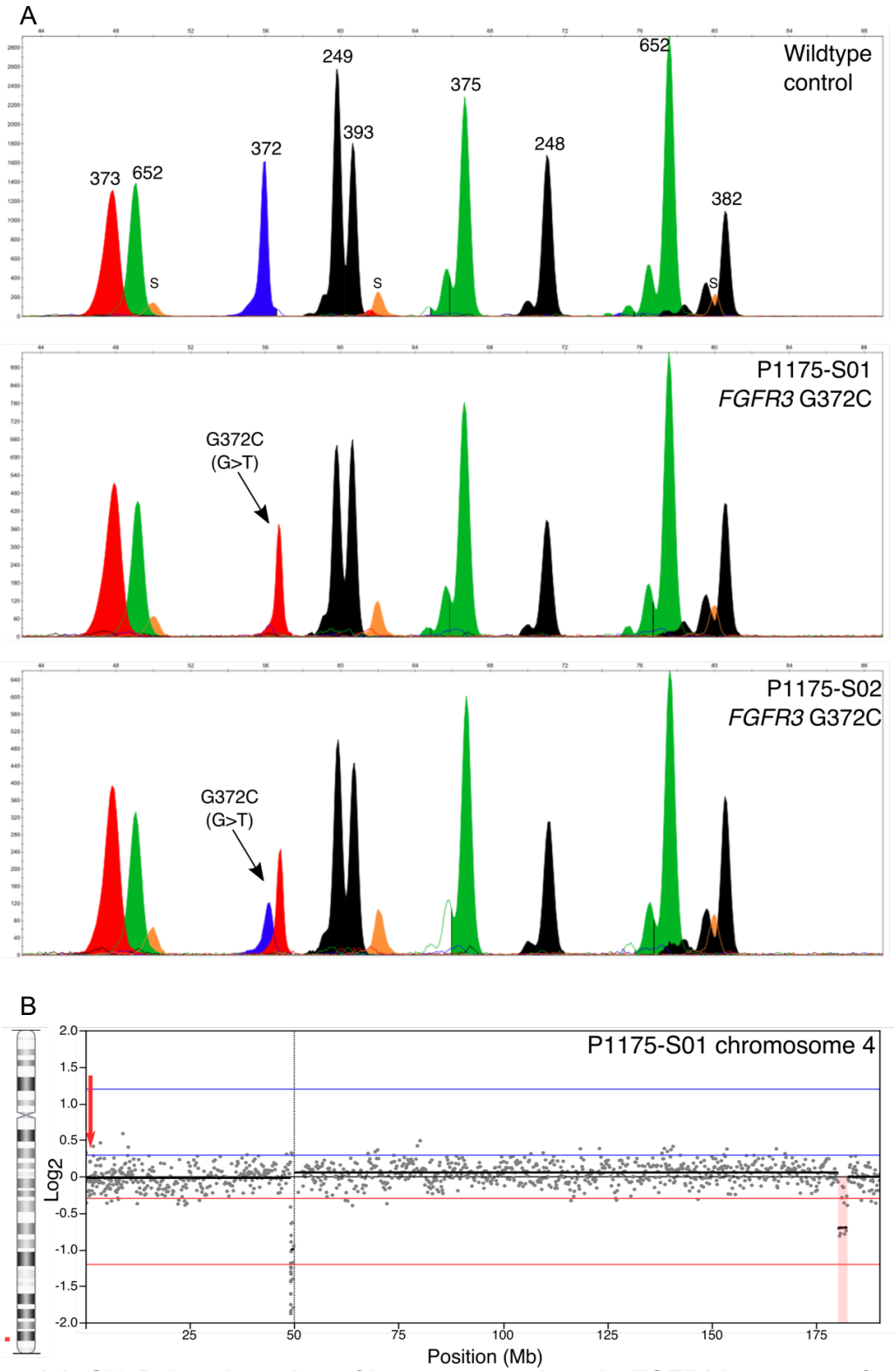


Figure 3.9: SNaPshot detection of hotspot mutations in *FGFR3* in tumours from P1175.

A) SNaPshot analysis of *FGFR3* in tumours from P1175 identified a G372C mutation that was homozygous in tumour P1175-S01 (middle) and heterozygous in tumour P1175-S02 (bottom). Top panel is a representative wildtype control trace. Bases are represented by the following colours: A = green; C = black; G = blue; T = red. Orange peaks (S) represent the internal Genescan-120LIZ size standards. B) Plot of CN data for chromosome 4 in P1175-S01. Approximate location of *FGFR3* is indicated by a red arrow.

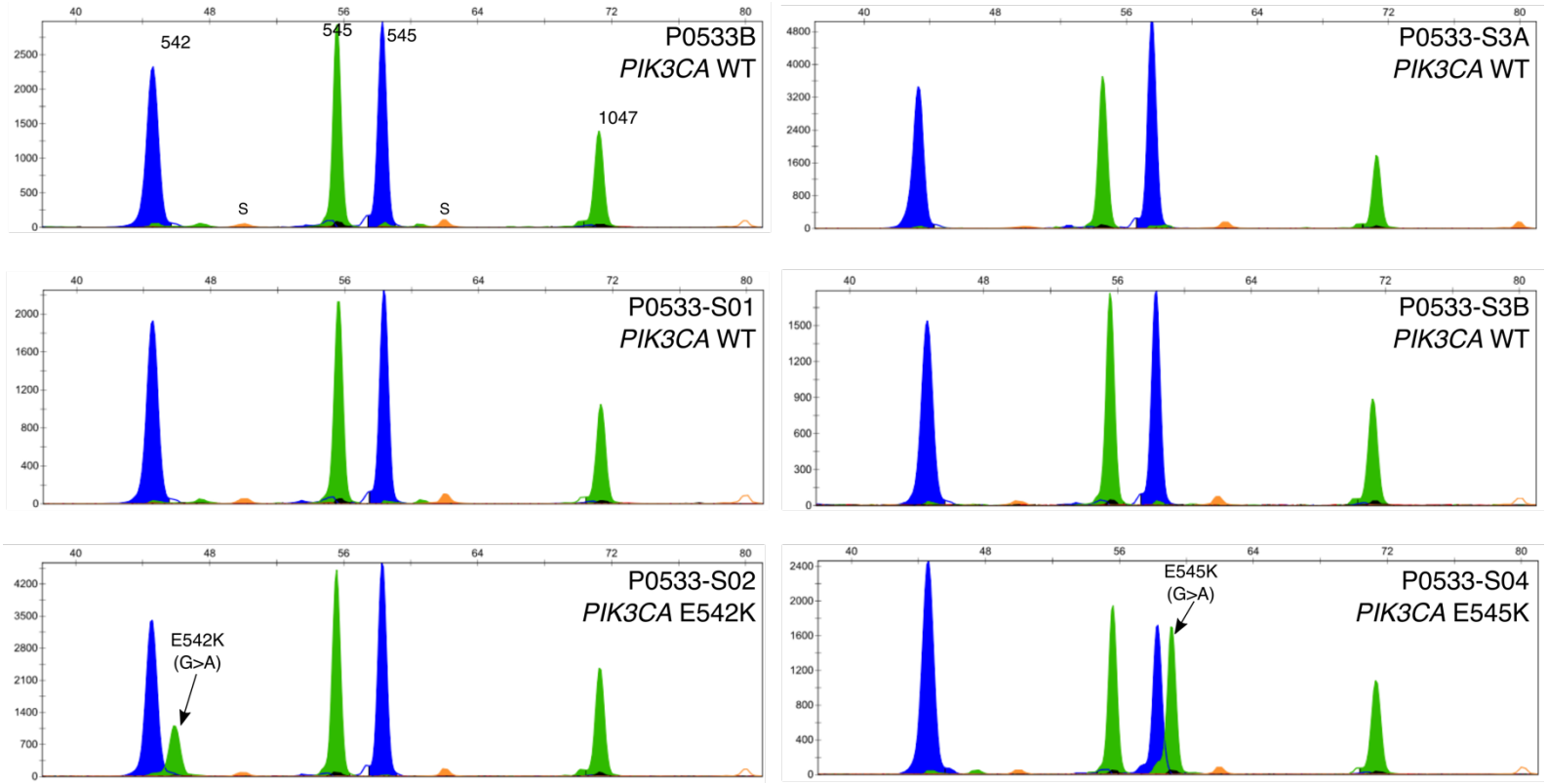


Figure 3.10: SNaPshot detection of hotspot mutations in *PIK3CA* in tumours from patient P0533.
Tumours from patient P0533 were assessed for hotspot mutations in *PIK3CA* using the SNaPshot assay. Two tumours (P0533-S02 and P0533-S04) contained different mutations whilst the other tumours from the patient were wild-type (WT). A WT profile was generated using the matched blood (P0533B) from the patient. Bases are represented by the following colours: A = green; C = black; G = blue; T = red. Orange peaks (S) represent the internal Genescan-120LIZ size standards

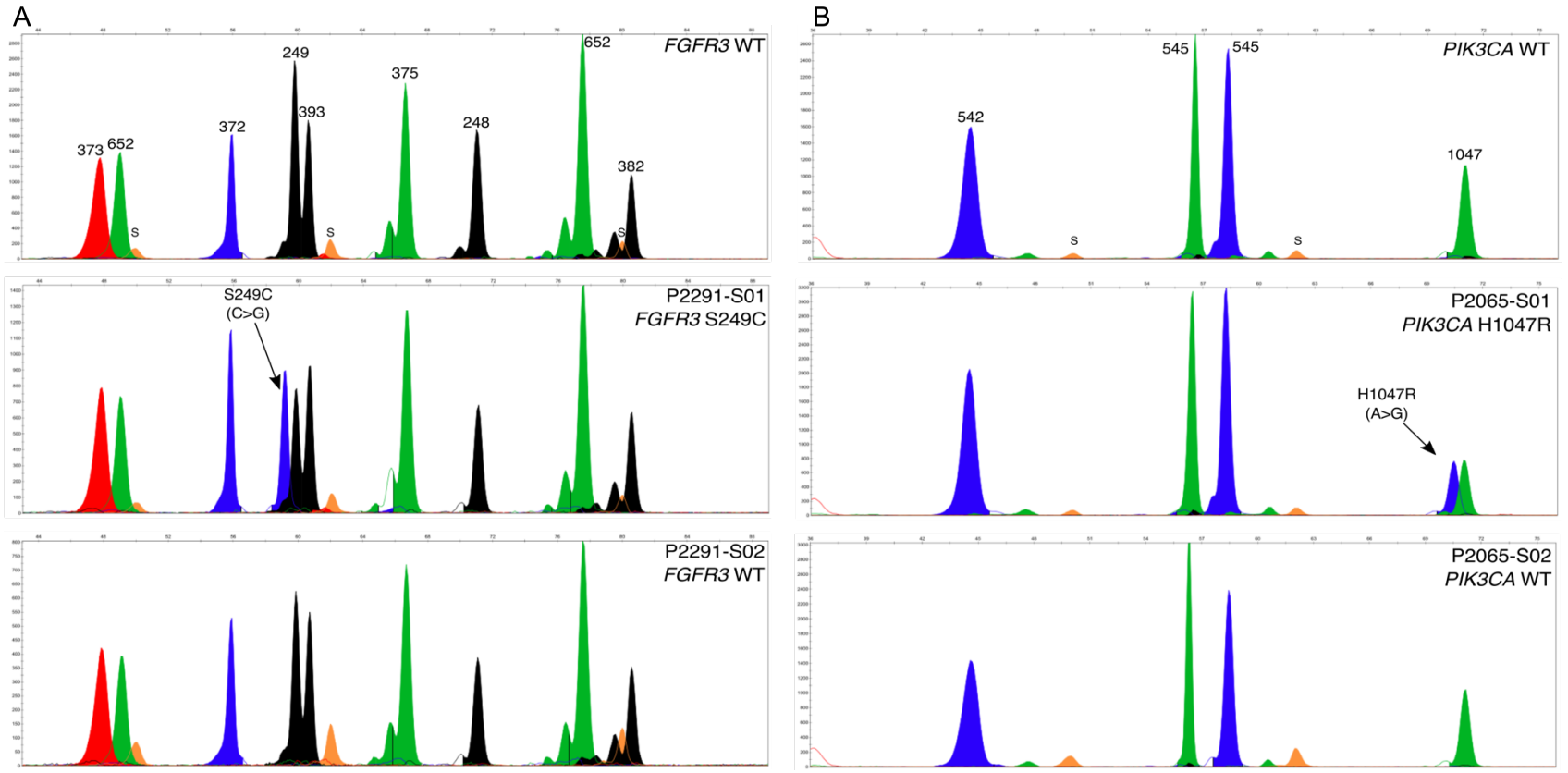


Figure 3.11: SNaPshot detection of hotspot mutations in *FGFR3* and *PIK3CA* in tumours from patients P2065 & P2291.
A) SNaPshot analysis of *FGFR3* in tumours from patient P2291. Top panel is a representative wild-type (WT) trace. An S249C mutation was identified in tumour P2291-S01 but not in tumour P2291-S02. B) SNaPshot analysis of *PIK3CA* detected a H1047R mutation in tumour P2065-S01 that was absent from tumour P0265-S02. Bases are represented by the following colours: A = green; C = black; G = blue; T = red. Orange peaks (S) represent the internal Genescan-120LIZ size standards.

3.2.7 CNAs and mutation status

Previous studies have assessed the relationships between the mutation status of key bladder genes and CNAs[125,259]. To investigate any relationships between mutation status and CNAs in our cohort, GWFPs were created for tumours according to the mutation status of *FGFR3, PIK3CA,* the *TERT* promoter and the RAS gene family (Figure 3.12).

Visual analysis of GWFP separating tumours according to *FGFR3* mutation status identified that *FGFR3* mutant tumours exhibit more losses of the whole of chromosome 9 (Figure 3.12A). In the *FGFR3* mutant tumours, 30/48 tumours had loss of the whole of chromosome 9 whilst only 3/19 tumours with wild-type *FGFR3* contained this alteration. Loss of *CDKN2A* was similar in both groups with 34/48 mutant tumours and 13/19 wild-type tumours containing this alteration. Gain of chromosome arm 8q appeared to be more frequent in *FGFR3* wild-type tumours than in mutant tumours (4/19 WT, 3/48 mutant), however this alteration was present only in a single patient (P0198).

Statistical comparisons undertaken using the Nexus software identified several regions exhibiting statistically significant differences in the frequencies of CNAs in *FGFR3* mutant and wild-type tumours. Interestingly a significant difference was identified at 4p16.3, containing the *FGFR3* gene. This region was gained in 29% of *FGFR3* mutant tumours and included tumours from 5 patients, whilst no wild-type tumours contained gain of this region ($p = 0.007$, Fishers exact test, differential threshold 25%). 32 significant regions of loss involving chromosome 9 were detected in the mutant tumours (p values ranging from 0.03 to < 0.0001). Visual inspection of the GWFPs identified that gain of the long arm of chromosome 8 was more frequent in *FGFR3* wild-type tumours compared to mutant tumours, but this frequency was not statistically significant at a 25% differential threshold. Two small regions on 8q that were identified as frequently gained in wild-type tumours compared to mutant tumours were located at 8q24.22 and 8q24.3 ($p = 0.004$ for both regions). Wild-type tumours also exhibited more frequent loss of 13q13.3-q14.3 ($p = 0.001$, Fishers exact test, differential threshold 25%).

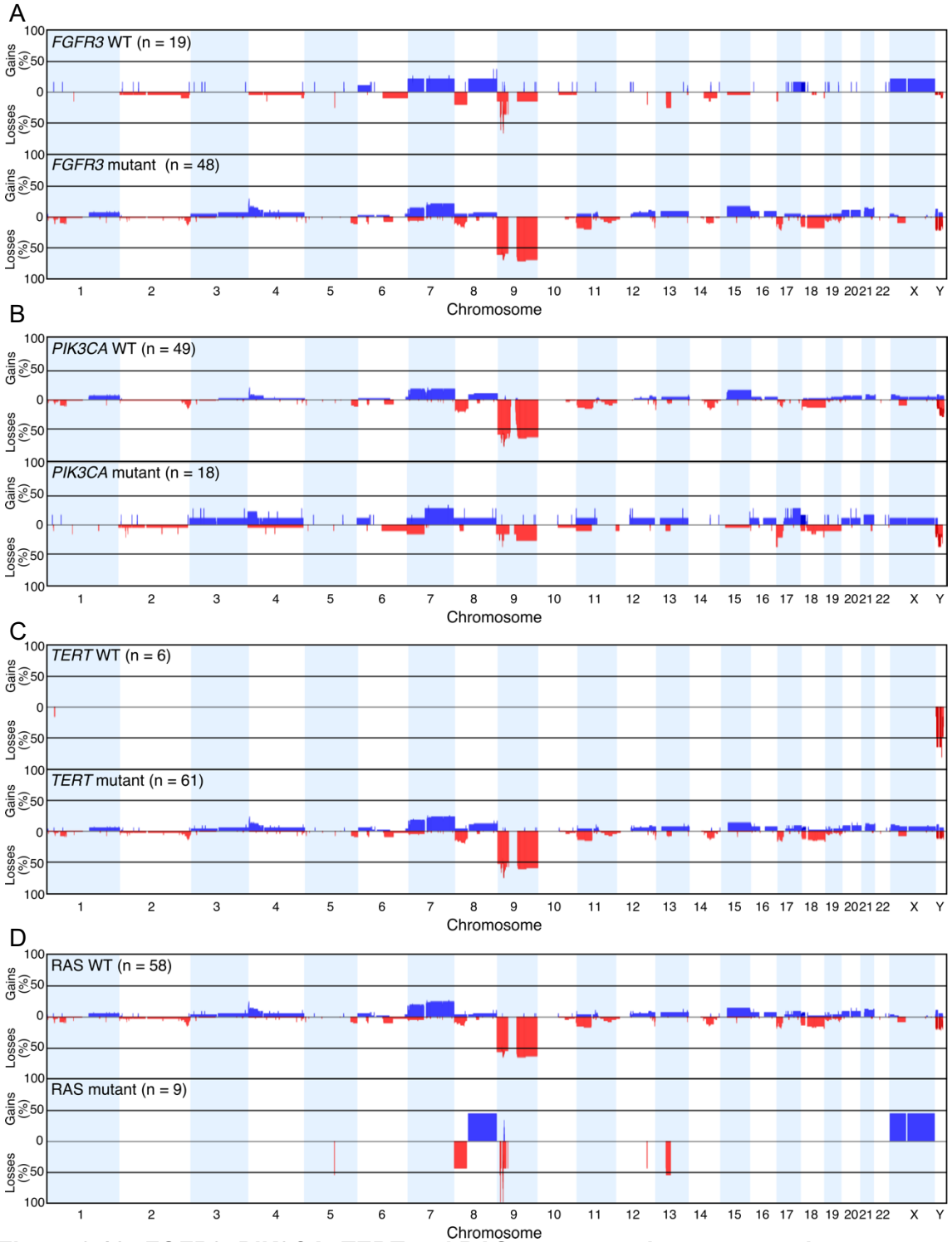


Figure 3.12: *FGFR3, PIK3CA, TERT* and RAS gene mutation status and genome-wide copy number alterations

Genome-wide frequency plots were generated for: A) *FGFR3* WT (*n*=19) and *FGFR3* mutant (*n*=48) tumours, B) *PIK3CA* WT (*n*=49) and *PIK3CA* mutant (*n*=18) tumours, C) *TERT* WT (*n*=6) and *TERT* mutant (*n*=61) tumours and D) RAS WT (*n*=58) and *RAS* mutant (*n*=9) tumours.

Analysis of FGA in mutant and WT tumours identified that *FGFR3* mutant tumours had a higher median FGA than WT tumours, but this was not statistically significant (wild-type median = 5.165 n = 19, mutant median = 7.015 n = 48, p = 0.2036, Mann-Whitney test). The higher FGA in the mutant tumours is likely due to the large number of mutant tumours with whole loss of chromosome 9. As these analyses were performed on multiple tumours from the same patient the results must be viewed with caution.

Separating tumours by *PIK3CA* mutation status identified several differences (Figure 3.12B). Loss of chromosome 9 was rare in *PIK3CA* mutant tumours with only 3/18 tumours showing loss of the whole chromosome. This is interesting as 13/18 of these tumours also contained *FGFR3* mutations, which was associated with chromosome 9 loss in this study. Statistical analysis identified 55 regions on chromosome 9 that were more frequently lost in wild-type tumours than in *PIK3CA* mutant tumours (p value range: 0.05 to 0.0002). Loss of 7p was identified only in *PIK3CA* mutant tumours but this was not significant at the differential threshold of 25%. Loss of 17p.13.3 – p13.2 and gain of 17q.22 – q25.3 were more frequent in mutant tumours (p = 0.0009 for both, Fishers exact test, differential threshold 25%).

Tumours lacking a *TERT* promoter mutation were seen to have very stable genomes with only one tumour (P2291-S01) exhibiting CNAs on any of the autosomes (Figure 3.12C). Of the 6 samples without a *TERT* mutation, 2 tumours (P2065-S02 and P2291-S02) did not carry any mutations in any of the hotspots assessed. In addition, these two tumours did not share any genomic alterations with the other tumours from the same patient. The remaining 4 tumours all had a *FGFR3* mutation (P0536-S01, P0536-S02, P2065-S01 and P2291-S01) and 3 tumours also had a *PIK3CA* mutation (P0536-S01, P0536-S02 and P2065-S01).

Analysis of GWFP from patients with RAS mutations showed that these tumours have few CNAs (Figure 3.12D). All RAS mutant tumours had HD of *CDKN2A.* Interestingly, for chromosome 9, none of the RAS mutant tumours exhibited loss of the entire chromosome or loss of an entire chromosome arm. Statistical analysis identified that all CNAs identified in the RAS mutant tumours, except for alterations on chromosome 9, were statistically more frequent in the mutant tumours. Caution must be taken in interpreting these results as only two patients had tumours with RAS mutations and all tumours in these patients shared the same CNAs. Nevertheless, the lack of 9q loss is interesting as this has been observed in other RAS mutant tumours (Carolyn Hurst, personal communication).

3.2.7.1 Inclusion of hotspot mutation data in the assessment of clonality

In section 3.2.5, tumours from each patient were assessed for clonal origins using CNAs. This identified a monoclonal origin for 15 of the 23 patients but could not infer a monoclonal origin for tumours from patients P0468, P0536, P0933, P0990, P2065, P2161, P2218 and P2291. It was thought that additional information from the hotspot mutation analysis may be able to assist in defining the relationships between tumours from these patients and confirm the monoclonal origin of the 15 patients identified as monoclonal by CN analysis.

Mutation analysis suggested a monoclonal origin for 6 of the patients for whom CN data alone had been insufficient (P0468, P0536, P0933, P0990, P2161 and P2218). However, due to the prevalence of many of the mutations in bladder cancer, this data alone cannot be used to completely assume monoclonality. Tumours from patient P0468 shared *TERT* promoter mutations (-124 G > A) but were wild-type for all other hotspot mutations. As mutations in the *TERT* promoter are highly prevalent in bladder cancer it is possible that these occurred independently, and therefore a monoclonal origin cannot be assumed in this patient from this data alone. Tumours from P0990 contained a *FGFR3* Y375C and a *PIK3CA* E542K mutation whilst tumours from patient P2218 shared a *FGFR3* R248C mutation. Tumours from both patients contained -124 (G > A) *TERT* promoter mutations. Combined, this suggests a monoclonal origin for these patients. Tumours from patients P0933 and P2161 carried an *FGFR3* S249C mutation as well as mutations in the *TERT* promoter. As these events are highly frequent in bladder cancer, it is again possible that these mutations developed independently. Tumours from patient P0536 shared *FGFR3* S249C and *PIK3CA* E545K mutations but tumour P0536-S03 contained an additional mutation in the *TERT* promoter (-146 G > A) not seen in the other two tumours. Again, these mutations are highly prevalent in bladder cancer and the presence of the *TERT* promoter mutation in a single tumour means that monoclonality cannot be absolutely confirmed. More extensive mutation analysis was carried out to confirm monoclonality in the tumours from these 6 patients; P2161 and P2218 underwent whole exome sequencing whilst the remaining tumours (P0468, P0536, P0933 and P0990) underwent NGS-based targeted sequencing using a gene panel of 140 genes identified as being frequently mutated in bladder cancer. The results of these analyses are discussed in Chapter 5 section 5.2.5.2.

No evidence for monoclonality was identified for two patients; P2065 and P2291. Both tumours from patient P0265 had losses on chromosome Y, but these were different:

P0265-S01 had loss of the whole chromosome whilst P2065-S02 had loss of only a segment of the chromosome. Neither tumour had any other CNAs so monoclonality could not be attributed using CNAs. Further evidence for possible oligoclonality comes from the fact that the two samples did not share any mutations in the hotspot regions targeted by the SNaPshot assays; P2065-S01 had mutations in both *FGFR3* (S249C) and *PIK3CA* (H1047R) whilst P2065-S02 had neither of these (Figure 3.11A). This evidence would suggest that the tumours are unlikely to be related. Unfortunately, P2065-S02 did not have enough DNA for NGS-based targeted exome sequencing to be carried out and therefore oligoclonality could not absolutely be confirmed. For patient P2291, each tumour had only one CNA and these were not shared. Tumour P2291-S01 also demonstrated an *FGFR3* S249C mutation that was not detected in tumour P2291-S02 (Figure 3.11B). These tumours were further assessed for monoclonality using targeted sequencing. The results of this analysis are discussed in Chapter 5 section 5.2.5.2.

3.2.8 Phylogenetic tree reconstruction

When analysing multiple samples from the same patient the identification of shared and unique alterations can predict the order of alteration acquisition. This can be used to identify events that occur early during the cancer evolutionary process versus events that occur later and may be tumour or region specific. To assess the temporal order of CNAs, and thus the relationship between tumours from the same patient, the TuMult algorithm was employed. TuMult uses the shared and unique breakpoints in tumours from the same individual to reconstruct the hypothesised sequence of events giving rise to each tumour. This information is then used to create phylogenetic trees. A phylogenetic tree was created for all patients using the scored CN data and this was supplemented with the mutation status of genes assessed using the SNaPshot assays. Any trees not shown in this section can be found in Appendix G. For some patients there were only a few CNAs shared by all tumours after which individual tumours diverged and acquired unique CNAs of their own. In other cases, the majority of CNAs were shared and there was very little individual divergence seen between tumours.

3.2.8.1 Linear and branching evolution

Tree reconstruction showed different predicted evolutionary trajectories amongst patients. Linear evolution was predicted in patients P0536, P0926, P0933, P0990, P1175, P1870, P2104 and P2440. Trees from these patients are depicted as straight lines, with tumours gaining more CNAs in each consecutive tumour (Figure 3.13A). Branching evolution was predicted in patients P0712, P0960 and P2218 where some

CNAs were common to all tumours, but each tumour also had unique events resulting in branching (Figure 3.13B). For some patients there appeared to be a mix of branching and linear evolution (P0418, P0533, P0712 and P1485). In patient P0712, for example, tumours P0712-S01, P0712-S02, P0712-S03 and P0712-S04 are predicted to branch off from common precursors, but tumour P0712-S05 is predicted to have developed linearly from tumour P0712-S04 due to the presence of a single additional CNA (+1q32) having been identified in this tumour (Figure 3.14A).

In some cases, all tumours from an individual were identical with respect to CNAs and hotspot mutations. For example, patient P0198 had 4 tumours, all of which shared identical CN events and hotspot mutations. Where there are no CNAs unique to an individual tumour TuMult depicts these as a linear tree in chronological TURBT order (Figure 3.14B&C). However, there is no way of knowing the order of events based on CNAs and hotspot mutation data alone in these cases.

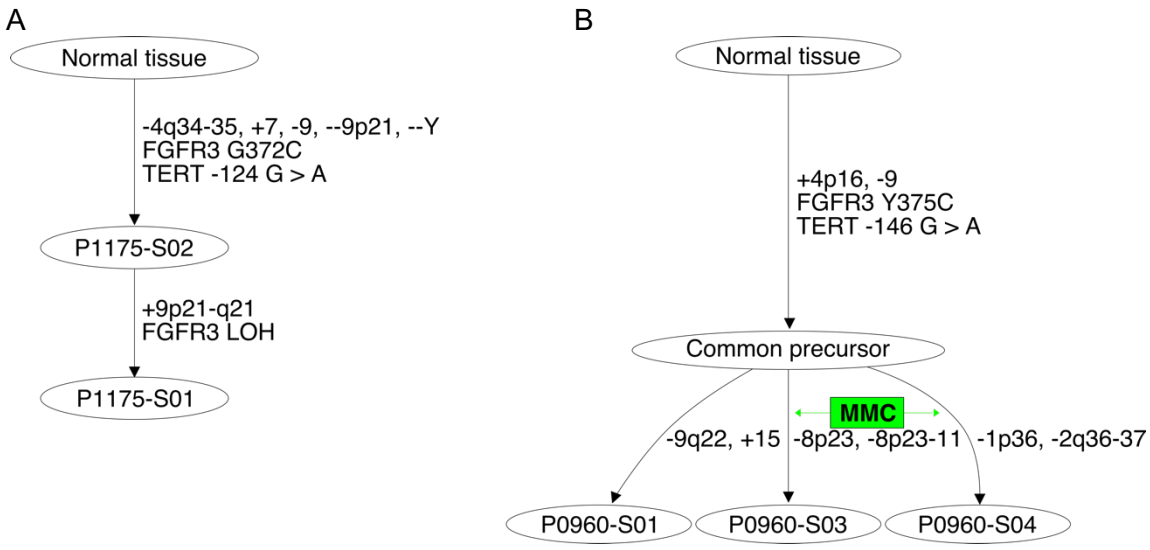


Figure 3.13: Phylogenetic trees for tumours from patients P1175 and P0960 demonstrating linear and branching evolution.

Trees were created using the TuMult algorithm which uses shared and unique breakpoints to reconstruct the evolutionary history of the tumours. A) Linear evolution was seen in tumours from patient P1175 according to CNAs and hotspot mutations. Tumour P1175-S01 is predicted to have evolved from tumour P1175-S02. B) Phylogenetic analysis of tumours from patient P0960 generates a branching tree. Tumours share some events then branch off with additional unique events.

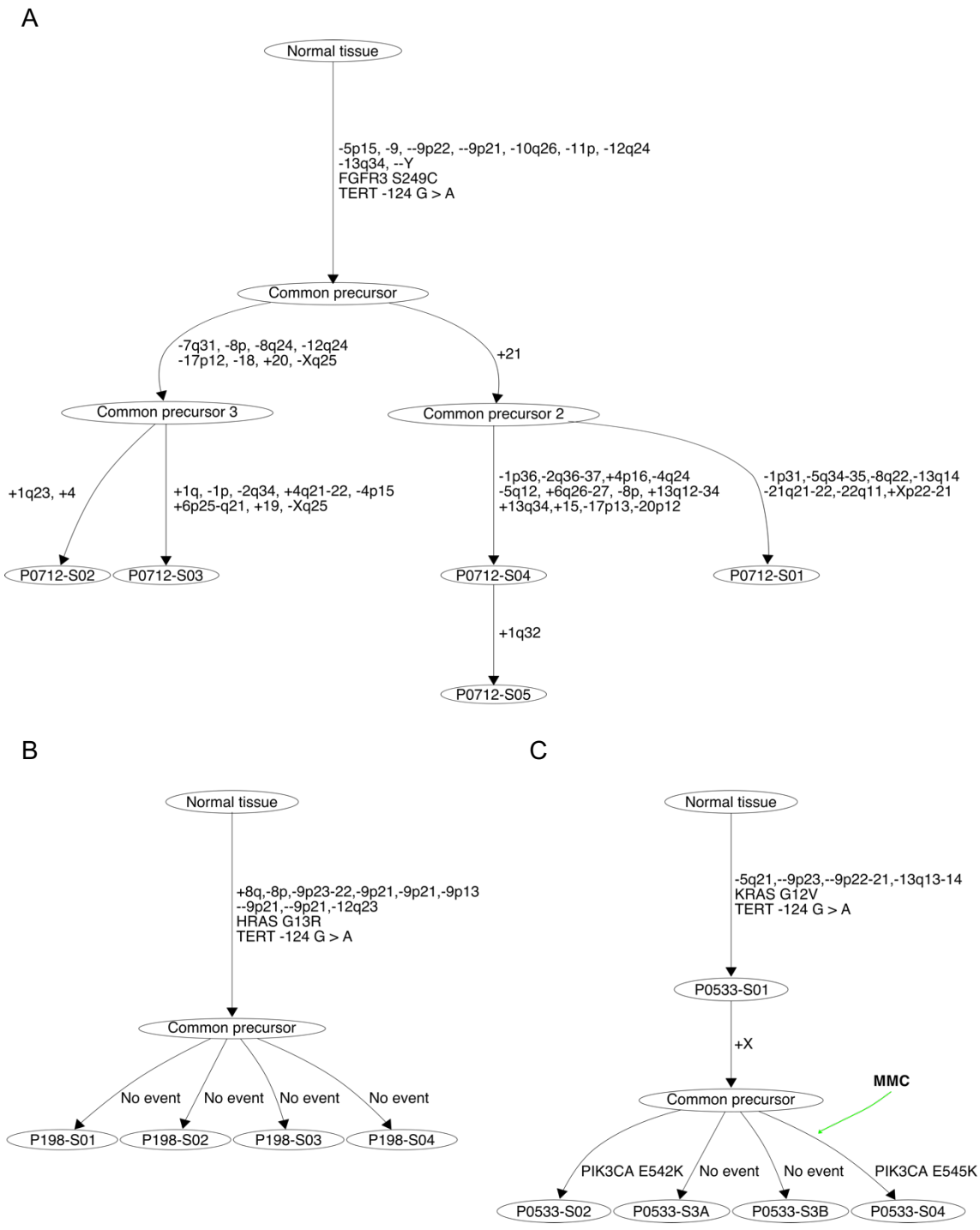


Figure 3.14: Phylogenetic trees for tumours from patients P0712, P0198 and P0533.

Trees were created using the TuMult algorithm which uses shared and unique breakpoints to reconstruct the evolutionary history of the tumours. A) Patient P0712 has some CNAs shared by all tumours followed by significant branching of the tumours. Tumour P0712-S05 is predicted to evolve in a linear fashion from tumour P0712-S04. B) Tumours from patient P0198 share all CNAs and hotspot mutations. There are no events unique to any tumour resulting in an uninformative tree. C) Tumours from patient P0533 share the majority of CNAs. However two different *PIK3CA* mutations are present in tumours P0533-S02 and P0533-S04. MMC treatment was not associated with any CNAs.

3.2.8.2 Chronology of tumour presentation

In some cases, the chronology of tumour presentation was not reflected in the order of genetic aberrations. In patients P0418, P0933, P1175 and P2104 the primary tumour was predicted to be a genomic descendant of the recurrent tumour and contained more CNAs. For P1485 the second tumour (P1485-S02) was predicted to be the first tumour in the tree with the primary and most recent tumours, P1485-S01 and P1485-S03, predicted to descend from this via a common precursor (Figure 3.15).

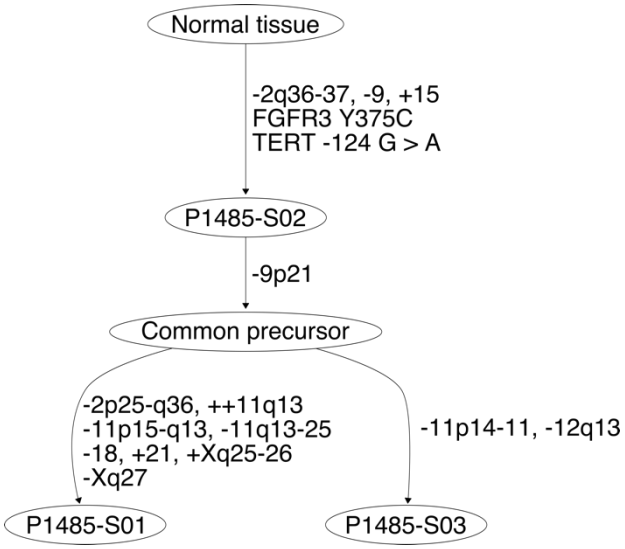


Figure 3.15: Phylogenetic tree for tumours from patient P1485.
The chronology of tumour presentation for tumours from patient P1485 does not reflect the order of genomic alterations. According to phylogenetic reconstruction, tumour P1485-S02 is genomically ancestral to tumour P1485-S01 despite being resected at a later date.

3.2.8.3 Recurrent trunk mutations

All trees were analysed for events occurring on the trunk of the tree. As patients P2065 and P2291 did not share any CNAs or hotspot mutations, and therefore had no truncal events, they were discounted from this analysis. The most frequent truncal event was a *TERT* promoter mutation which was seen on the trunk of the tree in 20/21 patients. Indeed, almost all hot-spot mutations were found on the trunk of the trees with the exception of one *TERT* promoter mutation identified in patient P0536 and the *PIK3CA* mutations identified in two of the tumours from patient P0533. *FGFR3* mutations were always predicted to be on the trunk of the tree.

CN losses were more common on the trunk of trees than gains. Alterations involving chromosome 9 were the most common CN events occurring on the trunk of the tumour trees (Table 3.3). Where patients had loss of the whole of chromosome 9, this always occurred on the trunk of the tree, suggesting that this may be an early event in bladder cancer development. This early event was identified in almost half of the patients (10/21). Only 3 regions of gain were seen on two or more trunks. Gain of the entire q-arm of chromosome 15 was seen on the trunk in two tumours; P1326 and P1485.

3.2.8.4 Mitomycin C treatment and tumour evolution

Sixteen of the patients in the cohort underwent a six-week course of MMC treatment (P0418, P0468, P0533, P0536, P0717, P0926, P0960, P0990, P1175, P1485, P1777, P1870, P2161, P2218, P2329 and P2440). Comparing the frequencies of CNAs in matched pre- and post-MMC treated tumours did not identify any regions of CNA that were differentially altered between the two groups. Likewise, no statistical difference in the FGA was identified between pre- and post-MMC tumours. We were interested to see whether placement of MMC treatment on the phylogenetic trees was associated with any recurrent alterations (Figure 3.16).

The predicted genomic evolution did not follow the chronology of presentation for two patients (P1175 and P1485), consequently it was not possible to indicate MMC treatment on the trees for these patients. For three patients, no changes in CNAs were identified in the post-MMC tumours (P1870, P1777 and P2161). In other patients, treatment with MMC coincided with a predicted branching event (Figure 3.16). CNAs that were uniquely detected post-MMC treatment or CNAs that were detected pre-MMC treatment but not post-MMC treatment were assessed for recurrent events. Tumours from 3 patients (P0468, P0926 and P2440) showed a loss of 9p21 containing *CDKN2A/B* after the MMC course compared to the tumours pre-MMC. For patients P0468 and P2440, tumours resected prior to MMC treatment contained loss of 9p21 whilst post-MMC tumours contained HD of this region. For patient P0926, loss of 9p21 was identified only in the post-MMC tumour.

Table 3.3: Recurrent events seen on the trunk of the phylogenetic trees.
Events seen on the trunk of the tumour trees were examined and events occurring in more than one patient have been listed in the table. *The overall frequency is calculated from the 21 patients with a trunk event.

Event	Number of patients (n=21)	Overall frequency (%)*
Mutations		
TERT	20	95
FGFR3	16	76
PIK3CA	4	19
RAS	2	9.5
Losses		
-9	10	47.6
--9p21 *(*HD *CDKN2A)*	7	33.3
-9p21 *(CDKN2A* loss*)*	3	14
-9p22	3	14
-11p	3	14
-Y	3	14
-6p21	2	9.5
-9p23	2	9.5
-9q	2	9.5
-12q24	2	9.5
-17p	2	9.5
Gains		
+7q11	2	9.5
+15	2	9.5
+16p13	2	9.5

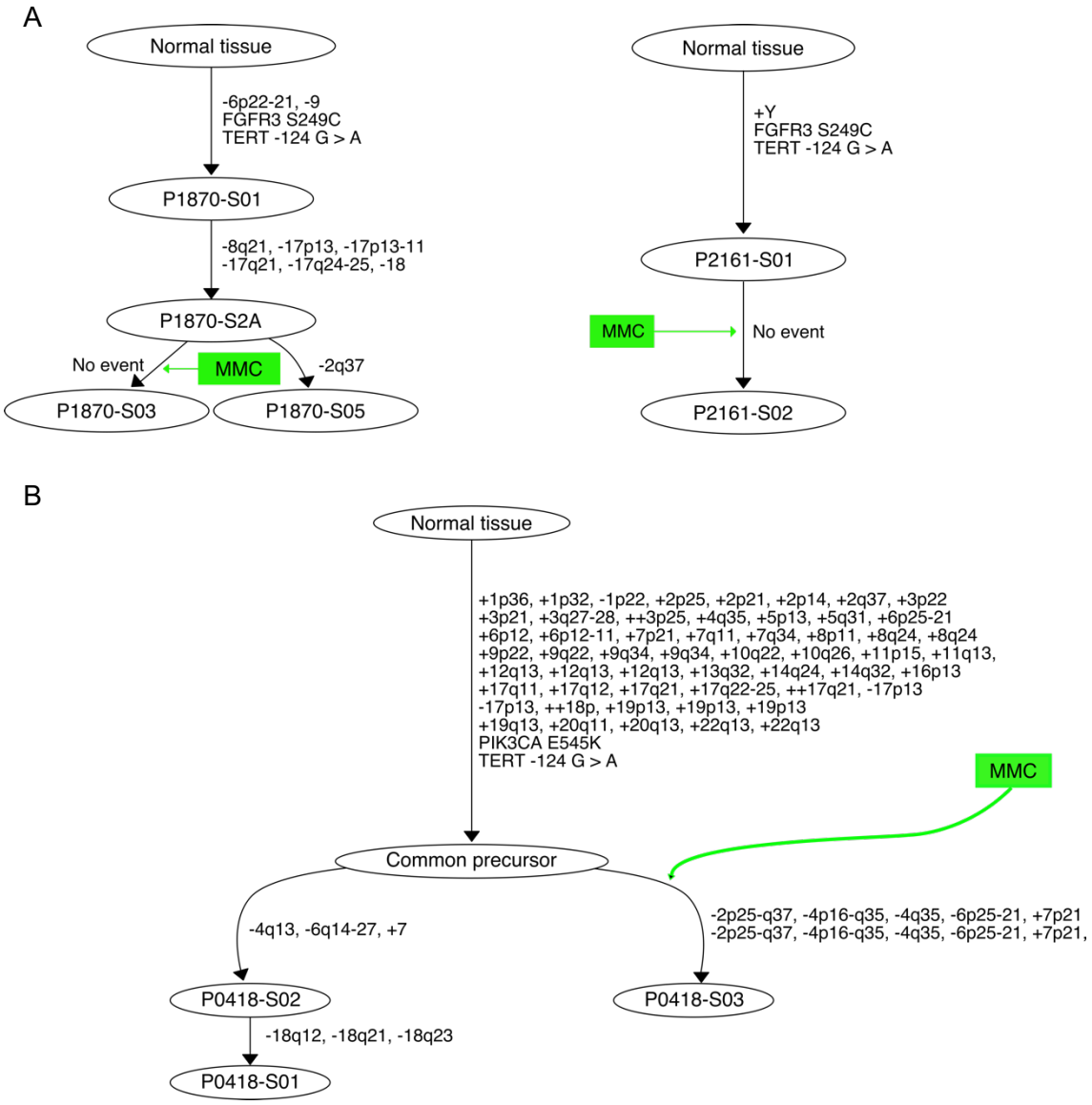


Figure 3.16: Examples of MMC positioning on phylogenetic trees reconstructed for tumours from patients P1870, P2161 and P0418.

A) MMC treatment was not associated with any changes in CN for patients P1870 and P2161. B) Tumours from patient P0418 were punctuated with many focal regions of CN gain that were shared between all tumours. MMC treatment coincides with tumour P0418-S03 branching off the common ancestor.

3.2.9 Hierarchical cluster analysis

In previous studies, hierarchical clustering of multiple tumours from the same patient has demonstrated that these tumours tend to cluster close together[163,167]. In our cohort, the majority of tumours from the same patient share CNAs, therefore we hypothesized that these would likely cluster together. One-way unsupervised hierarchical cluster analysis of copy number data from all 67 tumours from this cohort was performed. This identified two main subgroups of tumours; one characterised by loss of chromosome 9 and/or gain of chromosome 7, whilst tumours in the other group lacked these alterations (Figure 3.17). For the majority of patients, tumours tended to cluster next to each other. Seven patients (P0468, P0712, P0933, P960, P0990, P1485 and P1870) had one tumour cluster separately to the other tumours from that patient. This was due to CNAs not being shared by all tumours from an individual patient. For example, tumour P1870-S01 clustered separately to all the other tumours from this patient. This tumour did not contain loss of chromosome 18 which was identified in all other tumours from this patient. Similarly, tumour P0960-S01 clustered away from the other two tumours from this patient, likely due to a gain of chromosome 7 identified in tumour P0960-S01 but absent from the other tumours from that patient. In all seven cases, tumours clustered into the same main subgroup as the other tumours from the patient (Figure 3.17).

In NMIBC, genomic subtypes have been identified using hierarchical clustering analysis of CNAs[82,125]. These studies identified two genomic subgroups of stage Ta tumours; one (GS1) characterized by no or few CN alterations and the other (GS2) characterized by loss of chromosome 9. We were interested to identify which of these subgroups tumours from our cohort would cluster with.

One-way unsupervised hierarchical cluster analysis of copy number data from all 67 tumours from this cohort and 133 of the tumours from Hurst *et al.*[82] was performed. This produced the two main clusters (GS1 and GS2) as described by Hurst *et al.*[82] (Figure 3.18). The majority of tumours from our cohort clustered with the GS2 tumours. Only 11 tumours clustered into GS1, including all the tumours lacking a *TERT* promoter mutation. These tumours were all stage Ta grade 2 that were chromosomally very stable with few or no CNAs. The remaining 56 tumours clustered into GS2.

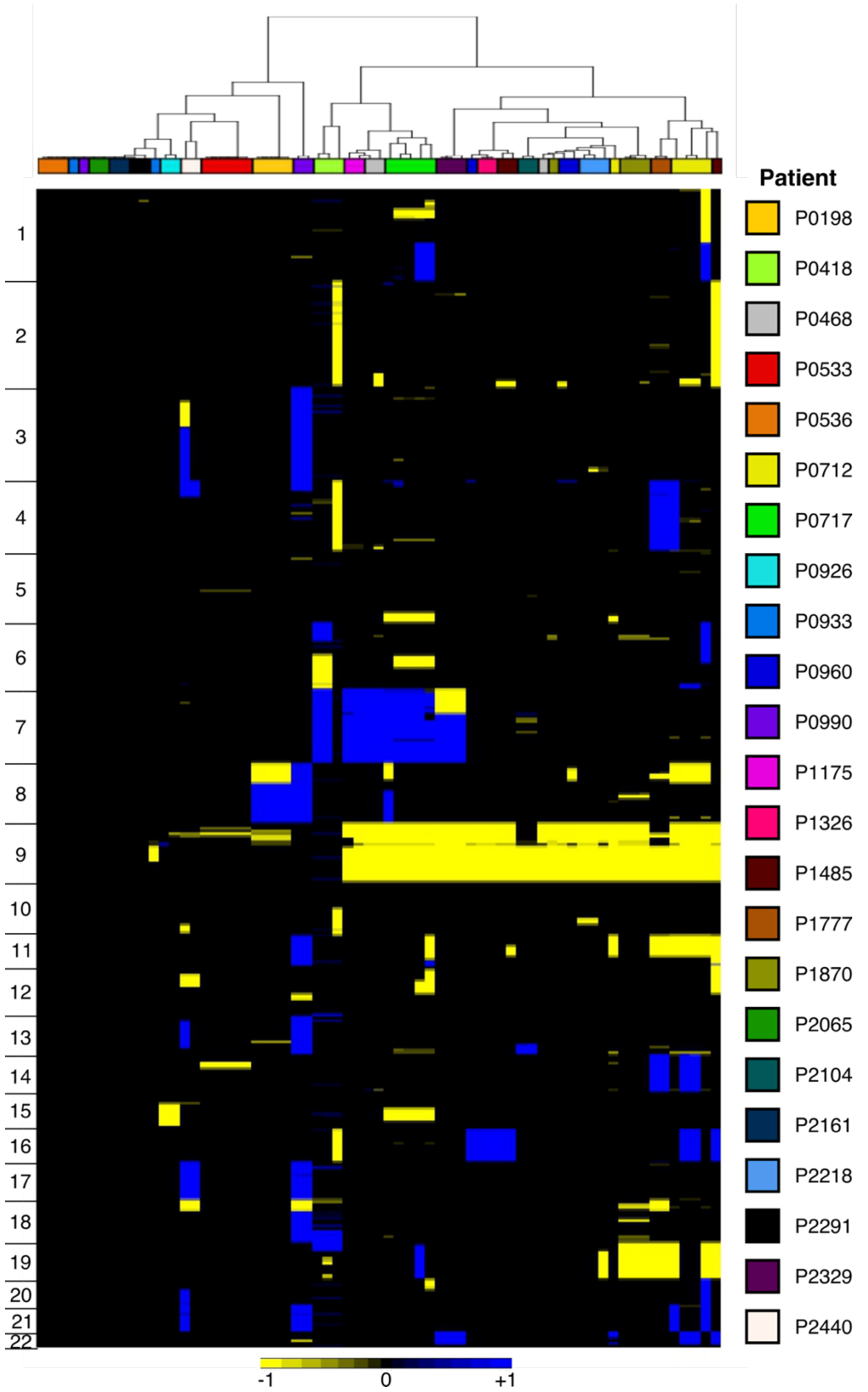


Figure 3.17: Unsupervised hierarchical cluster analysis
Unsupervised hierarchical cluster analysis of CN data from all 67 tumours. Columns represent tumours and rows genomic position. Blue shading is CN gain, yellow is CN loss and black represents no CN change. Patient ID for each tumour is designated by colour.

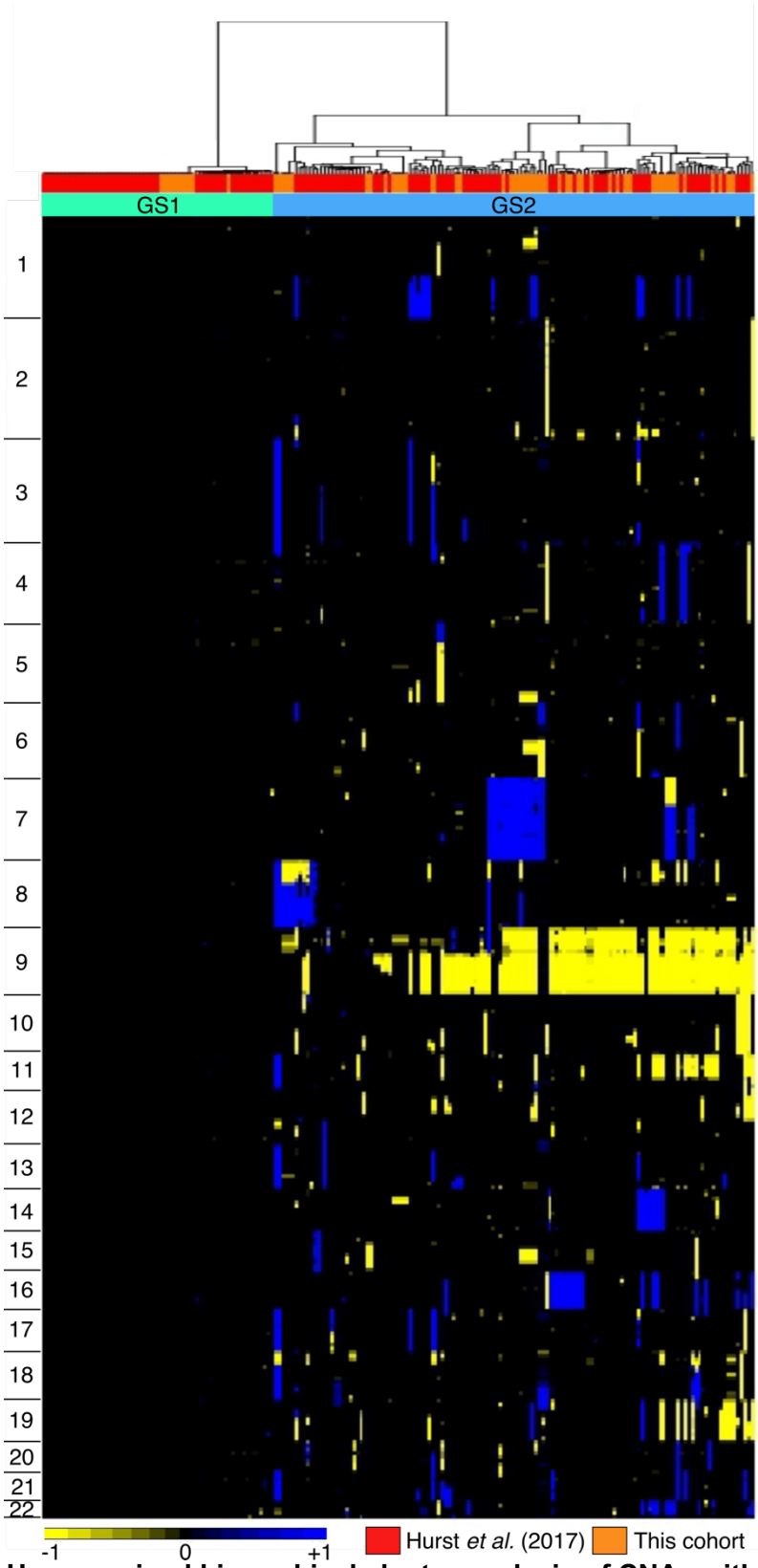


Figure 3.18: Unsupervised hierarchical cluster analysis of CNAs with tumours from Hurst *et al.*[82].

Unsupervised hierarchical cluster analysis of CN data from the 67 tumours from this cohort and 133 samples from Hurst *et al.*[82]. Columns represent samples and rows genomic position. Blue shading is CN gain, yellow is CN loss and black represents no CN change. Red bars indicate samples from Hurst *et al.* and orange bars represent samples from the current cohort of 67 tumours. The two main clusters from Hurst *et al,* GS1 and GS2, are indicated.

Two patients had tumours that clustered into different genomic subgroups (P0933 and P0990). For patient P0933, tumour P0933-S02 clustered into GS1 whilst tumour P0933-S01 clustered into GS2. Both tumours demonstrated stable genomes with no CNAs identified in P0933-S02 and only one CNA, loss of 9q21.11 – q22.33, was identified in P0933-S01. In this case, the loss on chromosome 9 led to tumour P0933-S01 clustering into GS2. For patient P0990, tumour P0990-S01 had no CNAs and clustered into GS1 whilst tumours P0990-S04 and P0990-S05 clustered into GS2. Interestingly, neither P0990-S04 or P0990-S05 contained loss of chromosome 9 or 9q, the characteristic losses of this subgroup, but these two tumours were more chromosomally unstable, containing several whole chromosome gains.

3.3 Discussion

Shallow-pass whole genome sequencing was used to assess CNAs in 67 tumours from 23 patients. Using multiple tumours from one individual patient can make it difficult to assess important CN changes as many of the tumours from an individual patient will share CNAs. This may artificially inflate the frequency of certain events. Likewise, if only one tumour per patient is assessed important events may be missed. To compensate for this GWFPs were created for all tumours as well as just the first or last tumour from each patient. These showed very similar CNA profiles with no significant difference in the frequencies of CNAs identified. Additionally, analysis of the FGA demonstrated that there no statistically significant differences between the first and last tumours from each patient. This suggests that bladder tumours are relatively stable at the CN level between recurrences.

Tumour stage and grade are key parameters used in the prediction of risk of recurrence and progression for NMIBC[182,260]. This cohort consisted of: 8 stage Ta G1, 42 stage Ta G2, 7 stage Ta G3, 4 stage T1 G2, 2 stage T1 G3, 1 stage Ta(x) G3, 2 stage T1(x) G3 and 1 possible low-grade urothelial carcinoma in which small, heavily diathermied fragments of tissue made grading difficult. Tumours with an (x) had insufficient sampling of the muscle layer and therefore invasion could not be ruled out. The one stage Ta(x) tumour was therefore omitted from the comparison of tumours by stage as it is possible that this tumour may have been of a higher stage.

Separating tumours according to stage and grade identified that stage T1 tumours and grade 3 tumours are more chromosomally unstable than stage Ta or grade 1&2 tumours. This is in agreement with previous observations[125,126]. We also used the fraction of genome altered (FGA) as a measure of chromosomal instability and this

showed that NMIBC with a higher stage or grade had a significantly higher median FGA (Mann Whitney U test, $p<0.005$ and $p<0.05$ for stage and grade, respectively). The number of tumours in each group was very different with the majority of samples being stage Ta grade 2 and this may have biased the results. However, the data presented here shows similarities to that of Hurst *et al,*[125] who looked at over 100 NMIBCs and also found that higher stage and grade tumours had a higher median FGA.

Statistical comparisons of the frequencies of CNAs in tumours of stage Ta and T1 identified many differences. Gains of 7p, 7q, 8q12.3 - q24.3 and 17q22 – q25.3 were more frequent in stage T1 tumours. Gains of 8q were reported as statistically more frequent in stage T1 tumours compared to stage Ta tumours in a study by Richter *et al.*[126], who also identified deletions of 2q, 8p and 11p and gains of 1q, 3p, 3q 5p, 6p and 10p as significantly different in frequency between stage Ta and stage T1 tumours. Gains on 3p, 3q and 5p were also significantly more frequent in our stage T1 tumours, but the size of these regions were much smaller than those previously identified[126]. Again, these differences may be due to the low number of stage T1 tumours included in this study.

Separation of CNAs in tumours according to grade showed a similar distribution to that generated when tumours were separated according to stage. Of the 8 stage T1 tumours, 4 were G3 and 4 were G2. Interestingly, the chromosome 7 gains seen in stage T1 tumours were equally divided between G2 and G3 tumours. Whilst this could suggest that this alteration may be associated with stage it is important to note that chromosome 7 gains were present in 4 stage T1 tumours from 2 patients only, P0717 and P0418. Analysis of the other tumours from these patients identified the presence of chromosome 7 gains in all tumours making it more likely to be a patient specific alteration rather than being associated with stage and grade. Within the literature there are conflicting reports on the relationship between chromosome 7 gain and stage and grade[261,262]. A correlation between chromosome 7 CN and increasing stage and grade was identified on one study[261] but in other studies, this correlation has not been observed[135,262].

The most common CN event identified was loss of the *CDKN2A* locus at 9p21.3. This is a region that is commonly deleted in bladder cancer and has previously been associated an increased risk of recurrence in NMIBC[132]. In a recent review by Knowles and Hurst[4], hemizygous deletion of *CDKN2A* was reported at 50-60% in both low stage Ta tumours and MIBC (≥T2) and HD was reported as being present in 15% of low

grade Ta tumours and 20-30% of MIBC. In our cohort of 67 tumours, 31 tumours had HD of *CDKN2A;* 22 TaG2, 5 TaG3, 2 T1G2 and 2 T1G3. On a per-patient basis, 10 patients had HD of this region in at least one of their tumours corresponding to 43.5% of patients overall. This is a higher rate of HD than previously reported for either NMIBC or MIBC. When assessing only the first and last tumours from each patient, HD of *CDKN2A* was present in 34.7% and 43.5% of tumours, respectively, suggesting that the high level of *CDKN2A* loss is reflective of this group of tumours, and not because of multiple recurrences containing the same CN event. This data therefore further supports the association between *CDKN2A* HD and increased risk of recurrence[132].

Meeks *et al*[173] identified an increased frequency of loss of the *CDKN2A* locus in high-risk NMIBC patients that progressed compared to non-progressors, however this was non-significant (p=0.54). They suggested that loss of the *CDKN2A* locus may be a late event during invasion and may be important for progression. In our cohort, where loss of *CDKN2A* occurred in a patient it was usually present in all tumours from that patient which may suggest that it might actually be an early event. Indeed, loss of *CDKN2A* (either as a focal loss or part of a whole chromosome loss) was seen on the trunk of the phylogenetic trees in 14 out of 15 patients. HD of the region was a trunk event in 7/10 patients. This may suggest that whilst loss of *CDKN2A* is an early event, HD of the region can occur later on and this may be the important event for progression. Our study looked specifically at recurrent tumours, with the majority of patients being of an intermediate risk of progression. It is therefore unsurprising that loss of *CDKN2A* was a common event.

In the current study, two patients' (P0926 and P0990) had progression from stage Ta to stage T1 disease and two patients (P0717 and P2440) later progressed to metastatic disease (see patient timelines in Appendix A). For patient P0926, loss of *CDKN2A* was identified only in the recurrent tumour. In this case the initial tumour was stage Ta grade 1, whilst the recurrence analysed was stage Ta grade 3. However this patient also had a T1G3 tumour between these two which did not have enough DNA for analysis. There was no loss of *CDKN2A* observed in any of the tumours analysed from patient P0990, but both patients with metastatic disease showed HD of *CDKN2A;* for patient P0717 this was present in all tumours from the patient whereas for patient P2440 only the recurrent tumour showed HD of this region. Whilst these observations are interesting, this study does not include enough patients with disease progression to confirm the association between *CDKN2A* loss and progression reported by Meeks *et al.*[173]. Interestingly both patients with metastatic disease had upper tract disease;

patient P0717 additionally had transitional cell carcinoma of the right renal pelvis and patient P2440 had bilateral nephrectomies due to transitional cell carcinoma of the renal pelvis. This could be an alternative route for progression. Clonality between bladder tumours and tumours in the renal pelvis has been identified by NGS[174]. It would be interesting to study these upper tract tumours to see if these were also clonally related to the bladder tumours.

CDKN2A HD has been associated with muscle invasion in *FGFR3*-mutated urothelial BC[263]. This could be a progression pathway for *FGFR3* mutant tumours. In this cohort, out of the 31 tumours with *CDKN2A* HD, 20 tumours from 7 patients (P0712, P0717, P1175, P1485, P1777, P2218 and P2440) carried an *FGFR3* mutation, including the two patients that progressed to metastatic disease. It would be interesting to analyse the metastatic tumours (as well as the upper tract tumours) to see if *CDKN2A* HD was present. Again, the lack of patient numbers makes it impossible to draw any firm conclusions concerning *CDKN2A* loss, *FGFR3* mutation and progression.

Amplifications detected in the current study were distributed throughout the genome whilst all HD events except for one were located on chromosome 9. These were all focal events ranging from 0.6-2.9 Mb in size except for one broad region of amplification of a 14Mb region on chromosome 18 seen in tumours from patient P0418. It is thought that focal CNAs are the product of errors in DNA repair whilst broader CNAs are the result of incorrect chromosome segregation during mitosis[264]. Focal amplifications in stage Ta BC have been associated with high-grade and recurrence[250,265] whilst HD are more frequent in patients with multiple tumours[250]. In this data set amplifications were only found in 5 tumours; 1 TaG2, 1 Ta(x)G3, 1 T1G2 and 2 T1(x)G3.

A single tumour from patient P1485 contained an amplification of 11q13.3, tumour P1485-S01. This is a region that is recurrently amplified in bladder cancer[135,250,266] and contains the Cyclin D1 (*CCND1*) gene as well as fibroblast growth factors (*FGFs*) 3, 4 and 19. CCND1 is often found to be dysregulated in cancer and forms part of an active complex that can phosphorylate the retinoblastoma (RB1) tumour suppressor protein, driving progression to S-phase[267]. The FGFs are the activators of *FGF* receptors, including FGFR3, and have roles in proliferation, migration, differentiation, angiogenesis and wound healing[268], all of which are frequently dysregulated in cancer. Amplification of this region has been shown to correspond with increased expression of *CCND1* mRNA but not *FGFs* despite the genes being within the same amplicon[265]. It

would be interesting to follow this up with further analysis of these candidate genes at the mRNA and/or protein levels.

P0468-S01 exhibited an amplification at 13q33.3 and part of this region was also gained in tumour P0468-S03. The region of gain in P0468-S03 included 3 of the 12 genes present in the amplified region seen in tumour P0468-S01 (*LINC00676*, *IRS2* and *COL4A1*). This could suggest that one, or all, of these three genes are the target(s) for this amplification. The concomitant focal deletions around this amplification in tumour P0468-S01 are suggestive of significant rearrangement occurring in this region. It is possible that this region was gained in a clone, then in a subclone there was significant chromosomal rearrangement leading to amplification of the region with deletions at the breakpoints. This potentially gave the subclone a survival or growth advantage, thus making it the first tumour to be detected. Of the 12 genes located in the amplification, insulin receptor substrate 2 (*IRS2)* is an interesting candidate. *IRS2* is an intracellular signalling adaptor protein involved in insulin signalling which can activate both the PI3K and ERK signalling pathways[269]. *IRS2* has been identified as a candidate driver oncogene in colorectal cancer[270] and overexpression plays a role in many cancers[269]. Analysis of the TGCA MIBC dataset[81] in COSMIC identified 7 patients with a gain of this gene and further inspection identified that this gain was actually a focal amplification in 3 of the patients, according to our criteria. However, *IRS2* was only found to be overexpressed in one of these patients (data not shown). It would be interesting to see if *IRS2* is indeed upregulated in the tumours from patient P0468 as this might implicate *IRS2* as a driver event in bladder cancer.

The majority of HD events were located on chromosome 9, with the *CDKN2A* locus being the most commonly deleted region as discussed previously. In patient P0533, a region containing protein tyrosine phosphatase delta (*PTPRD*) was focally deleted in all tumours in a separate event to HD of *CDKN2A*, which is on the same chromosome arm (9p) less than 12 Mb away. *PTPRD* is a tumour suppressor gene that is frequently inactivated in glioblastoma and other cancers[271,272]. This gene was also identified as a target of HD in bladder cancer in a recent study by Beoth *et al.*[250]. They investigated solitary, multifocal and recurrent tumours and identified that HD of *PTPRD* occurred more frequently in patients with multiple tumours as did HD at 9p22.3-p22.2, a region harboring the *C9orf93, BNC2* and *CNTLN* genes. This HD at 9p22.3-p22.2 was also identified in our cohort in all five tumours from patient P0712. Combined, these data suggest a role for these regions in multifocal and recurrent disease and warrant further investigation.

A region on chromosome X containing the genes *XIAP, STAG2, SH2D1A* and *TENM1* was HD in tumour P0712-S03 and there was CN loss in the same region in tumour P0712-S02. Of these genes, *STAG2* has been seen as frequently inactivated by mutations in bladder cancer[273,274]. *STAG2* forms part of the cohesin complex which is required to ensure correct chromosome segregation by maintaining sister chromatid cohesion following DNA replication[275]. Deletion of Xq25, the region that contains *STAG2*, has been reported in other cancers[276,277] and inactivation of this gene by mutation has been shown to cause aneuploidy[277]. Correlating with this, the two tumours with loss of *STAG2* have a higher FGA than the average in this cohort (18% in P0712-S02 and 24% in P0712-S03 compared to 6.08% median overall). Studies investigating the role of STAG2 in bladder cancer have reported conflicting results regarding aneuploidy[170,274,278]. The knockdown of *STAG2* by small interfering RNA (siRNA) resulted in an increase in aneuploidy in normal human bladder cells[278] but investigations into bladder tumours report few CNAs[170] and a lower FGA than WT tumours[274]. As our study contained only 2 tumours from one individual patient with loss of *STAG2* there is insufficient data to contribute to either argument as it is possible that that these tumours would be outliers in a larger study of *STAG2* CN loss in NMIBC.

Loss of *STAG2* expression is associated with low-stage and low-grade tumours[170,274,279,280]. In NIMBC, loss of *STAG2* is associated with a good prognosis and reduced risk of recurrence and progression[170,273,279,280]. Conversely in MIBC, loss of *STAG2* was associated with an increased risk of recurrence and cancer specific mortality[273]. The biological basis for this difference is currently unknown. In our cohort, loss of *STAG2* was only identified in two of the recurrent tumours from patient P0712 and these were both stage Ta of G2 or G3. Despite having many recurrences, this patient did not display any stage progression. With such low numbers in our cohort, no conclusions can be drawn regarding *STAG2* loss. Analysis of more recurrent tumours would help confirm or refute the observations regarding recurrences.

Patient P0712 is male and therefore only has one copy of *STAG2*, yet in tumour P0712-S02 the gene has not been completely deleted. This could suggest that it represents a sub-clonal loss in this tumour as it was highly pure. Taylor *et al.* [274] found that expression of *STAG2* was frequently chimeric suggestive the involvement of intra-tumour genomic evolution.

As part of their treatment a subset of patients underwent a 6-week course of intravesical MMC therapy. MMC has been seen to cause a variety of aberrations *in*

vitro including the induction of sister chromatid exchanges[281], giant cell formation with nuclear fragmentation[194], the formation of micronuclei[200] and non-random interchange breakages[199]. More recently, an *in vivo* study using the model organism *C.elegans* identified that MMC treatment mainly causes deletions[205]. We therefore hypothesised that there might be an increase in the number of CNAs detectable after treatment and these were more likely to be deletions rather than gains. It was thus interesting that no significant difference in CNAs from pre- and post-MMC treated tumours was identified.

Analysis of the CNAs surrounding MMC treatment identified a single region of CNA potentially associated with MMC treatment in multiple patients. This was loss of the 9p21 locus containing *CDKN2A* seen in post-MMC tumours from patients P0468, P0926 and P2440. For patient P0926 this was a change from CN neutral to loss of the region and for the other two patients it was a change from loss to HD of the region. Despite the fact that this occurred post-MMC treatment in all 3 cases, it is impossible to say that this was a direct result of MMC treatment. Patients P0468 and P0926 had a gap of 12 and 3 years respectively between tumours pre- and post-chemotherapy. Within this period they both also received rounds of BCG treatment. It is therefore not possible to say if this aberration is due to MMC, BCG or just down to tumour evolution over time.

Many of the early *in vitro* studies into the effects of MMC treated peripheral blood leukocytes with different concentrations of the drug before making cytological preparations to assess breaks and exchanges in metaphase spreads. One of the major findings of these early studies was that chromosomes 1, 9 and 16 were the most affected, specifically in areas of secondary constriction[196-199]. These regions of secondary constriction are the peri-centromeric regions of heterochromatin found specifically on these three chromosomes. These regions are gene-poor and have yet to be properly annotated with a sequence and therefore are not covered by NGS. This means that aberrations such as deletions, SNVs etc. cannot yet be identified in these regions. Another possible cause for the lack of an increase in CNAs after treatment is that any chromosomal interchanges occurring could be CN neutral and would therefore not be detectable by CN analysis. It is also possible that the large exchanges could be more lethal to tumour cells *in vivo* or that these do not provide a growth advantage and therefore do not grow out to form a significant clonal population detectable by shallow-pass WGS. The deletions identified by Tam *et al.* in *C.elegans* were also small, ranging from 2-318,826 bp in length. This could mean that any CNAs being generated by treatment may be smaller than can be identified with the resolution of this data.

Despite the lack of findings at the CNA level, it is still possible that MMC can affect the DNA. Studies have used sequencing to interrogate mutations caused by MMC *in vitro* and have identified a large proportion of substitutions and single base deletions[201,202]. WES performed on tumours pre- and post-MMC treatment should be able to identify any differences in the number and type of SNVs in the two groups. It may also provide an increased resolution for the analysis of CNAs in the coding regions, allowing smaller events to be discovered.

Bladder cancer has well characterised hotspot mutations in *FGFR3, PIK3CA,* the *TERT* promoter and the RAS gene family. These regions were probed using SNaPshot assays to detect the presence of any SNVs. Over 71% of samples were *FGFR3* mutant. *FGFR3* mutation has been associated with favourable disease parameters and a better prognosis[259-282]. Previous studies have seen a lower overall FGA in tumours with an *FGFR3* mutation compared to *FGFR3* wild-type tumours[125,259]. In this study there was no significant difference in FGA in *FGFR3* mutant verses WT tumours with mutant tumours having a slightly higher median FGA than WT tumours. This remained true even when separating the tumours by stage and grade, and when a single mutant and/or wild-type tumour from each patient only was analysed. The likely cause for the discrepancy in this data is the high frequency of chromosome 9 loss seen in the *FGFR3* mutant tumours. This study also specifically selected for patients with recurrences and for patients undergoing a treatment course of MMC. These patients tend to have a higher risk of recurrence and progression so it may be that this has selected for patients with a higher level of genomic instability than previous studies. It would be interesting to follow patients who had both *FGFR3* wild-type and mutant tumours to see which tumours recurred and if there was any progression within these patients. Analysis of the subsequent tumours would identify if wild-type tumours or tumours harbouring an *FGFR3* mutation led to the recurrent/progressed tumours. Interestingly, 29% of *FGFR3* mutant tumours demonstrated gain of 4p16.3, the region harbouring the *FGFR3* gene. It is possible that these tumours are heavily reliant on signalling through *FGFR3,* making this a possible therapeutic target for these patients.

Mutations in RAS genes were identified in only two patients; P0198 and P0533. Mutations in RAS genes and *FGFR3* are mutually exclusive events[142,143,283]. Consistent with this, none of the tumours with a RAS mutation carried an *FGFR3* mutation in this cohort. This mutual exclusivity is thought to be due to the two genes signaling through the same pathway in bladder cancer[146]. GWFP and WGPs of RAS mutant tumours showed that these tumours have relatively stable genomes with conserved CNAs

between tumours from the same patient. Neither of the patients had loss of the whole of chromosome 9, yet both had HD of *CDKN2A.* Interestingly, tumours harbouring RAS mutations frequently lack deletions of 9q (Carolyn Hurst, personal communication). Loss of 9q has been associated with upregulation mTORC1 signalling, likely through the loss of *TSC1,* a negative regulator of mTORC1[82]. Both RAS and mTORC signaling result in an increase in cell survival, proliferation and motility and the two pathways have been demonstrated to cross-talk and regulate one-another[284]. This could suggest that activation of both pathways is redundant in cancer, and may provide a possible explanation for the mutual exclusivity of these events.

The *TERT* promoter was the most frequently mutated hotspot region with 21/23 patients containing a mutation in at least one of their tumours. Other studies have demonstrated *TERT* promoter mutations in 56-83% of bladder tumours[156,157,285]. Restricting analysis to the first tumour from each patient identified that 87% (20/23) of patients carried *TERT* promoter mutations, a similar level to that reported previously. Only 6 tumours did not contain *TERT* promoter mutations and all of these tumours showed very stable genomes with loss of chromosome Y being the only recurrent event. Work by others has previously identified that bladder tumours with *TERT* promoter mutations have a significantly higher CNA burden and mutational load burden than wild-type tumours[286]. This correlates with our observations but the low sample size of wild-type tumours in our cohort means that results should be reviewed with caution. The lack of CNAs in tumours that do not carry *TERT* mutations suggests that mutation of the *TERT* promoter is an early event that precedes the acquisition of CNAs. This idea is reinforced by the observation that *TERT* promoter mutations are seen on the trunk of the phylogenetic tree in 20 out of 21 patients.

Some tumours did not share the same mutations with other tumours from the same patient. Patient P1175 had one tumour that was heterozygous and one tumour that was homozygous for an *FGFR3* G372C mutation. Analysis of the CN status at the *FGFR3* gene locus indicated no alterations in the region suggesting that tumour P1175-S01 may have undergone CN neutral LOH compared to tumour P1175-S02. It is also possible that tumour S02 contained more contaminating normal cells resulting in the detection of the wild-type allele in the mixture. However, the sample was macrodissected and consequently should have been close to 100% pure making this unlikely. It is possible that the LOH event is subclonal in this tumour, which would explain the observation of the wild-type allele in the SNaPshot analysis.

In patient P0533 two different *PIK3CA* mutations were identified in two tumours. These were the E542K and E545K hotspot mutations found in the helical domain of the protein[233]. These appeared to be later events in the carcinogenesis process in this patient occurring on the branches of the tree. These mutations have previously been linked to APOBEC mediated mutagenesis[83], especially when present subclonally[287] and they tend to occur later in carcinogenesis[288]. This might suggest that the *PIK3CA* mutations found in this patient are a result of APOBEC mutagenesis.

The prevalence of hotspot mutations in bladder cancer has resulted in studies attempting to use this information to create non-invasive urine-based tests for diagnosis and identification of recurrences. Bladder tumours routinely shed cancer cells and tumour DNA into the urine and this can be used as a target for tests that detect hotspot mutations[143,156]. Mutations in the *TERT* promoter have been assessed in the urine of bladder cancer patients for the detection of disease recurrence[156,289,290]. A recent prospective study used the detection of *TERT* promoter mutations in the urine as a marker of bladder cancer[290]. This study found that *TERT* promoter mutation detection was more sensitive than cytology with a high specificity when detecting bladder cancer. They also reported that *TERT* promoter mutations were a dynamic marker of recurrence as detection of these mutations mirrored the cytoscopic presentation of recurrent tumours in several patients. However, there are limitations to the efficacy of such urine tests and one such limitation is highlighted by the identification of recurrent tumours in our cohort that did not contain the hotspot mutations detected in the primary tumours. Relying on mutation detection in the urine alone would not have identified these tumours. To combat these issues, groups are investigating the use of panels of markers to improve sensitivity and specificity[291,292]. A recent systematic review of urine biomarkers reported that these multi-target panels have better diagnostic performance compared to single target biomarkers[293].

Whole genome plots were generated to enable comparison of CNAs between tumours from the same patient. CNAs were also used to reconstruct phylogenetic trees for all tumours from a given patient using the TuMult algorithm[35]. This uses the breakpoints of related tumours to infer relationships between the tumours. Events seen in multiple tumours are likely to have occurred earlier whilst events unique to certain tumours are likely to have occurred later in the evolutionary process. This data was supplemented with data from the single gene mutation analysis to aid in tree reconstruction.

The majority of patients shared at least some CNAs suggestive of a monoclonal origin. For patients P0198, P0536, P1326 and P1777, all tumours shared all CNAs with no

divergence seen, making tree building difficult. This shows a lack of inter-tumour heterogeneity with respect to CNAs for these tumours. It would be interesting to investigate whether these tumours had any unique mutations in order to assess whether the lack of heterogeneity is restricted to CNAs or if this is reflective of SNVs as well. For example, hotspot mutation analysis showed that tumour P0536-S03 contained an additional *TERT* mutation compared to the other tumours from this patient.

Cancer can follow different types of evolution. In the linear evolution model, a selective sweep follows the acquisition of a driver mutation, resulting in a dominant clone which can be seen at the time of profiling. Some patients (patients P0926, P0933, P0990, P1175, P1870, P2104 and P2440) showed apparent linear evolution where one tumour appears to be the direct descendant of another tumour, having acquired additional aberration(s). Interestingly, this linear evolution pattern was not always chronologically linear in terms of tumour resection. For example, for patients P0933, P1175 and P2104, the primary tumour was predicted to be a direct descendant of the recurrent tumour.

This lack of genomic evolution reflecting the chronological presentation of the tumours was not restricted to tumours displaying apparent linear evolution. Indeed almost 50% of cases in this study showed fewer CNAs in the recurring tumour compared to the primary tumour. This has been observed before in bladder cancer; a study profiling recurrent tumours with CGH showed that despite the tumours sharing some aberrations, others were different between the tumours and these were not compatible with a simple progression model where the recurrent tumour is directly descended from the primary tumour[167]. Other studies have also reported that the chronology of tumour presentation does not necessarily reflect the genomic evolution of bladder cancer[265]. A study looking at recurrences spanning periods of up to 17 years identified a clonal relationship in all the patients studied and also observed that the chronological order of tumour presentation did not parallel the genetic evolution of the tumour[163].

This reduction in genomic complexity as tumours are chronologically resected is likely a consequence of field change seen in bladder cancer. Once a tumour is detected it is resected using TURBT. The earliest tumours to be detected are the fastest growing ones and it stands to reason that these are likely to be the ones with the most genomic alterations. Once these have been resected the slower growing clones may be left resulting in a tumour with fewer genomic alterations than the previous tumour. This may also explain why recurrent tumours can be of a lower grade than the preceding

tumours[4]. The slower growing clones may be left due to incomplete resection, re-implantation or due to the spread of early clones throughout the urothelial layer, or potentially all three. Improvements in resection, such as the implementation of hexylaminolevulinate (HAL) photodynamic diagnosis (PDD) which is a procedure that uses a photoreactive molecule to identify neoplastic cells with blue light, will likely be instrumental in reducing recurrences. If all cancerous cells can be detected and removed then recurrence rates should drop. These practices are currently being trialed with promising results[294].

In two cases, no evidence for clonality was detected. These were patients P2065 and P2291. In each case the tumours from these patients did not share any CNAs or hotspot mutations at all. Tumour P2065-S02 was impure and required enrichment for tumour cells using LCM. In this case the stroma was removed from the slide leaving the tumour cells which were then scraped into a microcentrifuge tube for DNA extraction. This would have resulted in a highly pure DNA sample so it is unlikely that the hotspot mutations could have been missed due to contamination with normal cells. Tumour P2291-S02 had an estimated tumour content of 75% at the start of sectioning, but as cutting progressed, this purity dropped to approximately 60%. It was accepted that overall the purity of the cut sections would be around 70% so DNA was extracted from the two tubes that had been cut. Therefore, it is unlikely that the lack of clonality seen in these tumours are an artefact of tumour impurity.

Previous studies have identified synchronous and metachronous tumours that do not appear to be clonally related. Lindren *et al.*[167] used CGH, LOH and mutation profiling to analyse recurrent tumours and this indicated a clonal relationship in all but two patients. In one of the patients the tumours did not share any CNAs or regions of LOH but did share the same *TP53* mutation suggestive of a clonal relationship. The other patient's tumours did not share any events except for LOH in the region of *CDKN2A.* However these were likely independent events as different alleles were lost, suggesting that the samples were not clonally related.

Despite providing a good overview of the relatedness of the tumours from the same patient there are some flaws to the TuMult algorithm. For patient P0712 for example, TuMult predicted that gain of chromosome 21 occurred in a common precursor to tumours P0712-S01, P0712-S04 and P0712-S05 followed by loss of some regions of chromosome 21 in tumour P0712-S01. TuMult also separated the loss of 8p, seen in all tumours except P0712-S01, into two separate branches. It is much more likely that the loss of 8p was a common event whilst the loss of the different regions of

chromosome 21 were more likely different events. This suggests that the prediction of common breakpoints within the algorithm could be refined to make more biological sense.

Another limitation of the TuMult algorithm is that is uses the discretized data to infer the breakpoints. In most cases this is not a problem, as for most break points there is a change in the discretized CN status. However, in some cases there are CN changes present in multiple tumours from the same patient but in one of the tumours the change is not as great and does not change the discretized status. For example, tumours from patient P0418 shared 3 focal gains on chromosome 7. Additionally, two of the tumours (P0418-S01 and P0418-S02) demonstrated gain of the whole of chromosome 7. The whole chromosome gain likely resulted in two of the focal gains being classified as amplifications at 7p21.1 and 7q34 thus altering their discretized CN value. However, the third focal gain at 7q11.23 remained classified as a gain rather than an amplification resulting in no change in the discretized CN. As the algorithm cannot distinguish that this region is different from the whole chromosome gain it does not get automatically included as shared with tumour P0418-S03 and this alters the tree. This was corrected manually, but it represents a flaw in the algorithm.

Hierarchal clustering of the 67 tumours revealed that the majority of tumours from the same patient cluster together. This correlates with previous studies clustering multiple tumours from the same patient by CNAs[163,167], demonstrating that tumours from the same patient are relatively stable at the CN level. A lack of divergence at the CN level has also been demonstrated in muscle invasive and metastatic bladder cancer where tumours from the same patient tended to cluster into the same group despite containing some private CNAs[62]. Combined, this could suggest that the CN landscape of bladder tumours is set relatively early in cancer development, with only a small amount of evolution occurring between tumours.

Hierarchal clustering of the 67 tumours with tumours from Hurst *et al.*[82] showed that the majority of tumours from this cohort clustered into the more genomically unstable subgroup, GS2 whilst tumours with few-to-zero CNAs clustered into the GS1 subgroup. In the study by Hurst *et al.* more tumours clustered into the GS1 subgroup than the GS2 subgroup. They included only primary stage Ta tumours, the majority of which were of grade 1 or 2. Our cohort of 67 tumours contained some tumours of a higher stage and grade; 7 TaG3, 4 T1G2, 2 T1G3, 1 Ta(x)G3 and 2 stage T1(x)G3. This could explain why these tumours clustered into GS2 as tumours of a higher stage and grade often contain more CNAs[125,135]. Only 11 TaG2 tumours (from 6 patients; P0536, P0933,

P0990, P2065, P2161 and P2291) clustered into GS1 whilst 31 TaG2 and 8 TaG1 tumours clustered into GS2. The high number of tumours clustered into GS2 could suggest that recurrent tumours are more genomically unstable or contain a high frequency of chromosome 9 or 9q loss. Interestingly, Hurst e*t al.*[82] identified a trend towards decreased recurrence free survival for patients in GS2, but this was not formally significant (p = 0.25). It would be interesting to profile more recurrent tumours to see how they cluster.

Sample selection for CNA may have introduced some biases into the data. Patients were initially selected as they had multiple recurrences but obtaining sufficient DNA from many of the recurrent tumours proved challenging due to the small size of the tumours. Additionally, tumours that were heavily infiltrated or were difficult to microdissect were excluded from the analysis as their CN profiles would have been heavily influenced by the contaminating normal cells. This means that the tumours in the CN cohort were biased towards larger tumours that contained few contaminating normal cells. It is possible that smaller tumours may have displayed different characteristics; it would be interesting to compare larger and smaller tumours to see if there are any alterations associated with tumour size.

Infiltration of bladder tumours by lymphocytes has been identified as prognostic in muscle-invasive and metastatic bladder cancer, with heavy infiltration being associated with a longer overall survival[295-297]. However, in non-muscle-invasive tumours the relationship is less clear with dense tumour infiltrating lymphocytes having been shown to predict progression in an early study[295] but in other studies, a lack of correlation with clinical outcomes has been observed[296,298]. These differences may have been influenced by the scoring methods used in each study. It would be interesting to look at the mutation load in the heavily infiltrated tumours to see if it is different to the tumours included in our cohort. However the presence of the infiltrating lymphocytes means that the sequencing depth would need to be much higher to be able to identify variants with a low VAF.

In the subset of patients that underwent a course of MMC treatment, all 10 patients that had a tumour available both pre- and post-MMC treatment underwent CN analysis. Patients for whom a tumour was not available post-MMC treatment were not included and this subset included those patients whose disease did not recur after treatment. This patient subgroup would be interesting to investigate as it may be possible to identify features associated with a full response to a course of MMC chemotherapy.

CNAs were detected using shallow-pass WGS and a read depth approach. A pseudo-CGH algorithm, ngCGH (https://github.com/seandavi/ngCGH), was used. This algorithim uses a window size of 1000 reads to compare read depth between the tumour and normal, from which the log2 ratio is calculated. Each window can therefore be thought of as similar to a probe used in array CGH (aCGH) and the method produces a data set similar to that produced using aCGH methodologies. Previous work performed by the lab has demonstrated that the profiles generated by aCGH and the NGS pipeline display the same major features, however the NGS method has an improved resolution compared to aCGH due to the increased probe number and a lack of space between probes (see Appendix E).

A benefit of using shallow-pass WGS is that it can provide a higher resolution than aCGH: in aCGH the resolution is limited to the distance between probes and probe size (arrays used in our lab had a 1-Mb resolution and contained ~4000 probes[125]) whilst in WGS the resolution correlates with the depth of coverage. For the shallow-pass WGS, the raw depth of coverage was ~0.7x, and this equated to an average of ~12700 probes per patient. As coverage is not uniform in NGS data the algorithm uses a set number of reads for windowing rather than a set distance as this produces CN estimates that should have similar variance at each location. The limitation of this method is that window sizes can vary meaning that the resolution is not set and this makes it difficult to assign a value to the resolution granted. This makes it difficult to predict the size of CNAs that may be missed by the method, as these will be bigger in regions of low coverage. An alternative method would be to use a set window size (e.g 10 kb) but this would have its own challenges in low coverage regions.

Comparisons of profiles for the same tumour generated by shallow-pass WGS or aCGH suggests that shallow-pass WGS may be more sensitive than aCGH in detecting shallow CNAs (see Appendix E). However, it is also possible that intratumor heterogeneity and a lack of a matched normal may have affected analysis in the aCGH data set. As the shallow gain detected by WGS was detected in an additional tumour from the patient it would suggest that this event is likely to be real. A limitation of both aCGH and shallow-pass WGS is that they provide estimates of the relative CN, not absolute CN. Additional information from B-allele frequencies is required to estimate absolute CN. This could be improved with increased depth of sequencing for the WGS method or could be provided by using SNP arrays.

Another limitation of both methods is that other structural rearrangements, such as inversions or chromosome fusions, cannot be identified. Alternative methods such as

multicoloured Fluorescence In Situ Hybridization (FISH) could be used to identify rearrangements, but this would not be able to detect inversions. These could be detected by an alternative method such as directional genomic hybridization[299]. Information on structural rearrangements could also be generated by using deeper WGS and using information provided by mate-pairs[257]. However, this increased level of coverage would significantly increase the cost of the sequencing meaning that fewer samples could be assessed.

Overall, the shallow-pass WGS method employed for the detection of CNAs is an improvement on previous methods such as aCGH but it still has some limitations. If the price of WGS continues to fall, these could be countered by an increase in sequencing depth.

3.4 Summary

Overall, tumours from the same patient tend to share most CNAs and cluster together suggesting that the generation of CNAs is an early event in bladder cancer tumorigenesis. Some tumours will then go on to develop private CNAs. No difference in the frequency of CNAs or the fraction of genome altered was identified between tumours resected before and after MMC treatment. The majority of tumours from the same patient were predicted to be clonally related as they share many CNAs despite being separated by time and space. This supports the idea of the field change in bladder cancer where much of the urothelium contains aberrations[4]. The presence of the tumours that are potentially oligoclonal also suggests that the urothelium can have different genetic damage in spatially separated regions.

Chapter 4
Analysis of matched pre-MMC and post-MMC treatment tumours by whole exome sequencing

4.1 Introduction

The primary focus of this project was to investigate genomic alterations in bladder tumours associated with a course of mitomycin C (MMC) treatment. MMC is a potent DNA alkylator with the potential to form DNA-MMC monoadducts as well as intra- or inter-strand crosslinks. Next-generation sequencing studies investigating other DNA-alkylating chemotherapies have shown that these therapies generate mutations that can be identified in post-treatment tumours[176,300]. In glioma, treatment with the alkylating agent temozolomide has been shown to induce C:G > T:A transitions predominantly at CpC and CpT dinucleotides[70] whilst in MIBC, treatment with cisplatin-based chemotherapy was associated with an increase in C > T and C > A mutations[176].

Early *in vitro* molecular studies identified MMC-induced mutations, predominantly in GC rich regions, with C > A transversion substitutions occurring most frequently[201,202,204]. Additionally, an *in vivo* study identified an increased number of tandem substitutions in MMC treated mice[203], and this has also been identified *in vitro*[201,203]. To investigate the possible mutagenetic effect of MMC treatment, paired pre- and post-MMC treatment tumours from 8 patients were subjected to whole-exome sequencing to identify somatic variants.

Accurately calling somatic variants is inherently difficult[301]. Cancers are often impure and can contain rare subclones resulting in low frequency variants. These are therefore challenging to disambiguate from sequencing artifacts that are often present in low frequencies. There are now numerous variant callers available with a recent review identifying 46 publicly available callers[302]. Choosing the correct variant caller(s) depends on the type of variant that is of interest (SNV, indel) and expected allele frequency range.

Outputs of different variant calling pipelines can be highly divergent. This divergence is reflected by the results of a study that compared 4 major variant-calling methods. They identified that only 31% of the SNVs were identified by all 4 methods and there were numerous SNVs that were either unique to one caller, or missed by only one caller[303].

To attempt to identify the most accurate variant-calling methodologies the International Cancer Genome Consortium (ICGC)-TCGA Dialogue for Reverse Engineering Assessments and Methods (DREAM) Somatic Mutation Calling Challenge was launched (https://www.synapse.org/#!Synapse:syn312572/wiki/). This effectively crowdsourced the running and fine-tuning of algorithms. Overall, a clear trade-off was seen between precision (fraction of predicted SNV's that are true) and recall (number of SNVs detected)[304]. Using an ensemble of the variant-calling methodologies to create a consensus outperformed any individual method, especially when assessing more complex genomes[304]. As this data suggests that variant calling should perhaps be made using multiple variant callers, we decided to use an ensemble of three variant callers to identify variants. In this chapter we tested five different variant callers, in every possible combination of three callers, to identify the best combination.

The best practices in variant calling have so far focused on reducing the number of false positives. Analysing multiple tumours from the same patient has additional challenges as it is important to be able to identify mutations shared between tumours. Mutations that are present at a low level in one tumour yet are expanded in a second can give insight into genetic events contributing to progression, recurrence and resistance. Misidentifying a variant as unique to one tumour when it was actually present in both would alter the perceived evolution of those tumours allowing incorrect conclusions. It is therefore important in these samples to ensure that these shared variants are not filtered out. Droop *et al.*[242] describe a method to improve variant calling across multiple samples from the same patient. They reason that having multiple tumours from the same patient creates a type of internal validation for shared mutations: if a variant is seen as shared between samples then these are likely to be true variants regardless of the strength of evidence of any one call. The identification of shared mutations can therefore be completed at a lower stringency to ensure that all shared mutations can be identified. A second round of more stringent variant calling follows to identify variants found in only one sample. Using this method they were able to identify many more biologically relevant shared mutations. As our tumour samples are paired we investigated using this methodology in our variant identification pipeline.

By combining both shared mutation identification and ensemble calling, a two-stage variant calling pipeline was devised. This aimed to improve identification of shared variants whilst maintaining high specificity for unique variants. This pipeline was used to call somatic variants from whole-exome sequencing of paired tumours resected before and after a six week course of intravesical MMC treatment from 8 patients with

matched peripheral blood serving as a germline control. Individual tumours were assessed for mutation burden, base substitutions and mutational signatures and comparisons were made between pre-MMC and post-MMC treated tumours.

In this Chapter, tumour nomenclature follows that of Chapter 3. However, each tumour also has an additional identifier after the tumour number: BX - before MMC, PX – post-MMC, UX – unique variants post-MMC.

Two patients had more than one tumour at a timepoint: P2218 had two tumours pre-MMC and P0960 had two post-MMC. To avoid skewing the data through the inclusion of multiple tumours with similar genomes, only one tumour from each patient was included at a timepoint for comparisons. Tumour P0960-S03-PX was used as the post-MMC tumour for patient P0960 and tumour P2218-S1A-BX was used as the pre-MMC tumour for patient P2218 for Figure 4.6 A-E, Figure 4.7, Figure 4.8, Figure 4.9D and Table 4.3

Throughout this thesis, the terms ancestral/founding clone, subclone, subclonal and clonal are used. The ancestral/founding clone refers to the cluster of mutations that are present in every cancer cell of every tumour from a patient. They are found on the trunk of the cancer evolutionary tree. Subclone refers to a cluster of mutations that have branched off the ancestral clone whilst subclonal refers to mutations that are only present in some, not all, of the cancer cells of a tumour. Subclone and cluster therefore mean the same thing and are used interchangeably. Clonal means that the mutation is found in all the cancer cells of a particular tumour. A subclone can therefore be clonal, present in all cells of a tumour after making a clonal sweep, or subclonal with only some cancer cells containing the mutation(s).

4.2 Results

4.2.1 Variant calling pipeline

As discussed above, there are many caveats to variant calling. To create a variant calling pipeline, pre-MMC and post-MMC treated tumours from two patients were analysed initially; P0418 and P2161. Whole exome sequencing (WES) data for these two patients was initially used to determine the best way to identify the shared variants and determine which variant callers performed best for consensus calling.

4.2.1.1 Identification of shared variants

One of the most important processes in analysing paired tumours is the identification of shared variants. MuTect2 was selected to initially investigate the identification of shared variants due to its popularity within the literature and its ability to call low variant-allele frequency (VAF) variants. The raw data (i.e. all variants identified prior to filtration) can also be extracted from MuTect2 and this data can aid the identification of shared variants.

Shared and unique variants from the VCF files created by MuTect2 were identified using the "dplyr" package within the R software environment (Figure 4.1A). To ensure that unique variants were correctly annotated as such, the raw MuTect2 data was interrogated. This identified 6 variants in patient P2161 and 14 variants in patient P0418 that were actually shared between tumours but had been filtered out for one of the tumours. As these variants are present in both tumours these are likely to be real despite not being called in one of the tumours. The most common causes for filtering out of variants were "t_lod_fstar", which is where the tumour event does not meet the likelihood threshold for a variant to be called usually due to an insufficient number of reads, or "clustered_events". Visual inspection using the Interactive Genome Viewer (IGV) showed that these "clustered_events" were due to artefacts in the sequencing data around the variant site but the variants themselves were clearly real.

Droop *et al.*[242] used VarScan2 to identify their variants. The default parameters on VarScan2 are quite stringent and it requires a minimum allele frequency of 20% when run in somatic mode. This means that low allele frequency variants are missed. To combat this, Droop *et al.* used modified parameters changing the minimum coverage to 10 and setting the minimum VAF to 3.5%. To see if using VarScan2 with the adjusted parameters improved the identification of shared mutations, variants were called using both the default parameters and the adjusted parameters (Figure 4.1B). Adjusting the parameters led to an increase in the total number of variants identified for each tumour as expected. Counterintuitively, the number of variants identified as shared between tumours from patient P2161 was actually less than the number identified using the default parameters although a small increase in the number of shared variants was observed for tumours from patient P0418.

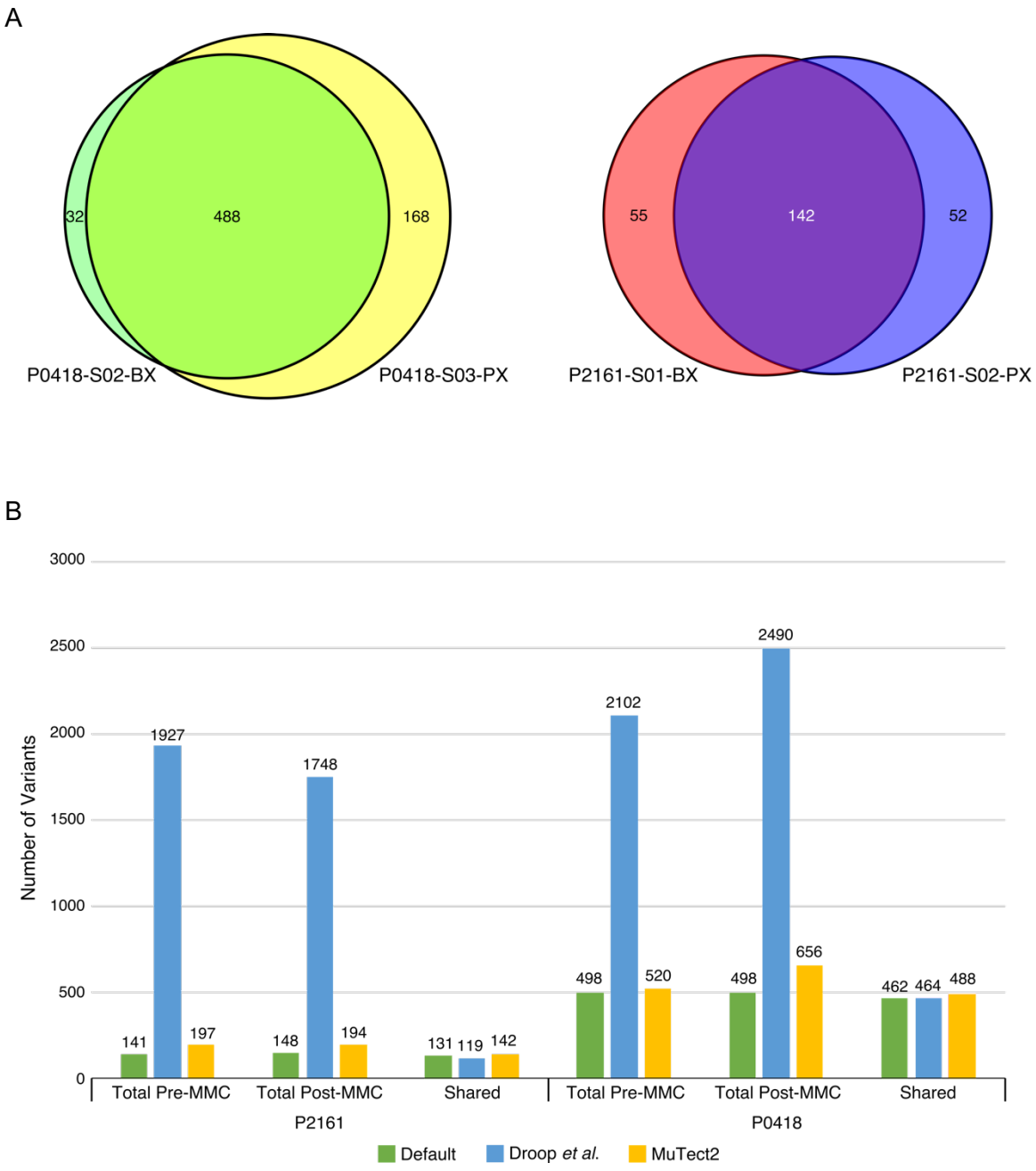


Figure 4.1:Variant calling optimisation.
A) Variants were called using MuTect2. Venn-diagrams show the shared and unique variants for pre-MMC and post-MMC tumours from patients P0418 and P2161. The unique variants for each tumour were checked against the raw MuTect2 data to identify any shared variants that may have been filtered out. These were included in the Venn diagrams. B) Variants were called using VarScan2 with either the default or less stringent parameters as described in Droop *et al.*[242] and compared to variants called by MuTect2. Bars show the number of variants identified by each method.

Overall MuTect2 outperformed VarScan2 with more variants identified for each tumour (compared to default) and more shared variants identified than either VarScan2 approach. Using the shared variant method, a further 20 variants were identified that may be important in the predicting the inferred evolution of the cancer. This method of identifying shared variants was therefore included in the pipeline.

4.2.1.2 Multiple variant callers

It has been shown in many studies that different variant callers will generate different outputs from the same sequencing data. This means that relying on just one variant caller may lead to variants being missed and false positives being called. Using the consensus from multiple callers is a way to improve specificity and recall. For this study it was decided that variants should be taken forward when reported by at least two out of three variant callers. Five variant callers were tested and the best combination of three of these was selected.

Variants were called using five variant calling algorithms; MuTect2[301], VarScan2[240], Strelka2[238], EBCall[239] and MuSe[241]. Strelka2 called the most variants whilst VarScan2 and MuSe called the least (Figure 4.2A). Of the five variant callers only MuSe does not call indels. It was noticed that indel calling appeared to be skewed for two of the variant callers; Strelka2 appears to call more deletions and MuTect2 more insertions. VarScan2 and EBCall appear to call both insertions and deletions fairly evenly (Figure 4.2B).

Shared mutations are internally validated so they do not need to be called by more than one caller. Shared mutations were therefore used to evaluate the 5 callers. As with MuTect2, VarScan2 and MuSe both generate pre-filtering variant data, facilitating the checking of unique variants against the raw data as described above. This identified a total of 4 extra shared variants for VarScan2 (2 per patient) and 3 extra shared variants for MuSe all in patient P2161.

We were interested to see if any of the “unique” calls made by a variant caller were called as a “shared variant” by an alternative variant caller (Table 4.1). Over 50% of the unique variants identified by VarScan2 were seen as shared by an alternative caller making it the lowest performer of the 5 callers. Strelka2 and MuTect2 performed best with fewer than 1% of calls being incorrect.

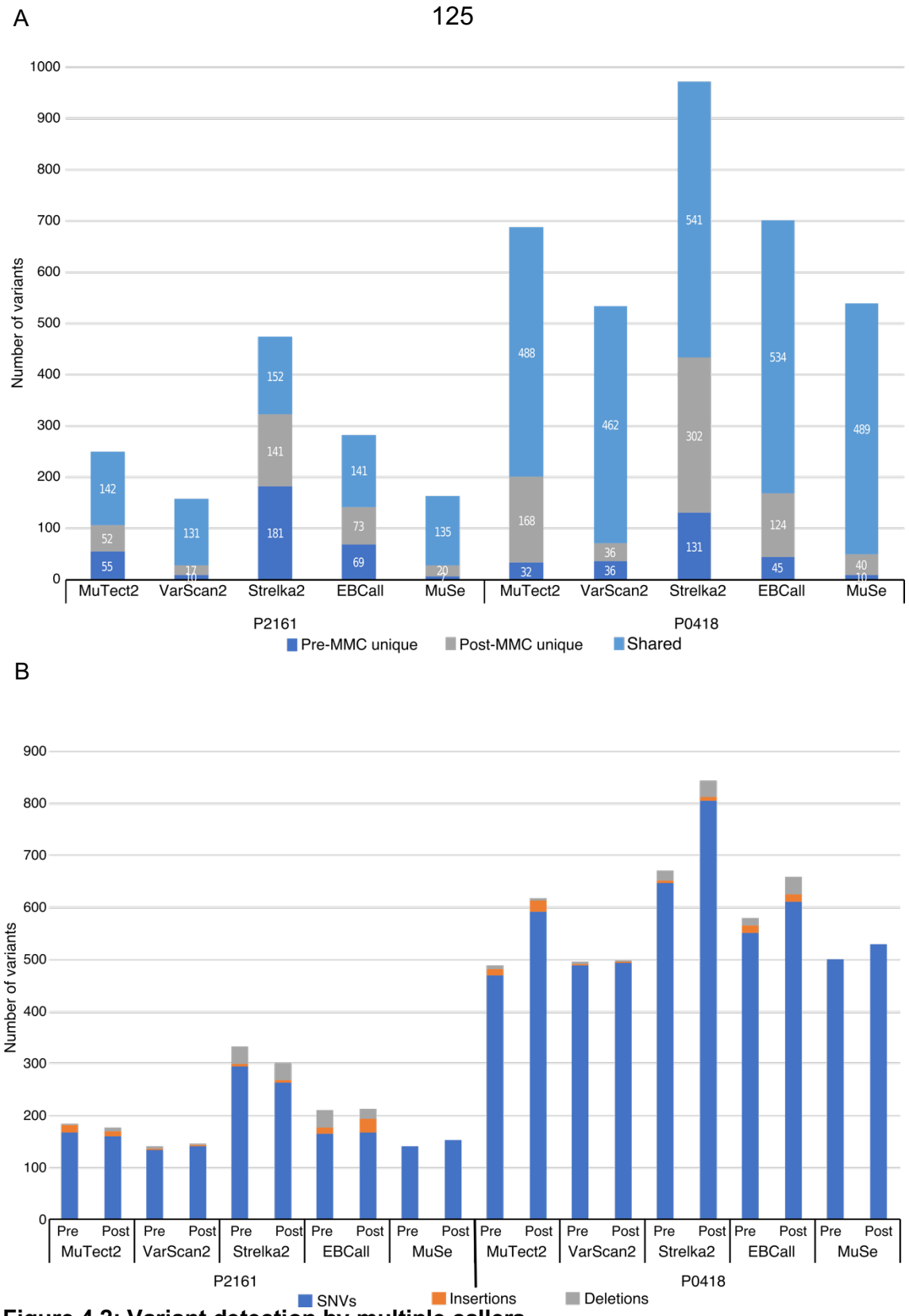


Figure 4.2: Variant detection by multiple callers.
Five variant callers (MuTect2, VarScan2, Strelka2, EBCall and MuSe) were used to detect variants from pre-MMC and post-MMC treated tumours from patients P0418 and P2161. A) Numbers of variants detected by each caller. Bars show both the unique mutations and shared mutations for each tumour pair. The x-axis identifies the results for each variant caller used for both patients, the y-axis shows the number of variants. B) Types of variants detected by each caller. The numbers of SNVs (blue bars), insertions (orange bars) and deletions (grey bars) are shown. Pre-MMC and post-MMC tumours are separated for each variant caller. The y-axis denotes the number of variants.

Table 4.1: Evaluation of unique variants using 5 different variant callers.
The unique variants identified by each variant caller for each tumour were interrogated to see if they had been identified as shared by a different caller. [a]Total number of variants identified as unique for both tumours from both patients P0418 and P2161. [b]Denotes the total number of those unique variants that were actually identified as shared variants by an alternative variant caller.

Variant Caller	Total no. of unique variants[a]	No. of variants incorrectly identified as unique[b]	% of incorrect variant calls
MuTect2	307	3	0.98
VarScan2	99	50	50.51
Strelka	755	7	0.93
EBCall	311	5	1.61
MuSe	77	4	5.19

Next, we wanted to identify the best three variant callers to be used in the pipeline. All 10 possible combinations of three variant callers were used to generate consensus calls using the vcftools function “vcf-isec” with “-n +2” specified so that only variants present in two or more input VCFs were kept. The combination of MuTect2, Strelka2 and EBCall identified the most variants and combinations involving VarScan2 and MuSe identified the least (Figure 4.3).

As the shared variants are internally validated it would be expected that these are all non-artefactual regardless of which variant caller identified them. Therefore, all the shared variants identified by each caller were added together to get the total possible number of shared variants for each patient. For patient P2161, a total of 161 variants were identified as shared and 560 were identified for patient P0418. Each combination of variant callers was then assessed for the number of shared mutations they each identified. The combination of MuTect2, Strelka2 and EBCall gave the most shared variants with 160/161 for patient P2161 and 557/560 for patient P0418. Further investigation into the variants not shared by this group of callers revealed that none of these were exonic.

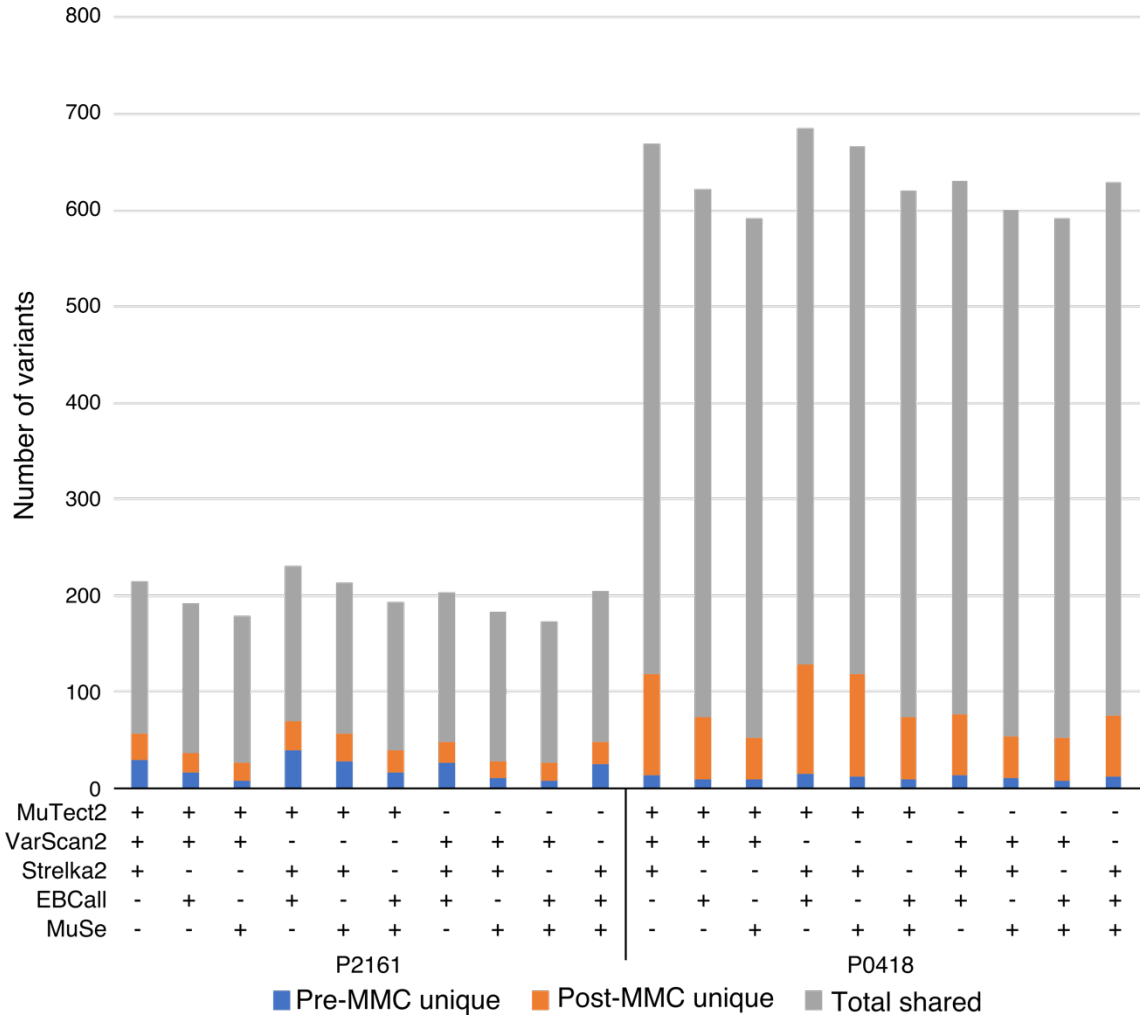


Figure 4.3: Consensus calling.
Variants were called with five variant callers. Each possible combination of 3 variant callers was used to generate a consensus with variants required to have been called by at least 2 of the three callers. Shared variants were included even if they were only called by one variant caller as these have internal validation. The use of a caller within a combination is denoted by a "+" in the table. All 10 possible combinations were tested for each patient.

Overall the best combination of callers was considered to be MuTect2, Strelka2 and EBCall. Each of these performed well individually with low false-unique call rates and when combined they gave the best identification of variants. The variant calling pipeline described here therefore consists of consensus calling using these three variant callers followed by the addition of any variants that were identified as shared by any one variant caller. Unique variants were checked against the raw MuTect2 data and any shared variants that may have been filtered out in one tumour were added back into the dataset.

4.2.2 Exome sequencing analysis

Whole exome sequencing was performed on a total of 18 tumours from 8 patients who underwent a course of MMC treatment. Matched pre- and post-chemotherapy biopsy tissue was sequenced for each patient with peripheral blood used as a germline control. Tumour DNA was sequenced to an average depth of 80x (range 70-90x) and matched germline DNA was sequenced to an average depth of 69X (range 48-88x). Variant calling was performed as described above. An average of 348 SNVs (99-636) were identified per sample within the covered regions as well as an average of 19 (10-24) indels per sample, corresponding to mean and median total somatic mutation rates of 5.90 and 5.75 mutations per Mb respectively. Of these, an average of 140 (64-241) SNVs and 6 (2-10) indels per tumour were non-synonymous (missense, nonsense, frameshift or mutations of the invariant dinucleotide at splice junctions) (Figure 4.4A). The number of variants did not correlate with sequencing depth (data not shown). A breakdown of this information per tumour can be seen in Table 4.2.

4.2.2.1 Changes in mutational load after chemotherapy

As MMC is a DNA damaging chemotherapy, it was hypothesized that treatment with this agent might lead to an increased number of mutations. Indeed 5 out of 8 patients showed an increase in the number of mutations after MMC treatment (Figure 4.4B) with a mean change of +74 mutations post-MMC. However this was not statistically significant (paired two tailed t-test, p=0.1086) (Figure 4.4C). A high proportion of mutations were shared between the two tumours of most patients (Figure 4.4D). When restricting analysis to non-synonymous mutations, a significant increase in the number of mutations post chemotherapy was observed (two-tailed paired t-test, p=0.0333) (Figure 4.4E).

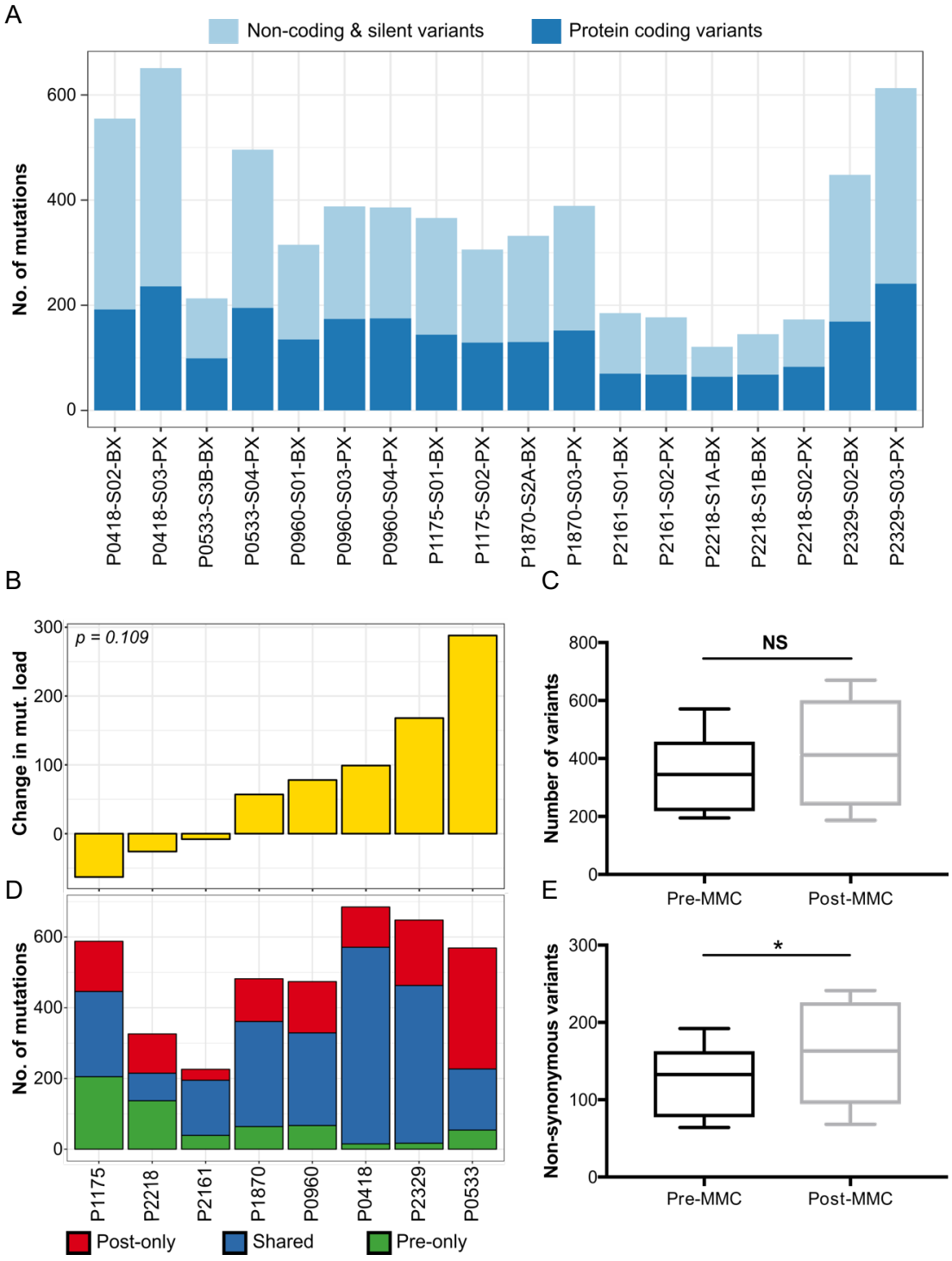


Figure 4.4: Overview of the number of mutations and mutational load.
A) Total number of mutations per tumour; dark blue indicates the number of potentially functional protein coding variants and light blue indicates the number of non-coding and synonymous variants. B&C) Changes in total mutational load per patient from pre-MMC to post-MMC treatment. There is no statistically significant change in mutation load after treatment (mean change = +74, paired t-test p = 0.109). D) Breakdown of mutations per patient. Private mutations for both pre-MMC and post-MMC tumours and mutations common to both tumours are shown. E) Boxplot showing the median and range of potentially functional mutations in pre-MMC and post-MMC tumours. There is a statistically significant increase seen post-MMC (paired t-test *p = 0.0333).

Table 4.2: Details of tumours and information regarding variants identified by whole-exome sequencing.

Grading is according to the 1973 World Health Organisation (WHO) guidelines. Variants were included if they were identified by at least two of the three variant callers. Additional variants that were seen in both tumours but were only called by one variant caller, or only passed the filtering for one tumour were also included. [a]Potentially functional variants include missense, nonsense, frameshift or mutations of the invariant dinucleotides at splice junctions.

Tumour ID	Stage	Grade	Total variants		Potentially functional variants[a]	
			SNVs	Indels	SNVs	Indels
P0418-S02-BX	Tx(a at least)	G3	556	15	190	2
P0418-S03-PX	Tx(1 at least)	G3	654	16	234	2
P0533-S3B-BX	Ta	G2	210	17	96	3
P0533-S04-PX	Ta	G3	491	24	192	3
P0960-S01-BX	Ta	G2	305	24	128	7
P0960-S03-PX	Ta	G1	379	26	166	8
P0960-S04-PX	Ta	G2	377	27	166	9
P1175-S01-BX	Ta	G2	423	23	140	4
P1175-S02-PX	Tx	G2	360	23	126	3
P1870-S2A-BX	Ta	G2	340	21	127	3
P1870-S03-PX	Ta	G2	394	24	148	4
P2161-S01-BX	Ta	G2	174	21	65	5
P2161-S02-PX	Ta	G2	175	12	63	5
P2218-S1A-BX	Ta	G2	114	23	55	9
P2218-S1B-BX	Ta	G2	136	25	58	10
P2218-S02-PX	Ta	G3	166	23	74	9
P2329-S02-BX	Ta	G2	441	21	160	9
P2329-S03-PX	Ta	G1	606	24	231	10

All patients had some private mutations in both pre-MMC and post-MMC tumours alongside mutations that were shared (Figure 4.4D). To account for differences in mutation number between patients, the number of mutations in each group (pre-MMC unique, post-MMC unique and shared) were presented as a percentage. The average percentage of mutations that were shared was 54% (24-81%). Overall, the average percentage of mutations seen as unique pre-chemotherapy was 17% (range 2.2-42%), whilst for post-chemotherapy unique mutations this was 29% (range 13.7-60.1%).

4.2.2.2 Mutational Profile

To gain an insight into the mutational landscape of these tumours the top 50 mutated genes were identified (Appendix H). The most frequently mutated genes were *FGFR3* (75%), *PIK3CA* (56%), *TTN* (56%), *DNAH2* (38%), *KDM6A* (38%) and *TSC1* (31%). A subset of genes with potentially functional (frameshift, nonsense or mutations of the invariant dinucleotides at splice junctions) or missense mutations predicted as potentially deleterious by SIFT and PolyPhen-2 is presented in Figure 4.5. Interestingly, divergent mutations in *TSC1* and *KMT2C* were identified for patients P2218 and P2161 respectively. This could represent possible convergent evolution within paired tumour samples.

Mutations in several chromatin modifier genes were identified including *KDM6A, KMT2C, KMT2D, CREBBP, EP300, ARID1A,* and *STAG2* (Figure 4.5, B-C). These mutations tended to be inactivating, as has been seen in previous studies[82,157,164]. *KDM6A* has previously been shown to have a female gender bias[82], but in this limited cohort *KDM6A* mutations were only seen in the males.

Overall, C:G > T:A transitions were the most common substitution, accounting for 48.8% of substitutions in pre-MMC tumours and 44.6% of substitutions in post-MMC tumours. Histograms showing the full mutation spectrum of substitutions were created to enable visualisation of the overall substitution patterns and look for asymmetry in the substitution spectrums of pre-MMC and post-MMC samples (Figure 4.6, A-B). Pre-MMC and post-MMC histograms showed very similar distributions with no signs of asymmetry. However there was a suggestion of an increase in C:G > A:T transversions in the post-MMC tumours. In order to focus on possible MMC related events, the post-MMC unique variants were extracted. These show a divergent mutation spectrum compared to the pre-MMC variants (Figure 4.6C) with a reduction in the number of C:G > T:A transitions from 48.8% to 36.7%, alongside an increase in the number of C:G > A:T transversions from 12.3% in pre-MMC variants to 32.4% in post-MMC unique variants.

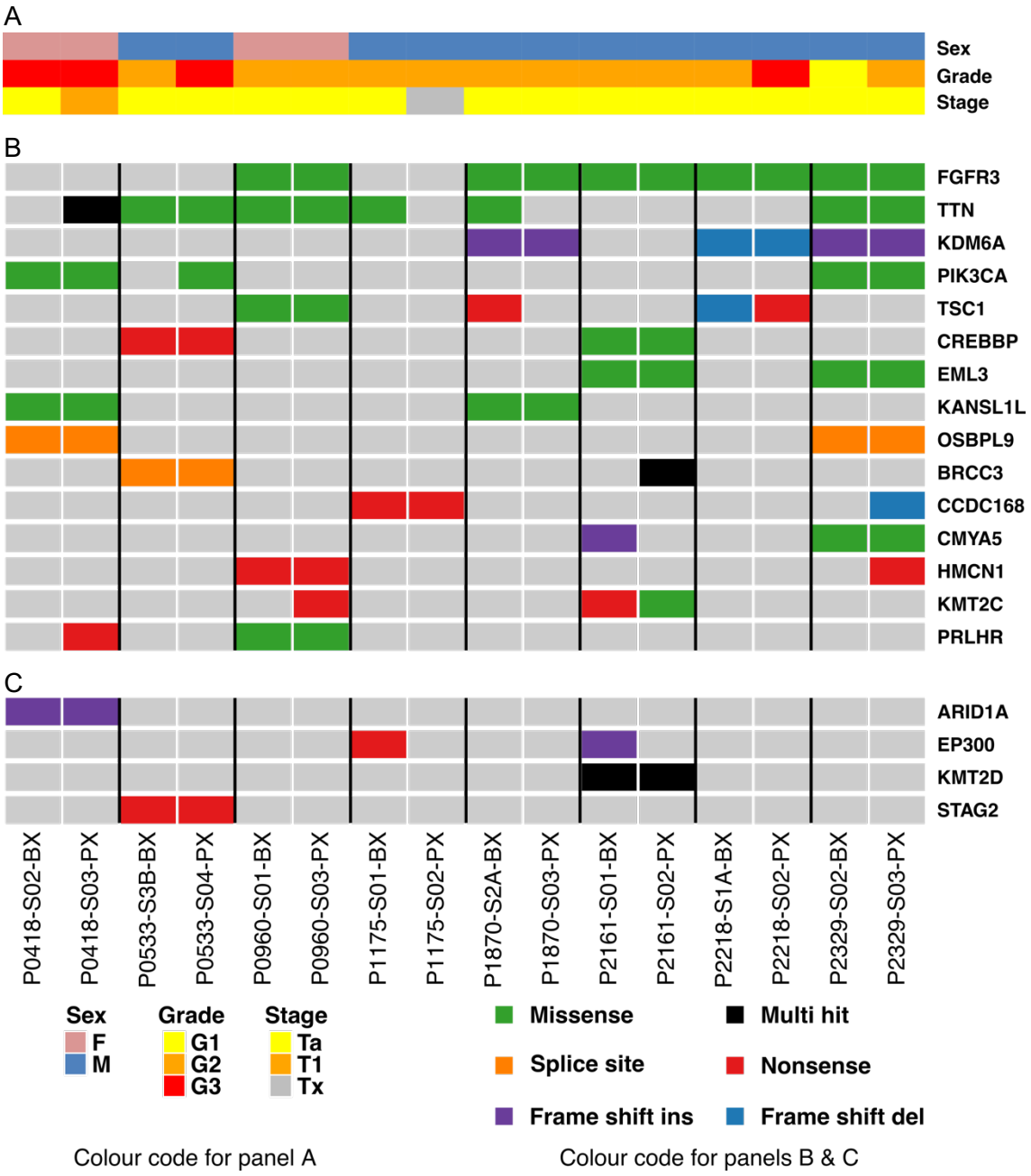


Figure 4.5: Oncoplot showing the distribution of potentially functional mutations identified in selected genes by whole exome sequencing.

A subset of the top 15 frequently mutated genes with loss-of-function mutations (frameshift, nonsense or mutations of the invariant dinucleotide at splice junctions) or missense mutations predicted as potentially deleterious by SIFT and PolyPhen-2 are shown. A) Clinical details for each sample. B) The top 15 mutated genes with potentially functional mutations from pre-MMC and post-MMC tumours. C) Potentially functional mutations in chromatin modifier genes not represented in (B) that are also frequently mutated in NMIBC.

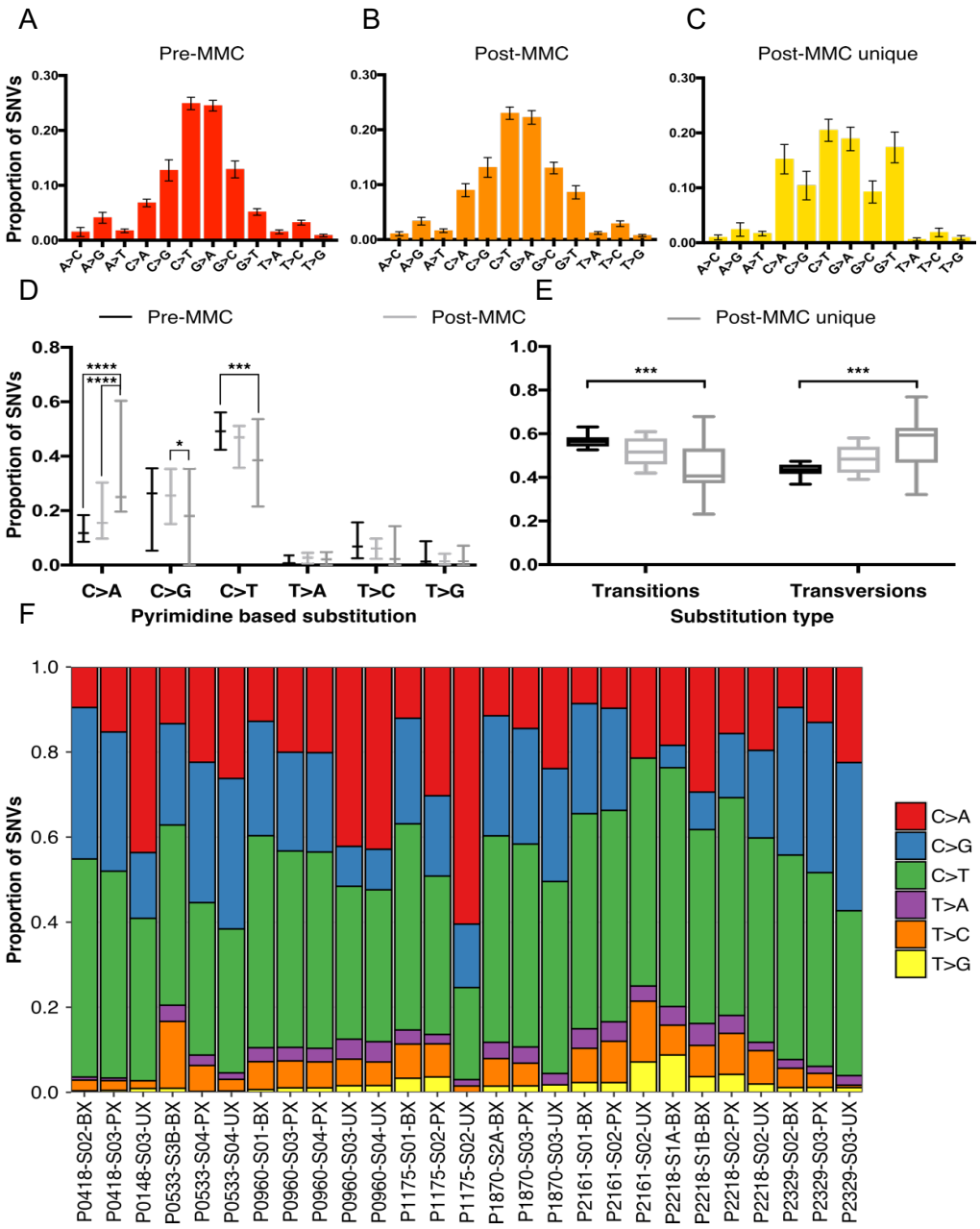


Figure 4.6: Base substitutions in pre-MMC and post-MMC samples.
Base substitutions were identified using the bcftools package with the "stats" command. A-C) Histograms of the substitution spectrum for each variant group (pre-MMC, post-MMC and post-MMC unique variants respectively). Substitutions are depicted as a mean proportion ± SEM. D) Substitutions categorised by the pyrimidine of the mutated base pair. Lines represent the median and range. The y-axis depicts the proportion of mutations with that event. Comparisons between the 3 groups of variants show a significant difference in the number of C > A, C > G and C > T substitutions in the post-MMC unique variants compared to the pre-MMC variants (two-way ANOVA with Tukey's correction for multiple comparisons: **** p = <0.0001, *** p = 0.006, * p = 0.0354). E) Boxplot showing the number of transitions and transversions identified in each variant group. Boxes show the median and interquartile range (IQR) with the lines showing the absolute range. There is a significant difference in both transitions and transversions between pre-MMC variants and post-MMC unique variants (two-way ANOVA with Tukey's correction for multiple comparisons, ***p =0.0010 for both). F) Sample specific mutation spectrum. The relative frequency of each substitution type (designated by the pyrimidine base) is shown for each sample.

To assess the extent of these changes, substitutions were categorised by the pyrimidine base and comparisons were made between the three groups using a two-way ANOVA with multiple comparisons. A significant decrease in the number of C > T transitions was observed between pre-MMC and post-MMC unique variants. Post-MMC unique variants showed a significant increase in the number of C > A transversions compared to both pre-MMC and post-MMC variants. A slight reduction in the number of C > G transversions was also seen in the post-MMC unique variants and this was significant compared to the post-MMC group (Figure 4.6D, two-way ANOVA with Tukey's correction for multiple comparisons: **** p = <0.0001, *** p = 0.006, * p = 0.0354). Despite the decrease in C > G transversions in the post-MMC unique variants, a significant difference was observed in the overall proportion of transitions and transversions between pre-MMC and post-MMC unique variants (Figure 4.6E, two-way ANOVA with Tukey's correction for multiple comparisons, p =0.0010 for both).

To ensure that the changes seen in the bulk analysis were representative of all the patients, the sample specific mutation spectrum was plotted (Figure 4.6F). For all patients except patient P2218 an increase in the number of C > A transversions and decrease in the number of C > T transitions can be seen in the post-MMC samples, and this change is exaggerated when looking at the post-MMC unique variants. Patient P2218 is the only exception to this rule.

Tandem substitutions, specifically at GpG bases, have been linked to MMC intrastrand crosslinks. Quantification of the number of tandem substitutions showed that post-MMC tumours had significantly more tandem substitutions than pre-MMC tumours (paired t-test, p = 0.0158). Overall, pre-MMC tumours exhibited a total of 8 tandem substitutions (per-sample median = 1, range = 0-2), whilst post-MMC tumours exhibited a total of 70 tandem substitutions (per-sample median = 7 range =2-19). Sixty-two of the post-MMC tandem substitutions were unique post-MMC events. To ensure an accurate representation, the numbers of mutations involved in tandem substitutions were converted to proportions (Figure 4.7). This confirmed that the proportion of tandem substitutions was significantly higher in the post-MMC group (p = 0.0441) and post-MMC unique group (p = 0.0224) compared to the pre-MMC group (repeated measures one-way ANOVA). Tandem substitutions predominantly occurred at GpG or CpC dinucleotides in the post-MMC unique variants (97%), whilst pre-MMC tandem substitutions occurred at these bases in only 63% of cases (Table 4.3). The increased proportion of tandem substitutions seen in the post-MMC unique events (mean =

10.78%, range: 1.10-26.95%) along with the specificity of GpG or CpC dinucleotides suggests that these are a MMC-related occurrence.

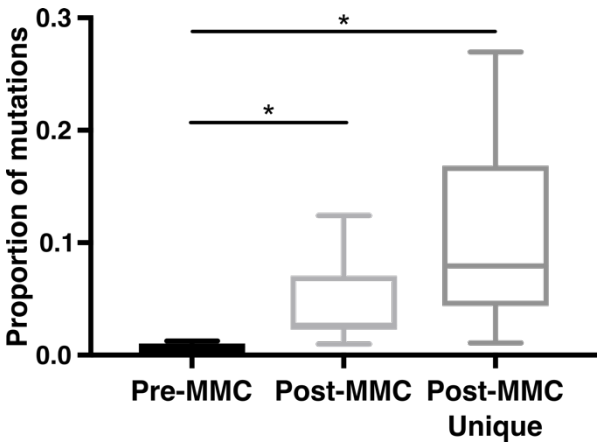


Figure 4.7: Tandem substitutions.
The number of tandem substitutions was identified within each variant group. The boxplot shows the median proportion of tandem substitutions for each group with the IQR and absolute range.*p < 0.05. There is a significant increase in the number of tandem substitutions in the post-MMC variants (p = 0.0441) and post-MMC unique variants (p = 0.0224) compared to pre-MMC variants (repeated measures one-way ANOVA with Dunnett's multiple comparisons test).

Table 4.3: Tandem substitutions.
Variants were combined into pre-MMC, post-MMC or post-MMC unique variant groups. The number of tandem substitutions in each group was identified. This was multiplied by 2 to give the total number of variants involved in tandem substitutions, and from this a percentage of variants involved in tandem substitutions was calculated.

Mutation group	No of tandem substitutions	Total variants	Variants involved in tandem substitutions (%)	Tandem substitutions at GpG or CpC dinucleotides (%)
Pre-MMC	8	2535	0.63	62.5
Post-MMC	70	3139	4.38	92.9
Post-MMC unique	62	1165	10.64	96.8

4.2.2.3 Mutational Signatures

To further characterise the mutational processes and evaluate mutational signatures within the tumour samples the 5' and 3' nucleotides surrounding each SNV were extracted (Figure 4.8). All 3 variant groups show a similar context distribution of C > G and C > T mutations with a high proportion of these mutations occurring at TpCpA and TpCpT sequences, consistent with APOBEC mutagenesis. In the post-MMC unique variants an increase in C > A substitutions is observed with peaks at ApCpG, CpCpA, GpCpG and TpCpC sequences whilst in the pre-MMC group, the few C > A mutations are primarily at TpCpA, TpCpC and TpCpT sequences. The mutational context per patient for pre-MMC, post-MMC and post-MMC unique variants can be seen in Appendix I.

The mutational context was used to create mutation signatures that were decomposed into the 30 signatures described by COSMIC using the mutation-signatures package from MSKCC (https://github.com/mskcc/mutation-signatures). In both pre- and post-MMC tumours, the most common signatures were signatures 2 and 13, both of which are APOBEC-related signatures (Figure 4.9A). Overall 49% of all variants had an APOBEC signature. Signature 1 was the next most common signature. This corresponds to spontaneous deamination of 5-methylcytosine and correlates with age at cancer diagnosis. The percentage contribution of the age signature did not correlate with age (data not shown), but the age range at diagnosis differed by only 9 years (range 67-76). Interestingly signature 7, which is associated with ultraviolet (UV) light exposure, was amongst the top signatures in both pre- and post-MMC tumours but this was not observed within the top signatures in the post-MMC unique variants (Figure 4.9A). An overview of all the mutation signatures identified in each sample can be seen in Figure 4.9B.

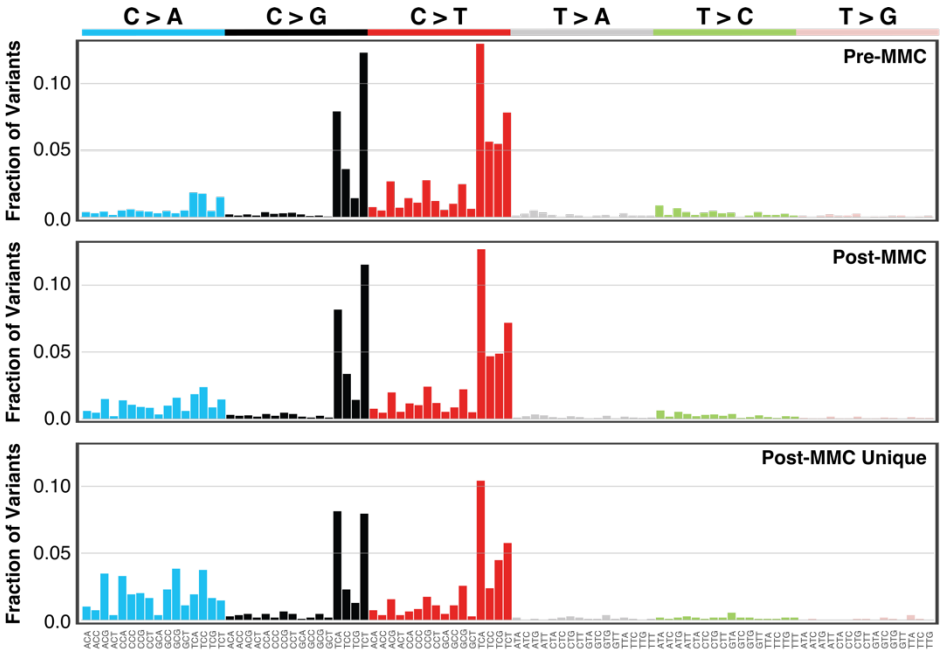
Using the results from the signature analysis, tumours were classified by the contribution of APOBEC signatures, where "high APOBEC" denotes tumours for which the combined contribution of APOBEC signatures 2 and 13 was greater than 30%, as reported in Lamy *et al*[61] (Figure 4.9C). Interestingly only one patient, P2218, had a "low APOBEC" score and this was for both pre- and post-MMC tumours. All other tumours had a "high APOBEC" score. When focusing on post-MMC unique variants, only 50% of tumours had a "high APOBEC" score. For patients P0418, P0960, P1175 and P2161 a "low APOBEC" score was seen in the post-MMC unique variants despite these tumours having an overall "high APOBEC" when looking at all variants. A reduction in APOBEC score was also seen in patients P1870 and P2329. However these patients

still had a high level of APOBEC contribution to post-MMC unique variants. This suggests that APOBEC mutagenesis was less prevalent for these tumours post-MMC compare to pre-MMC.

For two patients the APOBEC score increased post-MMC. For patient P0533 an increase in APOBEC contribution from 35% pre-MMC to 56% post-MMC was seen. Indeed 65% of post-MMC unique variants likely came from APOBEC derived mutagenesis for this patient. Most interestingly, for patient P2218 a “high APOBEC” score was seen in the post-MMC unique variants despite this patient having a “low APOBEC” score overall. This suggests that for these two patients APOBEC activity increased over time.

Signatures that contributed to more than 4% of the mutational spectrum for each group were compared using a two-way repeated measures ANOVA with multiple comparisons (Figure 4.9D). A significant reduction in the contribution of signatures 1 and 13 was seen in the post-MMC unique variants. Most interestingly, a very significant increase in the contribution of signature 24 to the mutational signature was seen in the post-MMC unique variants (p = <0.0001). Overall for pre-MMC variants, signature 24 contributed to only 4.5% of variants and this increased to 31% for post-MMC unique variants. Analysis of the contribution of each signature per patient shows an increase in the contribution of signature 24 in post-MMC unique variants for every patient except patient P2218. This large increase seen in post-MMC unique variants suggests that this could reflect the MMC mutagenic process. Signature 24 has been seen in a subset of liver cancers and has been linked to exposure to aflatoxin and is characterised by C > A mutations. Signature 4 also increased in contribution when analysing the post-MMC unique variants. Signature 4 is present at a low level in pre-MMC tumours, contributing an average of 4.1 % of mutations in this group and in post-MMC unique variants this contribution increased to 12%, but this increase is not statistically significant.

A



B

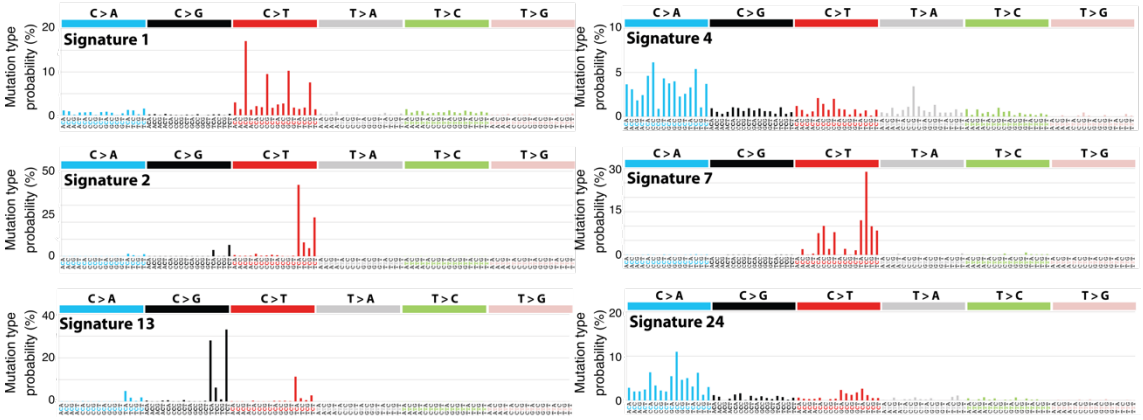


Figure 4.8: Mutational context in pre-MMC, post-MMC and post-MMC unique variants and COSMIC signatures.

A) Variants were grouped into pre-MMC variants, post-MMC variants and post-MMC unique variants. The 5' and 3' nucleotides surrounding each SNV were extracted and the sequence context plotted. B) COSMIC signatures identified within the samples (taken from https://cancer.sanger.ac.uk/cosmic/signatures).

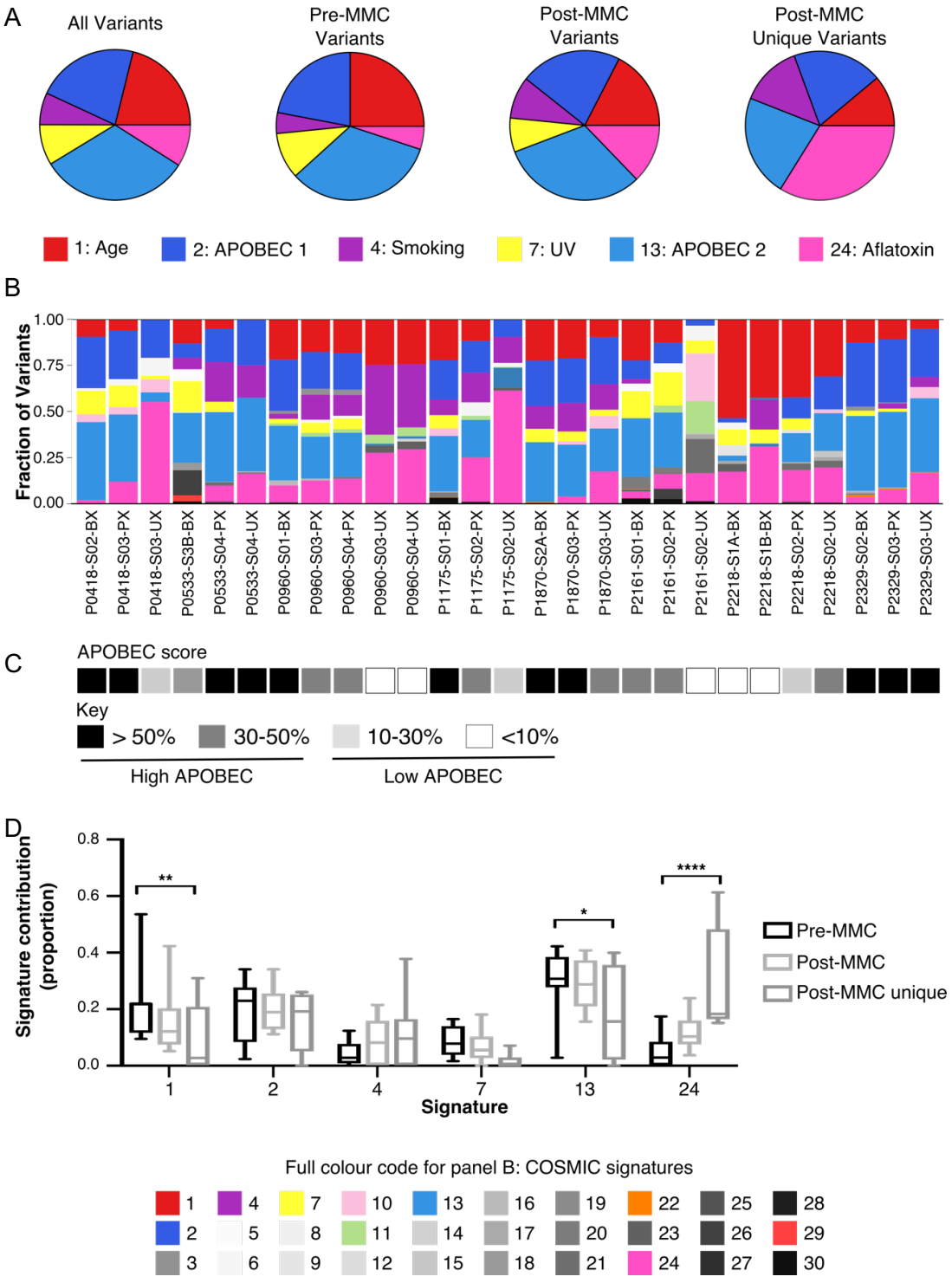


Figure 4.9: Mutational signature analysis.
Mutational signatures identified in the patient cohort. A) Signatures contributing to more than 4% of variants were assessed in the different variant groups. B) The contribution of each signature is shown per sample as well as for the post-MMC unique variants. Samples are grouped by patient. C) The contributions of the two APOBEC signatures were combined to create an APOBEC score. A score greater than 30% was designated “high APOBEC”. High and low scores have been differentiated further for greater clarity. D) Changes in contribution of the signatures were assessed using a two-way ANOVA with repeated measures by both factors. Post-MMC and post-MMC unique contributions were compared to pre-MMC contributions and Dunnett’s test was used to correct for multiple comparisons (* $p = 0.0106$, ** $p = 0.0047$, **** $p < 0.001$).

4.3 Discussion

This chapter investigated the mutational changes seen in paired tumours after a treatment course of mitomycin C (MMC) with the aim of identifying MMC chemotherapy-associated genomic changes. A variant calling pipeline was developed and used to identify somatic variants from 18 matched pre-MMC and post-MMC treatment tumours. These have been assessed for changes in mutational load after therapy, the types of base substitutions seen and mutational signatures, and frequently mutated genes have been identified.

4.3.1 Variant calling pipeline

The aim of best-practice bioinformatic pipelines is to decrease the number of false-positive and false negative variants detected in order to increase precision[242]. However, for the analysis of paired tumour data these practices can be too stringent and cause the filtering out of important low-frequency variants, often from only one tumour. To circumvent this problem an adapted variant calling methodology was adopted for the identification of shared variants. This methodology was an adaption of that described by Droop *et al.*[242] who suggest a two-stage approach with a round of reduced stringency calling for the identification of shared variants followed by a round of high stringency calling to identify unique mutations. To ensure adequate detection of shared variants, all variants that were identified as shared by a single caller were included and unique variants were checked against raw variants (variants detected prior to filtering) identified by the MuTect2 algorithm. Initial studies with the output from MuTect2 identified an additional 20 mutations for the two patients assessed.

Five variant callers were investigated to identify the best combination of three callers for consensus calling. Overall, VarScan2 was the least suited for this type of data analysis. VarScan2 is a high-stringency variant caller which uses a heuristic threshold methodology. It will identify potential variants that pass the thresholds (at least 4 supporting reads and 20% VAF) and employs a Fisher's exact test on the read counts between tumour and normal in a 2 x 2 contingency table to filter out any germline variants[302]. These settings are designed to filter out artifacts present at low levels but this will also filter out any variants with a low VAF that arise from subclones or as a result of impure input material. In our hands, relaxing the thresholds did not increase shared variant discovery so the default settings were used. This inability to identify low-frequency variants is the probable reason why so many of the unique variants identified by VarScan2 were actually called as shared by another caller. With appropriate

threshold settings, VarScan2 has been shown to be able to achieve sensitivity as low at 5% VAF[305], however this was with high coverage and requires optimization.

Of the remaining four variant callers, MuSe identified the fewest variants and had the highest percentage of incorrect calls. MuSe is also the only variant caller that does not identify indels. As the literature has suggested that an increase in indels may be observed in MMC treated samples, it is important that such variants are identified. A combination of MuTect2, Strelka2 and EBCall was therefore selected as this set of callers had the lowest incorrect unique call percentage and identified the most variants in combination.

MuTect2, Strelka2 and EBCall all use an allele frequency analysis approach which allows the discovery of minor subclones[301,302]. This makes them particularly useful for the investigation of intra-tumour heterogeneity or impure tumours. However, this ability to call low variant alleles may increase the number of false-positives due to the miss-calling of sequencing noise. By using consensus calling the number of false positives should be reduced as variants need to be called by at least two of the variant callers and are thus more likely to be real.

It is unlikely that there is a single “best” caller[303] and the type of variant caller to use largely depends on the question being asked. Benchmarking studies have shown that using an ensemble approach gives good results even when poor performing pipelines are included[304,306,307]. For this study it is important to be able to identify the shared variants as well as the variants unique to each tumour. The two-pronged method we have devised should increase the number of shared variants identified and reduce the number of false-negatives, whilst the ensemble calling should reduce the number of false positives in the unique data.

4.3.2 Mutation burden

In this cohort the total mutation rate (5.90 mutations per Mb) was higher than that reported previously for NMIBC by Hurst *et al.*[82]. In their study, Hurst *et al.* reported the clustering of low-stage low-grade tumours into two distinct groups using CNAs. These clusters were denoted genomic subtypes 1 and 2 (GS1, GS2). GS1 was characterized by no or few CNAs whilst GS2 was more genomically unstable with characteristic loss of chromosome 9. An overall mean somatic mutation rate of 2.41 mutations per Mb was identified. However, when assessing the individual subtypes, a higher mutation rate of 3.75 mutations per Mb was noted for GS2 tumours compared to only 1.85

mutations per Mb for tumours from GS1. In Chapter 3, tumours from our cohort were clustered by their copy number profiles alongside the tumours from Hurst *et al.*[82]. Tumours from this exome panel mostly clustered with tumours in GS2 . Only 2 tumours clustered with tumours from GS1. This higher mutation rate seen in GS2 tumours is still not as high as the mutation rate identified in our cohort. This may be partially due to the stage and grade of the tumours analysed: Hurst *et al.* assessed only primary, low-grade Ta tumours (though a handful of tumours were re-classified as high-grade when assessed by a single pathologist), whilst this cohort contained recurrent tumours of mixed stages and grades, including a stage T1 grade 3 tumour as well other grade 3 tumours. Additionally, Hurst *et al.* used a more stringent variant calling methodology which may have missed some low allele frequency variants leading to an underestimate of tumour mutational burden.

This difference in overall tumour mutation burden between our study and that of Hurst *et al.*[82] could be interesting. Few studies have specifically compared the mutation rate of high-grade and low-grade tumours. The study by Pietzak *et al.*[157] contained tumours with a mix of stages and grades; including stage Ta, stage T1 and muscle-invasive tumours. They identified a median mutation rate of 9 mutations per Mb in high-grade NMIBC compared to a mutation rate of 7 mutations per Mb in low-grade tumours. These values are much higher than previously reported, higher even than those reported for MIBC[81]. This may be due to the use of a targeted gene panel of cancer associated genes compared to a whole exome panel. These conflicting results identify the need for sequencing larger cohorts of NMIBC to gain insight into the mutational landscape of tumours of all stages and grades and exemplify the need for adjusting mutation rates identified by targeted gene panels. It would be interesting to see if somatic mutation rates follow a similar picture to that of CNAs, where more of the genome is altered by CNAs in tumours of a higher stage and grade[125].

All tumours contained a significant proportion of private mutations. However, most patients had a higher proportion of shared variants compared to private mutations. The average proportion of mutations private to pre-MMC tumours was 17% compared to 29% in post-MMC tumours. A significant increase in the number of non-synonymous mutations was also identified despite no significant increase in the total number of mutations. This is in contrast to what has been observed in MIBC treated with cisplatin where a reduced number of private variants was seen after treatment. This correlated with an overall reduction in variant number seen after treatment but this was not significant[176]. This could suggest that cisplatin is a less mutagenic chemotherapy agent

or that it is more toxic, resulting in an increased cell death and thus leaving fewer cells in which new mutations can be induced.

The significant increase in non-synonymous mutations after MMC treatment could be a reflection of the CpG specificity of MMC as CpGs are enriched in exons and promoter regions relative to the rest of the genome[206]. It would be interesting to use WGS to see if new mutations after MMC treatment cluster in the promoter and coding regions compared to the rest of the genome to confirm this pattern suggested by WES.

4.3.3 Mutational profile

FGFR3 and *PIK3CA* were the most frequently mutated genes and these were mutated at a level similar to that reported in other sequencing studies[82,157]. *KDM6A* mutations were seen at a lower level than in previous sequencing studies of stage Ta tumours. In those studies, the mutation frequency ranged from 50-65%[82,157,164], whilst in this study it was 38%. In previous studies, splitting stage Ta tumours into low and high-grade identified that whilst 52% of low grade tumours contained a *KDM6A* mutation[82,157], only 38% of high-grade tumours contained a mutation[157]. This is the same level as seen in our exome cohort. *KDM6A* mutation has previously been identified as having a female gender bias[82], but in this cohort mutations were only seen in the males. This could be due to the very low patient number in this cohort.

TSC1 was frequently mutated with 31% of patients carrying a mutation. This is much higher than seen previously across all stages and grades[81,82,157,164]. Hurst *et al.* saw a higher frequency of *TSC1* mutations in their GS2 subgroup (19%) compared to their GS1 subgroup (4%). As all but two of the tumours in this cohort clustered with the GS2 tumours this could go some way to explaining the high occurrence of these mutations. *TTN* was also seen to be frequently mutated with all missense mutations identified in this gene predicted to be potentially deleterious by SIFT and PolyPhen-2. *TTN* codes for a very large protein, Titin, which contains over 34000 amino acids. Its large size combined with its presence in late replicating regions of the genome contribute to an increased likelihood of DNA repair error and mutations in this gene are usually considered passenger events[308]. However, mutation frequencies of *TTN* differ amongst different cancers and mutually-exclusive mutation patterns have been identified, suggesting that the role of *TTN* still remains to be evaluated in cancer progression[309].

Investigations into the substitutions occurring in the samples showed some interesting results. C > T transitions followed by C > G transversions have been shown to be the

most common substitutions in both MIBC[81] and NMIBC[61,82,164] with C > T transitions accounting for approximately 51% of mutations in MIBC[81] and 47% in NMIBC[82]. In this data set, C > T transitions accounted for approximately 46.7% of substitutions overall, comparable to previous NMIBC studies[82]. C > G transversions were the second most common event overall. When analysing the post-MMC unique variants, C > T transitions remained the most common substitution, but the proportion of mutations with this substitution was significantly less, and a significant increase in the number of C > A transversions was seen. In this group C:G > A:T mutation was the second most common substitution event. A significant reduction in the proportion of C > T transitions was also seen by Lamy *et al.* when comparing the initial vs latest tumours from patients with NMIBC with progressive disease[61]. In their patients, an increase in C > G transversions accompanied this change but this was non-significant. No change in the proportion of C > A transitions was seen in that study. This suggests that different mutational mechanisms may be at play in tumours from Lamy *et al.*[61] compared to our dataset. This may be because they looked at progressive disease or because the tumours were treated differently. No treatment information was available for tumours in that study.

Only a few studies have characterised the mutations induced by MMC. In the study of Srikanth *et al.*[201], base substitutions accounted for 76% of the mutations identified. The majority of these (84%) were transversions, 69% of which were C:G > A:T. Transversions were also the most common substitution event in the study of Maccubbin *et al.*[202]. Again, C:G > A:T transversions were most common, but the overall contribution of such transversions was reduced compared to the level seen by Srikanth *et al.* making up only 29.3% of all substitutions. This evidence would suggest that the large increase in C:G > A:T transversions seen here in the post-MMC unique variants is likely due to MMC-induced mutagenesis.

Interestingly, 30% of the base substitutions seen by Srikanth *et al.* were tandem substitutions, all of which were identified at GpG dinucleotides. Takeiri *et al.*[203] also identified tandem substitutions, primarily at GpG dinucleotides, in their MMC treated mice. A significant increase in tandem substitutions, specifically at GpG or CpC dinucleotides was identified in the post-MMC unique events in our tumours. The tandem substitutions seen after treatment with MMC are thought to be the product of the intrastrand crosslinks[202]. Given the specificity of this adduct for GpG dinucleotides, combined with the observation that 97% of post-MMC unique tandem substitutions

were observed in a GpG or CpC, it is highly likely that these variants are caused by MMC-chemotherapy and may serve as a marker for MMC-related events.

It is interesting that the mutations identified in these tumours more closely reflect results from previous *in vitro* studies than *in vivo* studies. SNVs were not identified as increased compared to controls in two *in vivo* studies[203,205]. However both studies did see an increase in deletions. In Takeiri *et al.*[203] these ranged from 110 bp to 8 kb in length whilst Tam *et al.*[205] identified deletions ranging from 2 bp to 318.8 kb in length. This increase in deletions was not identified here in the MMC-treated tumours. This could be due to the size of the deletions as somatic variant callers tend to only be able to call short indels[302]. Indeed, Strelka2 will only detect indels up to a pre-defined value of 49 bp[238], and for Mutect2 and EBCall this value is not defined. It is therefore possible that we missed the majority of deletions caused by mitomycin-C due to their size being above the maximum detection limit of our exome pipeline, yet below the minimum detection limit for our shallow-pass whole genome sequencing (WGS) pipeline.

4.3.4 Mutational context and signatures

The sequence context of variants provides an insight into the mutagenetic processes that have taken place within a cancer. These can be delineated into signatures characteristic of each mutational process and multiple signatures may be identified within a tumour. By counting the number of mutations that contribute to each signature, a level of exposure to each mutational process can be determined[73]. Analysis of individual tumours can only give a historical overview of the mutagenic processes that have occurred and cannot tell which processes are still ongoing. Assessing multiple tumours over time and specifically analysing the “new” mutations can give important information about ongoing processes.

Six signatures were identified that contributed to more than 4% of variants in the pre- and post-MMC tumours. APOBEC related signatures, signatures 2 and 13, were the most prevalent with 49% of all variants coming from an APOBEC signature. In NMIBC, enrichment of APOBEC mutations has been associated with high risk tumours[61,310]. Lamy *et al.*[61] compared tumours from patients with progressive disease to those with non-progressive disease and saw that two thirds of patients with progressive disease were classified as “high APOBEC” compared to only a third of patients with non-progressive disease. Hurst *et al.*[82] also identified that APOBEC mutagenesis was increased in GS2 tumours relative to GS1 tumours. Out of the 8 patients analysed

here, 7 of them had “high APOBEC” scores in both tumours. However, the tumours with “low APOBEC” scores clustered into GS2.

When focussing on the unique mutations post-MMC treatment, half of the patients had a “low APOBEC” score despite all of these patients having a high score overall. This implies that fewer of the new variants were due to APOBEC mutagenesis. This may suggest that APOBEC mutagenesis could have been an early event in these tumours that was reduced or switched off later on, or that a different mutational process is contributing more to the process than APOBEC mutagenesis. Indeed, in the 4 patients for whom a change from high to low APOBEC scores were seen in post-MMC samples, a large induction in variants attributed to signature 24 was seen. Another possible contribution to the reduction in APOBEC mutations could be due to the short time frame between tumours. For a mutation to be identified, sufficient time is required for the cells containing the mutation to proliferate and become present in large enough numbers for detection in bulk sequencing. With the maximum time frame between tumours assessed in this cohort being 13 months, and with some patients having other tumours resected in-between analysed tumours, this leaves a relatively short time for new mutations to both occur and expand to a detectable level. This would also explain the reduction in the contribution of the age signature to the post-MMC unique events.

Previous studies have reported conflicting results regarding the timing of APOBEC mutagenesis[61,180]. Lamy *et al.*[61] suggested that it is a later tumour-specific event, whilst data from synchronous multifocal tumours identified more APOBEC mutations shared between tumours compared to tumour unique events, suggesting an early event[180]. Despite 50% of the patients in our study having a “low APOBEC” score in their post-MMC unique variants, 50% still had a high “APOBEC” score and for two patients the APOBEC score increased in the post-MMC unique variants. This data set therefore implies that APOBEC mutagenesis is a tumour/patient specific event that can follow multiple paths. It can occur early and reduce in prevalence over time as seen in the patients for whom APOBEC scores were reduced post-MMC (P0418, P0960, P1175 and P2161). It can occur early and remain active over time, as seen in the tumours from patients that maintained or increased their APOBEC scores (P1870, P2329, P0533) or it can be a later event, as seen in patient P2218. The evaluation of more tumours is required to understand if the timing of APOBEC mutagenesis has any links to prognosis.

Signature 1 was the next most common signature and was identified in all tumours. Signature 1 has been demonstrated to correlate with age of diagnosis in both adult and childhood cancers[70,311,312]. The mutational process underlying the signature is thought to be deamination of 5-methylcytosine at CpG dinucleotides creating C > T transitions[311]. In our data set no correlation between the age at diagnosis (or at resection of the pre-MMC tumour if this was not the diagnostic tumour) and the contribution of the age signature was identified. This could in part be due to the small age range in the patients, but could also be due to the recurrent nature of these tumours: some of the tumours are not the primary tumour and previous resections may have reduced the number of these mutations in later tumours. Indeed, the number and proportion of signature 1 mutations reduces in the post-MMC tumours from almost all patients. Additionally, few unique mutations post-MMC treatment (-UX) demonstrate an age signature and this could be a reflection of the short period of time between pre- and post-MMC tumours for the generation of new mutations.

The age signature is reported to be prone to artefacts; the deamination of 5-methylcytosine can be observed after formalin fixation of tumour samples[313,314], can occur during DNA isolation[315] and can be accelerated by heating[316]. As none of the samples in this cohort underwent formalin fixation, only errors caused during DNA extraction and library preparation are of concern. The presence of these errors in the dataset could contribute to the lack of correlation between the contribution of signature 1 and the age of the patient. With this in mind, it may be prudent to check low VAF C > T mutations at CpG dinucleotides with an orthogonal method to ensure that these are not artefacts from the sequence preparation procedure.

Signature 24 was found to contribute to significantly more mutations in the unique post-MMC variants compared to pre-MMC. Signature 24 was first discovered in 5 liver cancer samples by whole-exome sequencing and verified using ICGC and TGCA WES data which identified 6 further tumours with this signature[317]. All patients were African or Asian with six tumours containing a somatic *TP53* p.R249C. This mutation is typical of aflatoxin B_1 exposure[318] and suggested that aflatoxin B_1 may be the cause of signature 24. This has since been confirmed through the sequencing of a larger cohort[319], sequencing of a cohort from a high aflatoxin risk area[320] and experimentally using cell lines and mouse models[321].

Aflatoxin B_1 is a common contaminant of foods such as peanuts, corn and rice[322]. It has long been known to be a carcinogen and it targets guanine bases[323] creating G:C

> T:A transversions[324]. Aflatoxin preferentially binds at GGG sequences but will also bind at XpGpY where X is G or C and Y is G or T[324]. Aflatoxin exposure of these patients was unknown, and for seven out of eight patients, signature 24 only became a major mutational contributor post-MMC treatment. It is unlikely that all patients were exposed to aflatoxin at this time.

Signature 4 was also seen at a higher level in post-MMC unique variants compared to pre-MMC variants. Signature 4 is associated with tobacco smoke and whilst it is commonly seen in lung, liver and head and neck cancers it is not usually seen in bladder cancer[4,81,325]. Signature 4 exhibits transcriptional strand bias for C > A mutations and is also associated with CC > AA dinucleotide substitutions (https://cancer.sanger.ac.uk/cosmic/signatures).

MMC also targets guanine bases creating G:C > T:A transversions primarily at GpG or CpG sites and can cause CC > AA dinucleotide substitutions. Therefore it is unsurprising that mutational signature analysis has extracted signatures 4 and 24 as key contributors in the post-MMC variants. Indeed, comparisons of the peaks of C > A mutations within their trinucleotide context show that many of the peaks are at the same place. Signature 24 has peaks of C > A substitutions at GpCpC, CpCpA and TpCpC sequences, signature 4 has peaks at CpCpC, TpCpC and CpCpA sequences and in the post-MMC unique variants there are also peaks at CpCpA, and TpCpC sequences. However, there are some differences; the post-MMC unique variants also show peaks at ApCpG and GpCpG sequences, neither of which are large contributors to signatures 4 or 24. C > A mutations in a GpCpA context are also contributors to both signatures 4 and 24, but mutations were very infrequently seen in this context in the post-MMC unique variants. Perhaps with a greater sample number it would be possible to separate the MMC signature from signatures 4 and 24.

The are several limitations to the signature analysis performed in this study. One of the limitations is the use of WES sequencing data; whilst some of the tumours contained up to 654 mutations many contained much fewer, especially when looking at the unique post-MMC mutations. Indeed, patient P2161 had fewer than 30 post-MMC unique mutations making signature analysis unreliable for this particular mutational group (P2161-S0-UX). As the exome is a small percentage of the genome, only a fraction of the mutations present in a tumour may be identified. Whole genome sequencing on the other hand produces many more mutations, providing an increased power for signature decomposition resulting in a more reliable signature determination;

Alexandrov *et al.* [312] were able to detect a much larger number of signatures from a much smaller number of whole-genomes compared to exomes. Additionally, codon usage in exons may be constrained and this could mean that some signatures are less likely to be identified in exons whilst others may be more likely[326] and some mutations may be under selection[327], potentially skewing the results generated from WES. A study comparing the signatures identified from the whole-genomes of 323 liver cancers with signatures identified using only the mutations from the exomes of those samples demonstrated that use of only the exome mutations resulted in the identification of fewer signatures with a higher background noise. These signatures also differed from those identified using the whole genome data, questioning the validity of using exome sequencing for mutational signature analysis[327].

As the sample set was small, with relatively small numbers of mutations, signatures were identified using a signature fitting method from Memorial Sloan Kettering Cancer Centre (MSKCC). The package, mutation-signatures, extracts the sequence context for each mutation creating a mutation profile per sample. The algorithm then attempts to identify the mutational signatures underlying the observed mutational profile using the 30 signatures described by COSMIC as a reference and an iterative multiple linear regression strategy[328]. This restricts the analysis to known signatures, excluding the chance of identifying new signatures. This decomposition method used has limitations highlighted by the identification of the UV signature as well as the likely misclassification of MMC-related mutations to signatures 4 and 24 as discussed above. It is highly unlikely that mutations identified within a bladder tumour are caused by exposure to UV. A separate signature fitting decomposition method, deconstructSigs implemented through mSignatureDB[329] (available at http://tardis.cgu.edu.tw/msignaturedb/), demonstrated the same signature identification as the MSKCC method including signature 7 (data not shown). Additionally, evaluation of the TCGA bladder cancer data using this portal also demonstrated the presence of signature 7. This data could suggest that there may be a signature present in bladder cancer similar to signature 7. However, a known limitation of signature fitting methods is that mutational profiles arising from a combination of signatures may be best fitted by an alternative single signature that resembles the mutational profile but does not actually contribute to the process. This is particularly applicable to C > T transitions as a third of the signatures in COSMIC are mostly composed of these transitions[328]

An alternative approach to identifying mutational signatures would be to use the non-negative matrix factorisation (NMF) method as described by Alexandrov *et al.* [70].

Signatures identified using NMF can be compared to the known signatures using cosine similarity, allowing for the identification of both known and novel signatures. However, using NMF for the identification of mutational signatures requires a large number of variants, either from a large sample number with few variants or from a small sample number with a large amount of variants. For example, 25,000-50,000 mutations are needed to decipher 5 mutational signatures depending on the number of mutations per sample[330]. With fewer than 4000 different mutations identified in total in our cohort (some mutations were present in multiple tumours so have only been counted once in this instance) this significantly limited the potential for identifying any but the strongest mutational signatures within our cohort using NMF methods, therefore the signature fitting method was used.

Since performing the analysis new methods have been published including a package called MutationalPatterns that combines NMF and signature fitting methods to identify both new signatures and quantify the contribution of previously known signatures[331]. Perhaps a greater sample number and WGS combined with a more accurate signature identification using both NMF and signature fitting would enable the identification of a MMC signature that can be distinguished from signatures 4 and 24.

Despite the wealth of information generated by the computational analysis of cancer genomes in the detection of mutational signatures, it is clear that further analysis is still required to fully elucidate the mutation spectrum and delineate partially correlated signatures. The work presented in this Chapter has highlighted that different mutational processes can produce similar signatures. The Pan-Cancer Analysis of Whole Genomes network (PCAWG) has recently used over 84 million somatic mutations from 23,829 samples of different cancer types in the largest analysis to date. They have identified 49 single base substitution, 11 doublet substitution, 4 clustered base substitution and 17 indel mutational signatures. The data is as-yet unpublished, but the paper is available as a pre-release[312]. It would be interesting to see if this increased power would delineate the MMC signature further.

Further validation of this signature of MMC chemotherapy is required through the analysis of more tumours with exposure to the drug. With the advances in technology since the studies of Maccubbin *et al.*[202] and Srikanth *et al.*[201], further experimental analysis of the mutation signature should also be attempted to confirm the activity of MMC. This could be done on cancer cell lines or in mouse models. This is important as

similar signatures could be generated by several different mutagenic processes and confirmation that the MMC signature is similar to the aflatoxin signature is important.

4.4 Summary

Overall the data presented in this Chapter has shown that a 6-week course of MMC into the bladder can induce new mutations unique to post-treatment tumours. Analysis of these mutations has identified a mutational signature consisting of an increase in the number of C > A transversions and an increase in tandem substitutions, specifically at CC or GG bases. These mutations identified in the post-MMC treated tumours match mutations reported in early studies investigating the effect of MMC on DNA making it highly suggestive that these mutations are reflections of MMC induced mutagenesis. In the next Chapter the evolutionary path of the cancers will be determined and the clonality of the MMC-induced mutations will be assessed.

Chapter 5
Subclonal composition of pre-MMC and post-MMC tumours and targeted sequencing of additional tumours for the determination of clonality

5.1 Introduction

Cells within a tumour are constantly competing for space and resources, following a Darwinian based evolutionary model[17]. This often results in tumours which consist of multiple cell populations that are genotypically distinct[53,332,333]. The comparison of shared and unique mutations provides key information about differentially altered genes across samples and may identify key genes that are mutated in a certain subgroup of samples, such as after treatment. However, this does not provide information on the frequency of a mutation which could be present in all the tumour cells or just a subpopulation. Analysis of mutations that have expanded from a subclonal to clonal prevalence can provide an understanding of genetic events that contribute to progression, recurrence and resistance[242]. Likewise, clonal populations that have been eradicated after treatment can provide information about alterations that confer sensitivity to treatment. Measuring the prevalence of a clone over time or space therefore provides an indication of the fitness of that clone[334]. In Chapter 4, we identified shared and unique mutations in each pre-MMC and post-MMC tumour pair as well as mutations associated with MMC treatment. In this chapter, we analyse those mutations to identify potential clonal and subclonal populations present within the tumours.

Mutations can be classified as clonal or subclonal through the analysis of the variant allele frequency (VAF)[110]. The VAF is the fraction of reads supporting the variant allele rather than the reference allele and this can provide an estimate of the proportion of the cells containing the mutation[335]. Clustering mutations with a similar VAF can identify tumour subclones as mutations present at similar proportions in a population are likely to be present in the same cells[335]. Additionally, using multiple related samples can further delineate clusters through the identification of mutations that shift together when comparing tumours[246]. Unfortunately, the VAF can be affected by changes in copy number (CN) or regions of loss of heterozygosity[336] as well as contaminating normal cells[110], meaning that analysis of VAF alone can only provide a rough estimate of subclonal populations within tumours.

There are many algorithms available for the identification of subclones and intratumour heterogeneity from bulk tumour sequencing[246,336-338], yet many of these do not take copy number alterations (CNAs) into account. To provide a more accurate measurement of subclonal populations, some algorithms incorporate CN data and tumour purity estimates into their predictions. In this project, clonal clustering was performed using PyClone[246]. This is a Bayesian clustering method that groups somatic mutations into putative clonal clusters whilst accounting for both CN and tumour purity[246]. To enable accurate clonal clustering PyClone requires allele-specific CN estimations for each mutation. Allele-specific CNAs cannot be identified using shallow-pass whole genome sequencing (WGS) therefore we utilized the whole exome sequencing (WES) data from Chapter 1 to generate allele-specific CN data for each of the tumours and these results were compared with CNAs from the shallow-pass WGS.

In this Chapter, the VAF and allele-specific CN of each SNV was used as the input data for PyClone[246] in order to estimate the proportion of cells harbouring a mutation (referred to as “cancer cell fraction (CCF)”). Mutations with similar CCFs were clustered to identify clones present within each tumour. Temporal ordering of the clones was performed using ClonEvol[247] to infer the order of evolution. This data was then used to understand the evolutionary dynamics of each tumour and examine the clonality of MMC-associated events.

Four of the eight patients in the WES cohort had additional tumours available for analysis; patients P0418 and P2329 both had an additional tumour prior to the pre-MMC tumour, patient P0533 had 3 additional tumours prior to the pre-MMC tumour and patient P1870 had one tumour before the pre-MMC tumour and one tumour after the post-MMC tumour. To supplement the information obtained from WES, these tumours underwent next-generation-based targeted sequencing of a panel of 140 genes that have previously been identified as mutated in either muscle-invasive or non-muscle-invasive bladder cancer[82,100,164,170-172,325]. A list of the genes can be found in Appendix D.

In Chapter 3, single gene mutations and CNAs were used to assess clonality between tumours from the same patient. However, for 8 patients this data was insufficient to accurately discern a monoclonal or oligoclonal origin of their tumours. In this Chapter, we used targeted sequencing of the gene panel described above to provide a more in-depth examination of clonality for these patients.

5.2 Results

5.2.1 Allele-specific copy number estimation

Mutations found in regions of CN loss, CN gain or CN neutral LOH will have an altered VAF which needs to be considered during clonal clustering and ordering. Clonal clustering programs, such as PyClone, are more accurate when used with allele-specific CN information but shallow-pass WGS does not provide enough depth for the detection of allele-specific CNAs. In order to accurately reconstruct the clonal evolution of the tumours the package FACETS[245], an allele-specific CN analysis tool, was utilized with the whole-exome sequencing data. The output from FACETS was compared with the output from the shallow-pass WGS CN data generated by Nexus to ensure that all major CNAs were detected by the FACETS package. The allele specific CN data from FACETS was then used for clonal clustering.

Overall FACETS identified significantly more regions of CNAs than were detected by analysis of the shallow pass sequencing using Nexus (Figure 5.1A). This included regions of subclonal CN change, CN neutral LOH and regions of smaller CNAs that are below the limit of detection with shallow-pass WGS. Comparisons of the size of the CNAs identified by Nexus and FACETS showed that shallow WGS rarely identifies CNAs less than 500 kb in size whilst FACETS can detect CNAs as small as 10 kb (Figure 5.1:B-C).

In order to understand the types of CNAs that were called by FACETS but not Nexus, the CNAs identified uniquely by FACETS were investigated. These unique CN calls were subdivided into clonal and subclonal CNAs where clonal CNAs had an estimated cellular fraction greater than 0.75 whilst subclonal CNAs had a cellular fraction less than 0.75. Clonal regions of CNA identified uniquely by FACETS tended to be smaller in size, often below the lower limit of detection of the shallow-pass WGS (Figure 5.1D). Subclonal regions of CNA showed greater variance of size distribution with both large and small subclonal events detected (Figure 5.1E). Some larger CNAs were uniquely identified by FACETS that were clonal but many of these were CN neutral LOH events that shallow-pass WGS cannot detect. Very few CNAs were identified by Nexus that were not also detected by FACETS, indeed only 19 of the 247 CNAs identified by Nexus were not also identified by FACETS.

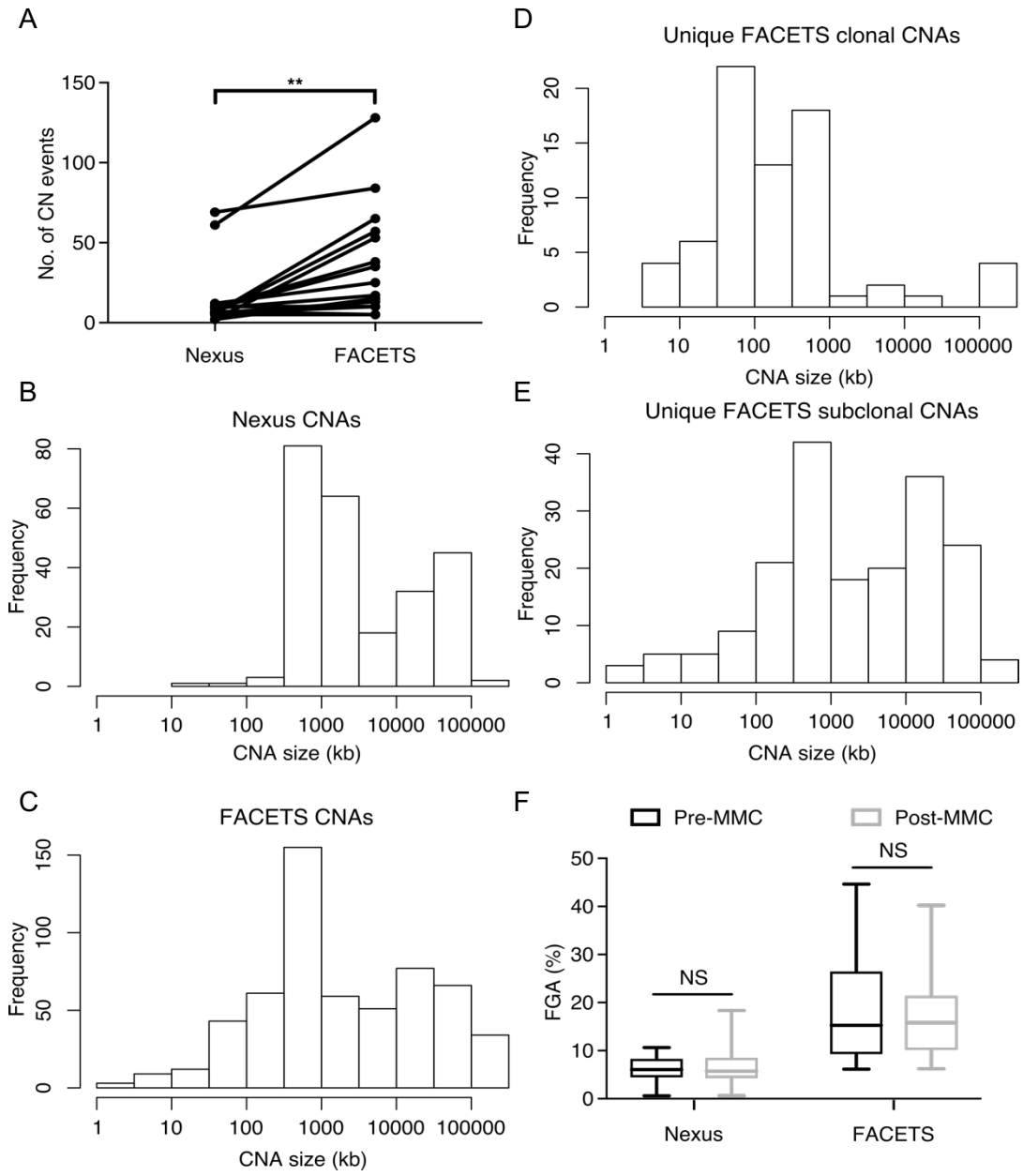


Figure 5.1: Comparison of copy number analysis carried out on shallow-pass WGS and WES data.

Shallow-pass WGS CNAs were called using the Nexus software whilst CNAs from WES data were called by FACETS. A) The number of CN events detected per sample was significantly higher using WES and FACETS (paired t-test, $p = 0.0023$). B-E) Size distribution histograms of CNAs. The x-axis is a $\log_{10}$ scale of CNA size and the y-axis is the frequency (no. of events). B) Size distribution of CNAs called by Nexus in the shallow-pass WGS data. Shallow-pass WGS identifies larger CNAs. C) Size distribution of CNAs identified by FACETS in the exome data. FACETS identifies CNAs of all sizes. D) Size distribution of clonal CNAs uniquely identified by FACETS. CNAs with a cellular fraction estimate greater than 0.75 were designated clonal. The size distribution is skewed towards smaller CNAs. E) Size distribution of subclonal CNAs uniquely identified by FACETS. CNAs with a cellular fraction estimate less than 0.75 were designated subclonal. FACETS identifies subclonal CNAs of all sizes. F) The Fraction of Genome Altered (FGA) was calculated for each method per sample. There was no significant difference in FGA seen between pre-MMC tumours and post-MMC tumours in either dataset (2-way ANOVA with Sidak's multiple comparisons test).

CN analysis of the shallow-pass WGS data had previously shown that there was no significant difference in the fraction of genome altered (FGA) between pre- and post-MMC tumours (see Chapter 3, section 3.2.4.3). To investigate whether this was also the case with the allele-specific CNAs, the FGA was calculated for each sample using the FACETS CN data. FGA was calculated by dividing the sum of size of the CNAs per sample by the size of the genome and multiplying by 100 to get a percentage. No significant difference was identified between pre- and post-MMC tumours for either CN method (2-way ANOVA with Sidak's multiple comparisons test) (Figure 5.1F). However, a much higher FGA was identified in patients P1175 and P2161 when it was calculated using the WES data with FACETS as compared to the shallow-pass data analysed by Nexus. In the pre-MMC tumours, there was a 22% increase in FGA detected in P1175 and a 44% increase in FGA detected in P2616. In the post-MMC tumours these increases were 32% and 21% respectively. Analysis of these additional events showed them to be largely subclonal and therefore not identifiable by shallow-pass CN analysis. For the other tumours, the difference was not as pronounced with an average increase in FGA of 5.3% as detected by FACETS. This data correlates with the increase in CNAs identified by FACETS.

Overall, FACETS appears to identify the vast majority of CNAs identified by the analysis of shallow-pass WGS data using Nexus and its use with whole-exome sequencing data also enabled the identification of smaller regions of CNA as well as CN neutral LOH and subclonal events. The use of this CN data was therefore considered suitable for subclonal deconvolution .

5.2.2 Identification of subclonal populations using kernel density plots

In CN neutral regions of the genome the VAF can be used to provide an estimate of clonal dynamics within a tumour. If clusters of mutations are identified with different VAF within these CN neutral regions then this is evidence of multiple clones present within the tumour. To identify the presence of subclones, CN neutral SNVs were extracted for each tumour and a kernel density plot of the resultant VAFs was generated within R using the ggplot2 package. Clonal heterozygous mutations have a VAF greater than 0.25 usually with a peak near 0.5 whilst subclonal mutations have a VAF of less than 0.25. All tumours in this cohort were estimated to have a purity of at least 70% tumour cells and should therefore demonstrate this distribution.

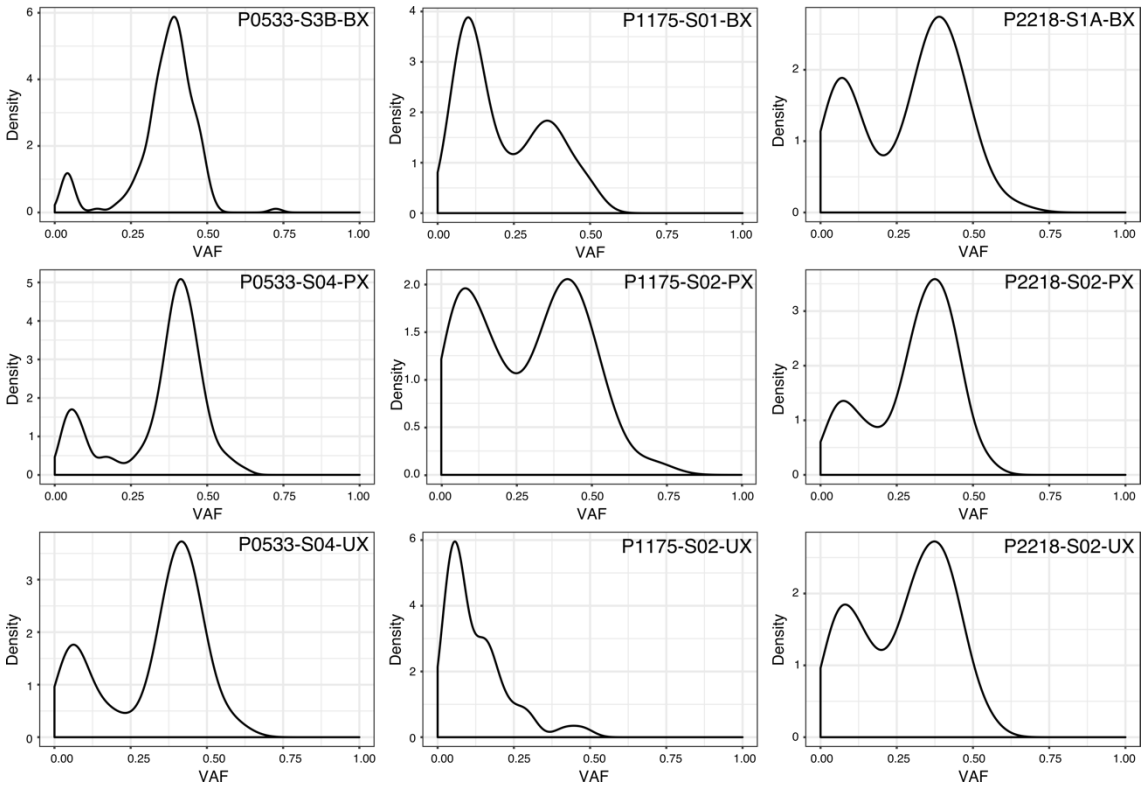


Figure 5.2: Assessment of intratumour heterogeneity using kernel density plots of tumour variant allele frequencies.

Variants from CN neutral regions were extracted for each tumour and kernel density plots were generated. The VAF (x-axis) is plotted against the density of the VAF (y-axis) generating a curve, the area under which is equal to 1. A single peak around a VAF of 0.5 would suggest a lack of subclonal populations whilst the presence of more than one peak is indicative of populations with different genotypes and therefore intratumour heterogeneity. The top row shows the pre-MMC tumour plots, the middle row shows the post-MMC tumour plots and the bottom row shows the post-MMC unique variants for each tumour.

An example of kernel density plots for tumours from 3 patients (P0533, P1175 and P2218) can be seen in Figure 5.2. Plots for all other patients can be found in Appendix J. All tumours showed at least two separate peaks suggestive of at least two distinct clonal populations. Patient P1175 shows an interesting pattern with a high density of subclonal mutations identified in the pre-MMC tumour. This was the only patient to display more subclonal mutations pre-MMC than clonal mutations. The post-MMC tumour had almost equal densities of clonal and subclonal mutations but when focusing on the post-MMC unique variants, almost all of these were subclonal. Overall, patient P1175 had a reduction in mutational load post-MMC. Taken together the data suggest that it was the subclonal mutations from the pre-MMC tumour that were lost and these were replaced by a smaller set of subclonal mutations post-MMC. This high level of subclonal mutations in this patient is reflective of the high level of subclonal copy number changes identified in this patient. The high number of subclonal mutations may also be indicative of neutral evolution within this patient.

Patients P0533 and P2218 both had more clonal mutations than subclonal. For these patients a mix of clonal and subclonal mutations were observed in the post-MMC unique variants (Figure 5.2). Only patients P1175 (Figure 5.2) and P0418 (see Appendix J) had very low levels of clonal mutations generated post-MMC treatment whilst the remaining patients had a mix of clonal and subclonal mutations. Some patients showed more than 2 peaks suggestive of multiple subclone clusters.

5.2.3 Subclone analysis and tumour evolution

To understand patterns of clonal selection related to MMC treatment, subpopulations of potentially functional variants were inferred from the paired tumours using PyClone-0.13.0[246]. PyClone can account for changes in CN state, including LOH, and uses this information to assign cellular prevalence estimates to individual mutations. Mutations with a similar cellular prevalence are then grouped together as a clonal cluster in which each clone is defined as a set of mutations. The package ClonEvol[247] was then utilized for clonal ordering and tree reconstruction.

An ancestral clone (set of mutations) was identified that was present in all cells of all tumours for each patient. Each tumour also contained additional subclones with a median of 2 additional subclones per tumour (range 1-4). The ancestral clones carried mutations in several genes that have been identified as cancer driver genes[47]. Potential driver events in ancestral clones included mutations in *FGFR3, PIK3CA, KDM6A, CDKN1B, KMT2D, ARID1A* and *KRAS*. Some tumours demonstrated a very simple

subclone structure such as patient P1175. Kernel density analysis had shown that both tumours from this patient had many subclonal mutations as well as some clonal mutations. Plotting the raw VAF of the variants as a scatterplot indicated that the variants present in both tumours ranged in VAF from approximately 0.25 to 0.9 whilst variants unique to either tumour had VAFs of less than 0.4 with clusters around 0.125 (Figure 5.3A). This correlates well with the CN neutral variants shown in the kernel density plots (Figure 5.2). The scatterplot shows three variants that were likely to be homozygous in both tumours; *FGFR3* p.G372C, *HTT* p.Q659E and *MICA* p.I116S. These were all in regions of CNA and therefore were omitted from the kernel density analysis.

For patient P1175, PyClone analysis grouped the variants into 3 clusters (Figure 5.3B). The mutations shared between the two tumours were clonal in both tumours suggesting that these mutations likely represent the ancestral founding clone. Each tumour also contained a single subclone derived from the ancestral clone (Figure 3C) and these were different between the pre- and post-MMC tumours. A representative diagram of each tumour illustrates the proportion of the tumour that is made up by each subclone (Figure 3C).

Overall the data suggests that for patient P1175, cluster 1 represents the ancestral trunk of the tree from which the two tumours branched off with different subclonal events (Figure 5.3D). This implies that the selective pressures of resection and/or MMC treatment may have eradicated cluster 2, whilst additional mutations that arose between resections generated cluster 3. This can be visualized in Figure 5.3E, where cluster 2 is not present after resection whilst cluster 1 retains its high prevalence. Cluster 3 is a new small subclone derived from cluster 1. Clusters were assessed for the presence of potential driver gene mutations using a list of 299 genes identified as drivers in various types of cancer[47]. Only mutations that had a predicted effect on protein composition (frame-shift, nonsense, splice-site) or missense mutations predicted as deleterious by SIFT and/or PolyPhen were included on the tree. Potential driver events were identified in both the ancestral branch and the pre-MMC subclone (clusters 1 and 2 respectively). The event affecting the potential driver gene *AJUBA* in cluster 3 was a tandem substitution hitting two separate codons; one variant was silent (p.P192P) and the other predicted to be tolerated (p.A193S). Despite the lack of a prediction for a detrimental effect, these mutations were included on the tree as they were likely a MMC specific event.

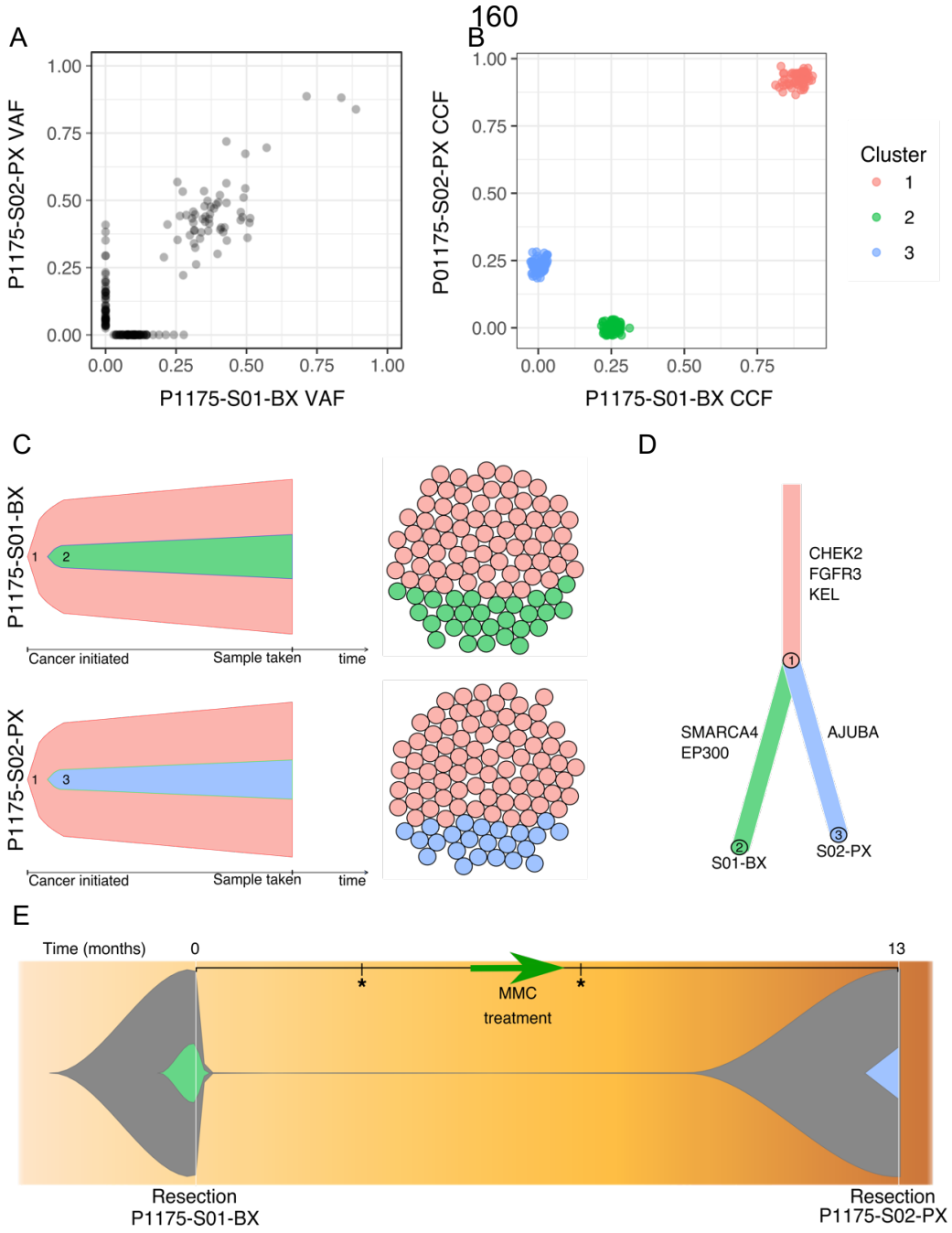


Figure 5.3: Clonal clustering and ordering for patient P1175.
A) Scatterplot of the variant allele frequency (VAF) of mutations found in the pre- and post-MMC tumours. Shared mutations can be identified in the middle of the graph with mutations unique to a tumour found at x or y =0. Variants were plotted at 30% opacity to show where mutations are clustering. VAF has not been adjusted for CN. B) PyClone was used to cluster mutations with a similar cancer cell frequency (CCF). Three clusters were identified; cluster 1 which is a clonal cluster common to both tumours, cluster 2 which is a subclonal cluster found only in tumour P1175-S01-BX and cluster 3 which is a subclonal cluster found only in tumour P1175-S02-PX. Points were plotted with a reduced opacity and jitter for better visualisation. C) ClonEvol was used for clonal ordering. Bell plots can be seen showing the development of each tumour with a representative depiction of the clonal makeup of each tumour. D) Clonal evolution tree depicting the evolution of the two tumours. Cluster ID can be found within the circles at each node. Branch colour reflects cluster composition. E) Fishplot bringing together the clonal models for the two tumours. * represents tumour events for which fresh-frozen material was not available. Clonal composition between the two time points is unknown and therefore represented by a line. Grey fill = cluster 1, green = cluster 2 and blue = cluster 3.

Patient P2161 also had a simple subclone structure. The raw VAF plot suggests that there was a clonal cluster shared between the two tumours then a subclonal population unique to the pre-MMC tumour and two subclonal populations unique to the post chemotherapy tumour (Figure 5.4A). One mutation appears to be potentially homozygous for both tumours and this was the *FGFR3* p.S249C hotspot mutation. SNaPshot analysis suggested this was a heterozygous mutation due to the presence of a wild-type allele. Reanalysis of the SNaPshot profiles indicates that this allele peak was very small and therefore consistent with the VAFs identified by WES. Clonal clustering performed by PyClone identified 5 different clusters, two of which were common to both tumours. Of these, cluster 1 was likely the ancestral clone as this remained at a stable high cellular prevalence in both tumours (Figure 5.4B). Cluster 2 was near-clonal in both tumours but increased in prevalence slightly after treatment. Cluster 3 was a subclonal cluster unique to tumour P2161-S01-BX, the pre-MMC tumour, and clusters 4 and 5 were subclonal clusters unique to tumour P2161-S02-PX, the post-MMC tumour. Clonal ordering suggested that both tumours evolved in a linear fashion (Figure 5.4C) and tree reconstruction indicated that divergence occurred from cluster 2 (Figure 5.4D). Cluster 5 was the first cluster to grow out after therapy, rising to be prevalent in about 75% of cancer cells. Interestingly no potential driver gene mutations were identified in this cluster but a mutation in *KMT2C* (p.L975P) was identified in cluster 4, and this was different to the pre-MMC *KMT2C* mutation (p.S1480*) identified in tumour P2161-S01-BX.

For some patents, multiple possible evolutionary trajectories were identified. Patients P0533, P1870, P2218 and P2329 each had two possible evolutionary trajectories. For patient P0533 an ancestral cluster, cluster 1, was identified from which tumour P0533-S3B-BX developed in a linear fashion with two subclones (Figure 5.5A). Both of these subclone clusters were eradicated post-MMC treatment. A new major clonal cluster, cluster 7, developed from the ancestral cluster post-MMC treatment and became present in almost all cells of tumour P0533-S04-PX. Clusters 6 and 8 were subclones of cluster 7 however it could not be determined if these developed linearly or if they were separate branches from cluster 7 (Figure 5.5A).

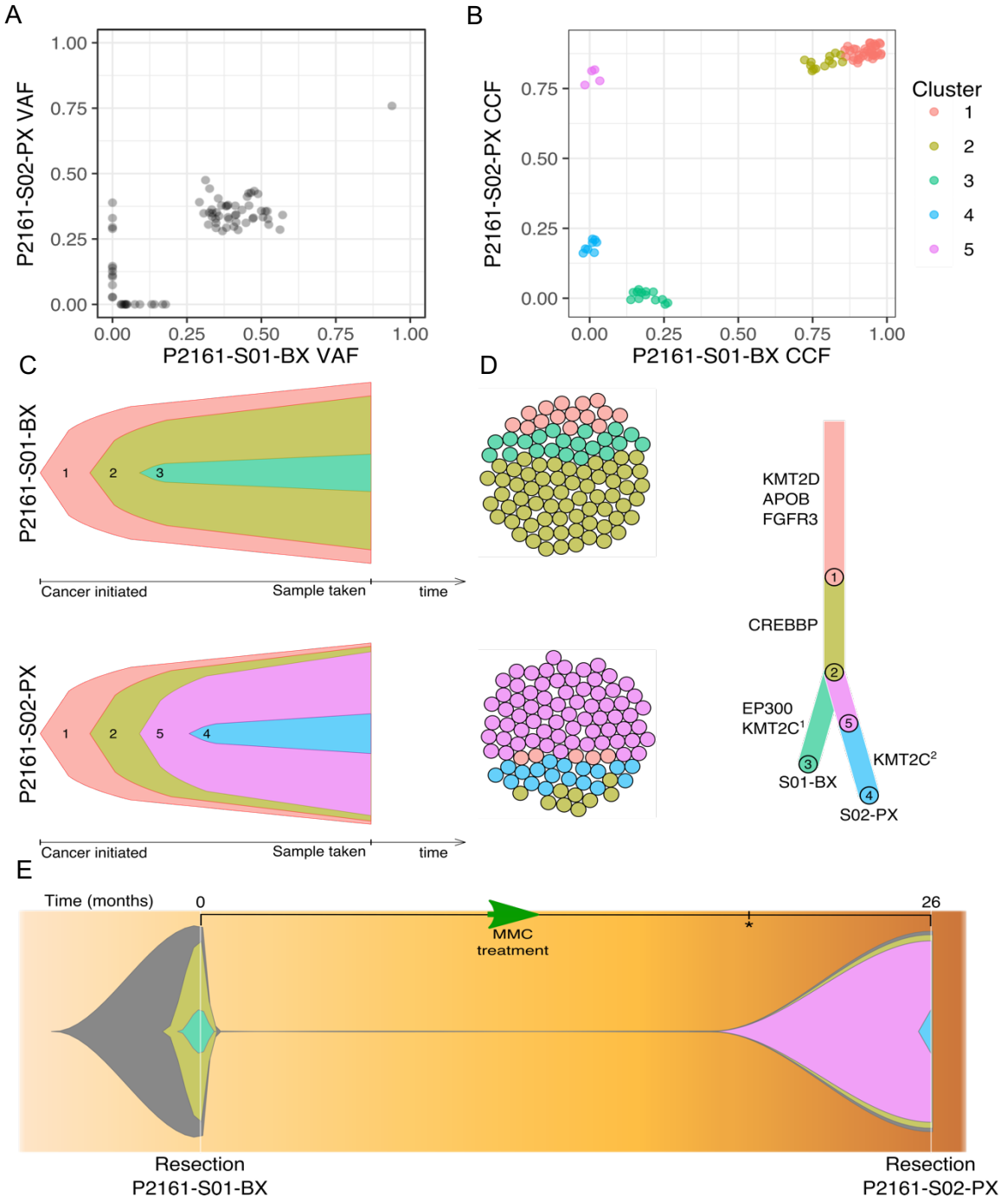


Figure 5.4: Clonal clustering and ordering for patient P2161.
A) Scatterplot of the variant allele frequency (VAF) of mutations found in the pre- and post-MMC tumours as per Figure 5.3. B) PyClone was used to cluster mutations with a similar cancer cell frequency (CCF). Three clusters were identified; cluster 1 which is a clonal cluster common to both tumours, cluster 2 which is a subclonal cluster in tumour S01 which rises to be almost clonal in tumour S02, cluster 3 which is a subclonal cluster unique to tumour S01 and clusters 4 and 5 which are subclonal clusters unique to tumour S02. Points were plotted with a reduced opacity and jitter for better visualisation. C) ClonEvol was used for clonal ordering. Bell plots can be seen showing the linear development of each tumour. D) Clonal evolution tree depicting the evolution of the two tumours. Cluster ID can be found within the circles at each node. Brach colour reflects cluster composition. [1]nonsense mutation p.S1480*, [2]missense mutation p.L975P. E) Fishplot bringing together the clonal models for the two tumours. * represents tumour events for which fresh-frozen material was not available. Clonal composition between the two time points is unknown and therefore represented by a line. The ancestral clone is grey whilst the subclones are coloured as per B.

Interestingly, both major subclones found within the two tumours from patient P0533 (cluster 4 and cluster 7) carried mutations in the *RB1* gene. Cluster 4 carried a splice site mutation affecting the splice donor site (p.X565_splice at c.1695+1G>A) whilst cluster 7 carried a nonsense mutation (p.S576*, c.1727C>G). Both of these mutations would lead to non-functional proteins. As all tumours from patient P0533 carried a deletion of the part of chromosome 13 that contains *RB1* this would have eradicated RB function in both tumours. The convergence of mutations on the *RB1* gene could suggest that loss of this pathway was critical for tumour outgrowth in this patient.

The CCF of two subclonal clusters from tumour P1870-S03-PX, clusters 4 and 5, made it impossible to determine if these subclones developed linearly or branched off separately from the ancestral cluster (Figure 5.5B). Interestingly neither of these clusters contained a potential driver gene mutation. For cluster 5, the lack of driver combined with the low CCF of this cluster could suggest that it developed neutrally. Cluster 4 was present in over 75% of the cells from the post-MMC tumour suggestive of an increased fitness despite the lack of an identifiable driver gene mutation. This could either suggest the presence of a driver gene that has yet to be identified, that a non-coding mutation drove growth of this cluster or that the ancestral mutations were sufficient for tumour recurrence.

Tumours from patient P2329 demonstrated simple clonal architectures (Figure 5.5C). An ancestral cluster, cluster 4, was identified in all cells. In tumour P2329-S02-BX an additional subclone containing just 2 coding variants was identified (cluster 3). In tumour P2329-S03-PX two additional subclones were identified; clusters 1 and 2. Again it was impossible to determine if the two subclones developed linearly, with cluster 2 developing as a subpopulation of cluster 1, or if they branched off the ancestral cluster (cluster 4) separately, resulting in two possible evolutionary trees (Figure 5.5C).

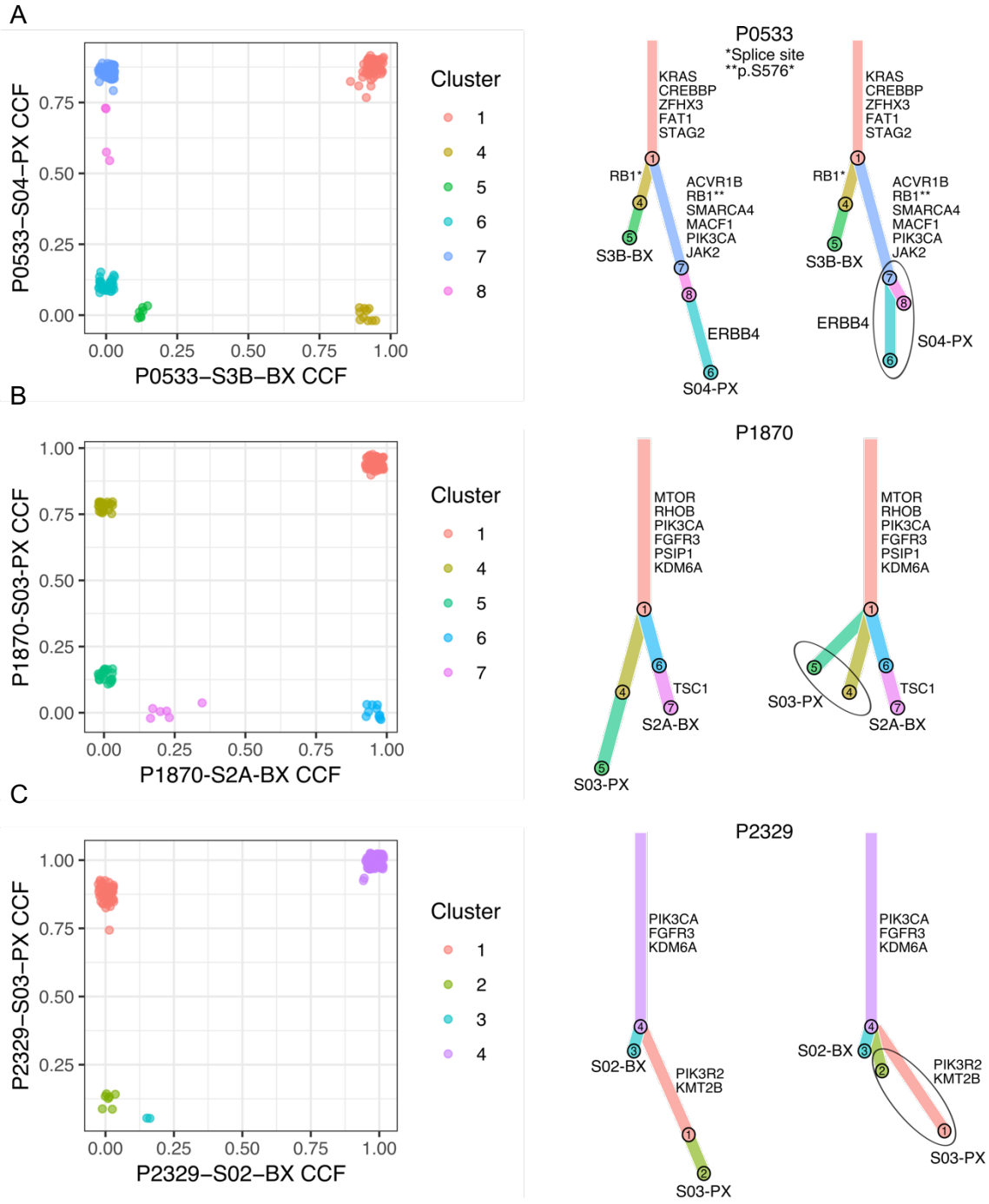


Figure 5.5: Clonal clustering and ordering for patients P0533, P1870 and P2329. PyClone was used to cluster mutations with a similar cancer cell frequency (CCF). These are plotted as a scatter graph with pre-MMC tumours on the x-axis and post-MMC tumours on the y-axis. Clusters compromising of a single mutation were removed from the analysis. Clusters unique to pre-MMC tumours can be found where y=0 and clusters unique to post-MMC tumours can be found at x=0. Clusters were plotted with added jitter and increased transparency to give distinction between clusters containing many variants and clusters containing few variants. ClonEvol was used to order the clusters. Two trees are possible for the post-MMC tumours for each patient and the branching or linear evolutionary paths that could have been followed are shown in each case. Ovals are drawn around branching subclonal clusters that are present in the same tumour. A) Results for patient P0533. B) Results for patient P1870. C) Results for patient P2329.

Patient P2218 had two multifocal tumours prior to MMC treatment and a single tumour post-MMC treatment. All three tumours contained cluster 8 which likely represented the ancestral clone. The pre-MMC tumours shared a small cluster that made a clonal sweep (cluster 7), from which the other tumour specific subclones evolved. Tumours P2218-S1B-BX and P2218-S02-PX followed a linear evolution pattern however the evolutionary pattern for tumour P2218-S1A-BX could not be determined resulting in two possible evolutionary trees (Figure 5.6). Parallel evolution was displayed with tumours P2218-S1A-BX and P2218-S02-PX showing independent mutations of the tumour suppressor gene *TSC1* (p.E174K fs*36 and p.Q55*) (Figure 5.6). Interestingly tumour P2218-S1B-BX contained an additional *FGFR3* mutation to the R248C hot-spot mutation identified in all 3 tumours. This was a frameshift insertion (p.V245R fs*36).

Two patients, P0418 and P0960, had complex genetic structures which made clonal ordering difficult. For patient P0418 the presence of a cluster that transitioned from subclonal to almost clonal confounded ordering of the clusters (Figure 5.7A). This cluster, cluster 1, contained 6 variants including the *PIK3CA* p.E545K hot-spot mutation. Investigations into the CN status of these 6 variants identified that they were in regions of subclonal CN loss that was not present in the post-MMC tumour. It was hypothesised that the subclonal nature of these CN events confounded the calculation of CCF and that these variants were actually part of the ancestral cluster. Removing the subclonal CNAs by restoring them to a CN neutral state resulted in these mutations being clustered together with the ancestral cluster and reduced the number of clusters identified (Figure 5.7B). Using this data for clonal ordering with ClonEvol generated 4 possible evolutionary trees. These all differ with respect to cluster 6 which has a very low CCF that makes it impossible to determine where this cluster branches off the tree (Figure 5.7C).

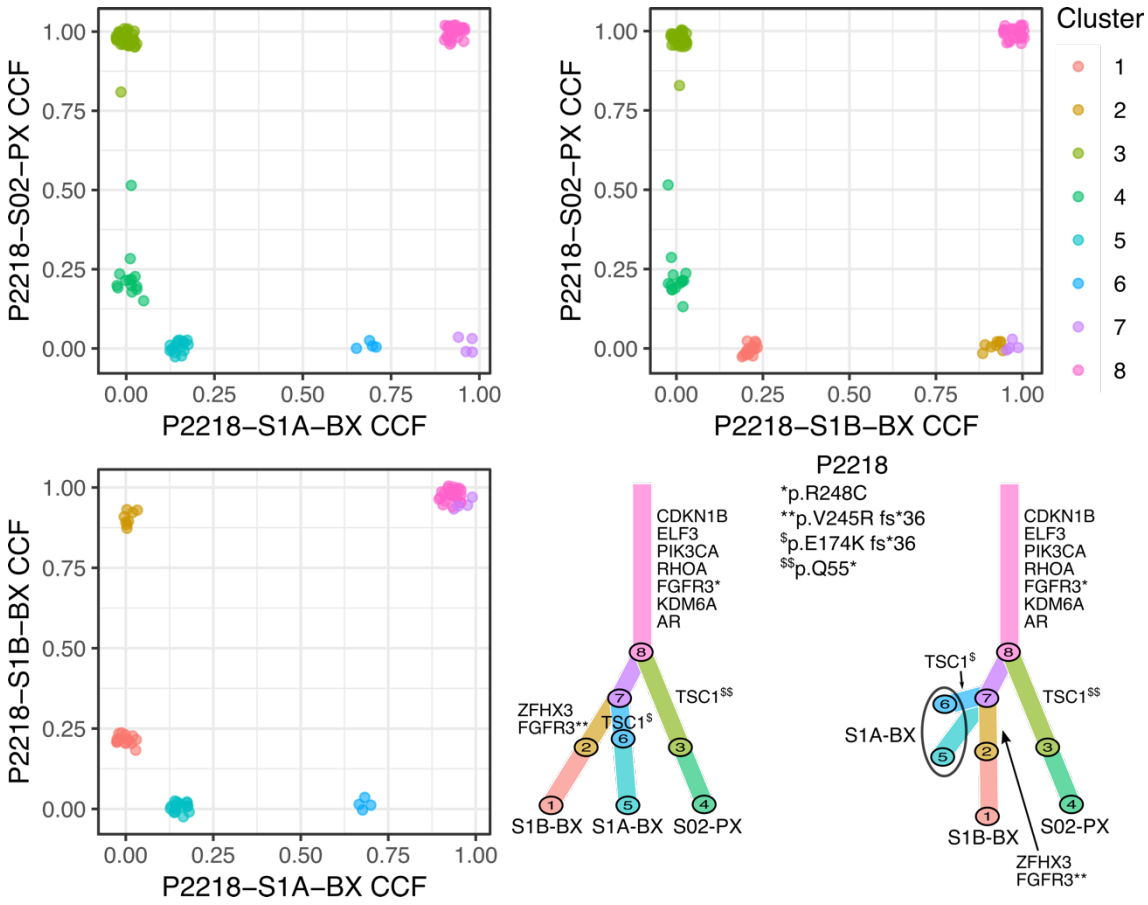


Figure 5.6: Clonal clustering and ordering for patient P2218.
PyClone was used to cluster mutations and scatterplots were generated as per Figure 5.3. As patient P2218 has multiple tumours all possible combinations have been plotted. ClonEvol was used to order the clusters. Two trees were possible for P2218 as branching or linear evolutionary paths for tumour P2218-S1A-BX could have been followed. Ovals are drawn around branching subclonal clusters that are present in the same tumour.

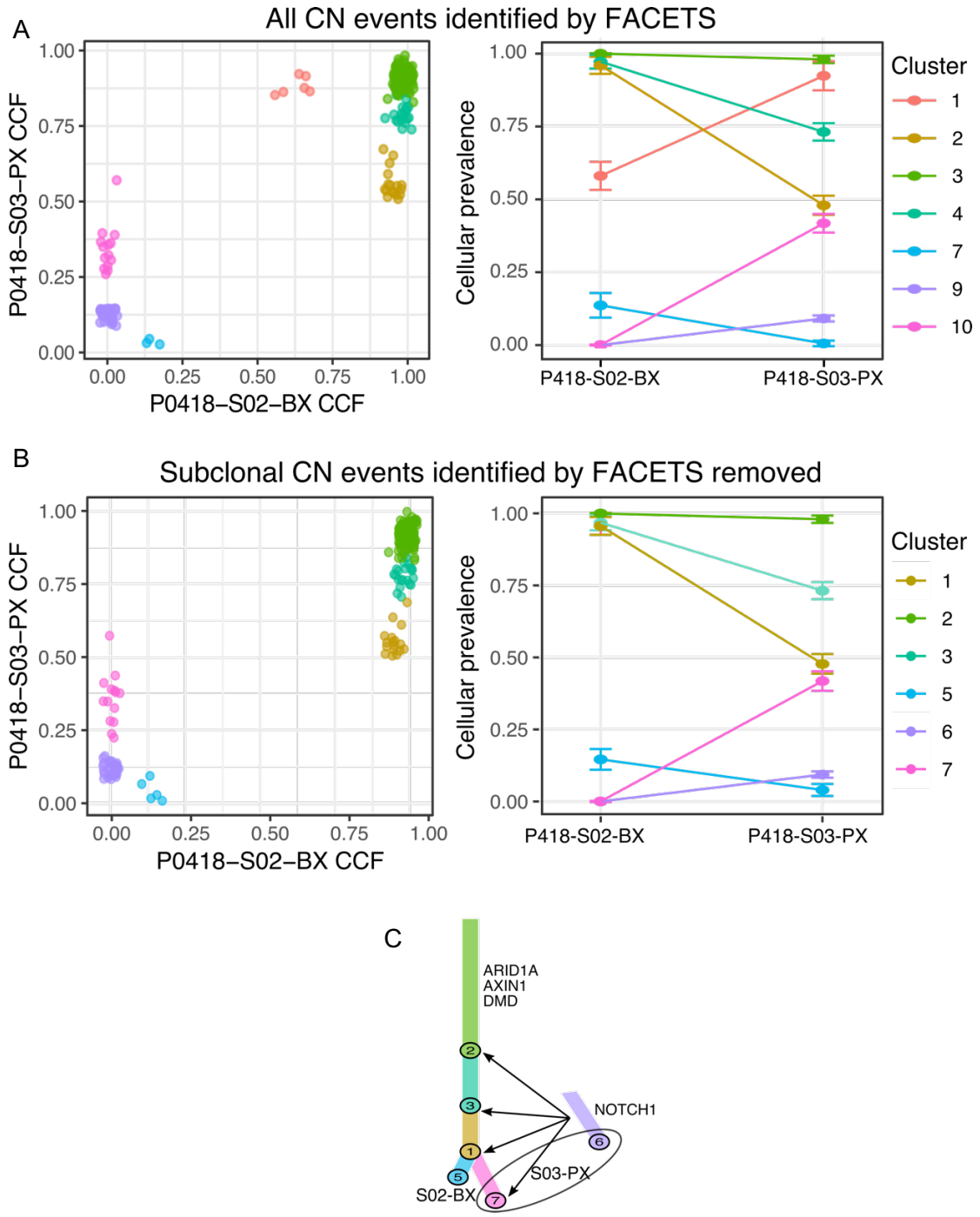


Figure 5.7: Clonal clustering for patients P0418.
PyClone was used to cluster mutations and scatterplots were generated as per Figure 5.3. Parallel coordinate graphs have also been generated for each patient showing the mean cellular prevalence for each cluster with the standard deviation. A) Clonal clustering for patient P0418 with all CN events identified by FACETS. Cluster 1 does not appear to make evolutionary sense. These variants are all in regions of subclonal CN identified by FACETS B) Clonal clustering for patient P0418 with subclonal CN events identified by FACETS returned to a normal CN status. Cluster 1 from A now clusters with cluster 2, the ancestral cluster. C) Tumour evolution tree for patient P0418 using clustering from B for clonal ordering with ClonEvol. Cluster 6 has a very small prevalence and therefore the branching point of this cluster cannot be identified. All possible positions for cluster 6 have been indicated with an arrow. Ovals have been drawn around branching subclonal clusters that are present in the same tumour..

Clonal ordering was difficult for patient P0960 as cluster 3 did not follow a cellular prevalence pattern that fitted with the evolution of the other clusters (Figure 5.8). Cluster 7 is a new cluster that becomes clonal post-MMC meaning it cannot be a subclone of cluster 3. However, this means that cluster 3 should not theoretically be present at the high clonal prevalence it is found to be in the post-MMC tumours. Clonal ordering of the post-MMC tumours with ClonEvol places this cluster as a subclone of cluster 7, however this is unlikely as cluster 3 was present in the pre-MMC tumour. Cluster 3 contained 11 coding variants making it an important cluster that cannot be overlooked but there were no mutations identified in potential driver genes. Removing any subclonal CNAs from the analysis did not change the cluster dynamics observed (data not shown). Cluster 2 was likely the ancestral cluster and this contained several mutations in potential driver genes including missense mutations predicted to be deleterious in *NCOR1* (p.K835I)*, MGA* (p.E1337K) and *TSC1* (p.M142V) as well as a hot-spot mutation in *FGFR3* (p.Y375C)*.* No other clusters contained mutations predicted to affect protein function in any driver genes, however a tandem substitution in *KMT2C* (p.G2213L) was identified in cluster 9. This had a MMC signature mutation of CC > AA however this was predicted to be benign using PolyPhen-2.

5.2.4 Mitomycin C associated mutations are predominantly subclonal

In Chapter 4, a number of mutation characteristics were associated with a course of mitomycin C treatment. These included: an increase in the number of C > A substitutions, a decrease in the transition/transversion (Ts/Tv) ratio and an increase in the number of tandem substitutions, specifically at CC or GG dinucleotides. Kernel density analysis identified the presence of both clonal and subclonal mutations in the unique post-MMC variants (see Figure 5.2 and Appendix J). Clustering of the coding mutations with PyClone identified clonal post-MMC unique clusters in 6 patients and subclonal post-MMC unique clusters in all patients. We were interested to identify if the MMC-associated variants were clonal, subclonal or a mix of the two.

The bcftools “stats” command was used to investigate the base-substitution pattern of post-MMC unique clonal and subclonal variants. Clonal clusters contained C > T or C > G substitutions whilst the subclonal clusters contained predominantly C > A substitutions (Figure 5.9A and B). The clonal variants had a Ts/Tv ratio of 0.92 whilst the subclonal variants had a Ts/Tv ratio of 0.25. These data suggest that the mutations associated with mitomycin C treatment tended to be subclonal.

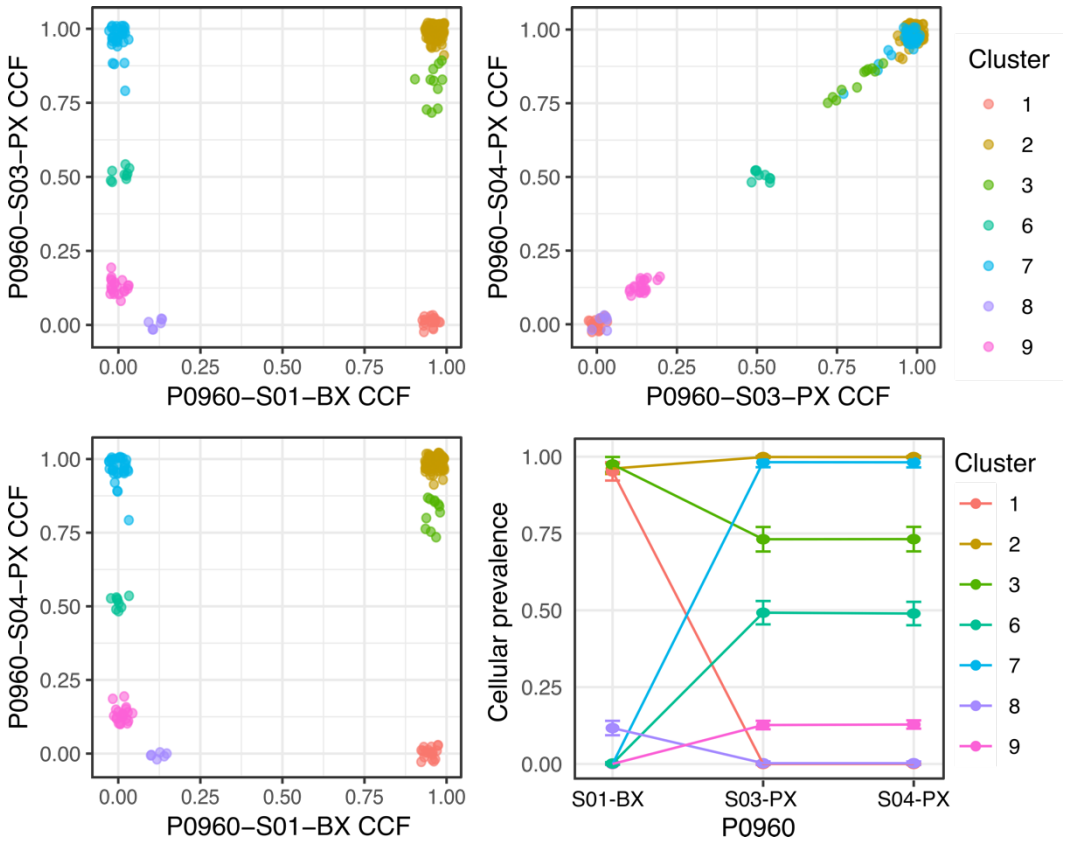


Figure 5.8: Clonal clustering for patient P0960.
Clonal clustering for patient P0960. Clusters with 3 mutations or fewer were removed. Scatterplots comparing the CCF for each post-MMC tumour against the pre-MMC tumour, as well as comparing both post-MMC tumours were generated. The results have been displayed on a line-chart to visualise the changes in CCF between tumours. Clonal ordering was attempted by ClonEvol however no consensus trees could be identified due to cluster 3.

A large increase in the number of tandem substitutions was identified post-MMC treatment. These are highly likely to be MMC-induced events and can therefore be used to represent such events. Of the 62 post-MMC unique tandem substitutions, 36 were in coding regions. All coding, post-MMC unique tandem substitutions from CN neutral regions were plotted in a kernel density plot of VAF against density (Figure 5.9C). This shows that the majority of these tandem substitutions were subclonal. Analysis of the clonal and subclonal clusters for each patent identified only 2 tandem substitutions within a clonal cluster and these were both from patient P0533. Overall the data suggests that variants induced by MMC treatment were mostly subclonal.

5.2.5 Targeted next generation sequencing

Targeted next generation sequencing was used to supplement the WES data described in Chapter 4. A panel of 140 genes (Appendix D) was sequenced in additional tumours from patients who previously had selected tumours included in the WES cohort. Targeted sequencing was also used to discern clonality of tumours from patients for whom tumour clonality could not be determined by CN and single gene mutation analysis alone.

5.2.5.1 Targeted sequencing of additional tumours from patients in the WES cohort

Four patients from the WES cohort had material available from tumours additional to those immediately surrounding treatment. These were patients P0418, P0533, P1870 and P2329. It was considered that analysis of these tumours could provide additional information about shared and unique mutations identified within the exome cohort. These additional tumours underwent targeted sequencing using the bladder cancer gene panel described above. Summary data of the variants identified can be seen in Table 5.1.

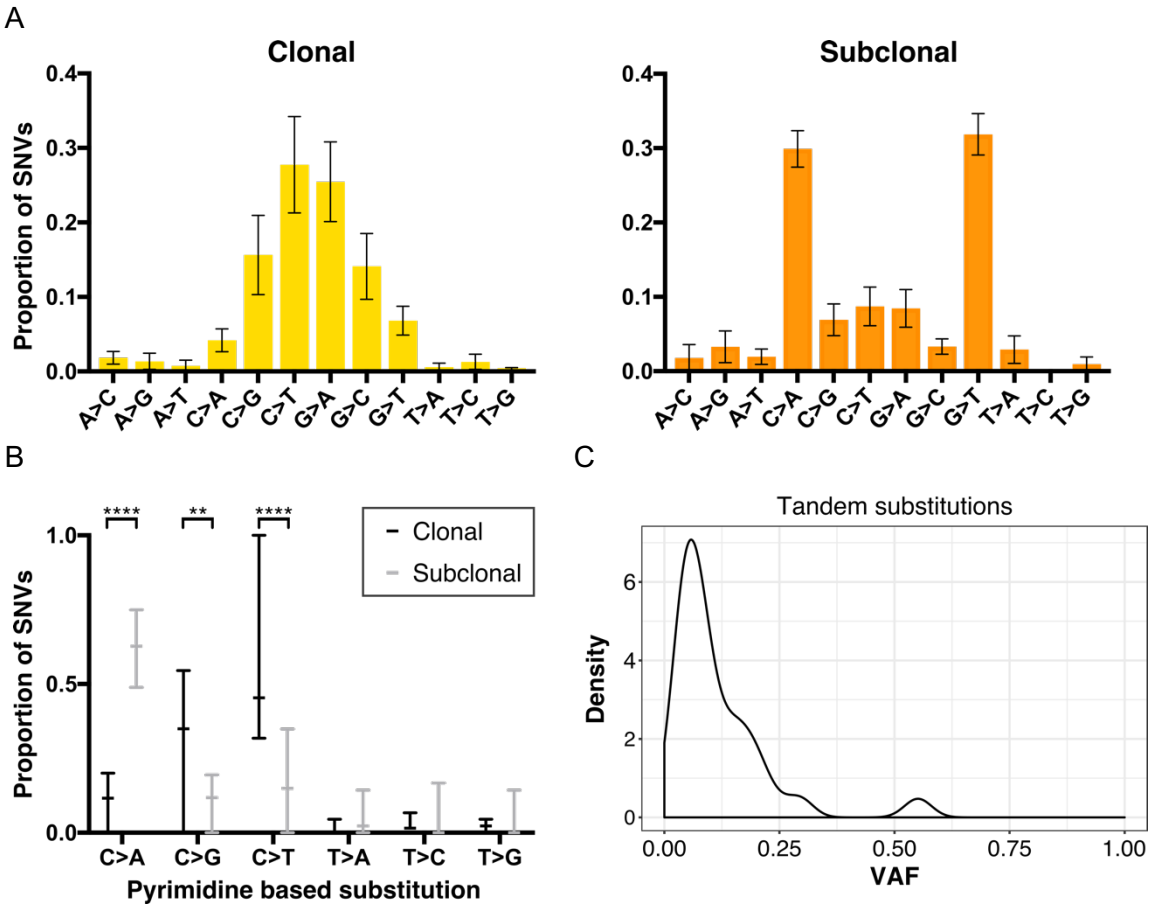


Figure 5.9: Mutation spectrum of clonal and subclonal post-MMC unique variants.

Summary of mutation spectrum data for clonal and subclonal variants unique to the post-MMC tumours. A) Base substitutions were identified using the bcftools package with the "stats" command. Histograms of the substitution spectrum for clonal and subclonal SNVs can be seen. Substitutions are depicted as a mean proportion ± SEM. B) Substitutions categorised by the pyrimidine of the mutated base pair. Lines represent the median and range. The y-axis depicts the proportion of mutations with that event. Comparisons between clonal and subclonal variants show a significant difference in the number of C > A, C > G and C > T substitutions (two-way ANOVA with Sidak's correction for multiple comparisons: **** p = <0.0001, ** p = 0.0090). C) Kernel density plot of coding tandem substitutions from the post-MMC unique variants. The VAF (x-axis) is plotted against the density of the VAF (y-axis). The majority of post-MMC unique tandem substitutions tend to be subclonal.

Patient P0533 had 3 additional tumours available for targeted sequencing all of which were resected prior to MMC treatment: P0533-S01 resected almost 12 years prior to MMC treatment, P0533-S02 resected 8 months prior to MMC treatment and P0533-S3A resected 2 months prior to MMC treatment, at the same time as tumour P0533-S3B-BX. Analysis of the three tumours identified 8 variants that were shared between all 3 tumours: *BRCA2* (p.E3152Q), *CREBBP* (p.Y1539*), *FAT1*(p.Q3929*), *KRAS* (p.G12V), *KRAS* (p.A146T), *NAT10* (p.D90N), *STAG2* (p.Q573*) and *ZFHX3* (p.P2282L). These variants were also shared with both tumours from the exome sequencing data and were part of the ancestral clonal cluster. These variants included all the key driver variants depicted on the branch of the tree in Figure 5.5A which were nonsense mutations in *CREBBP* (p.Y1539*), *FAT1* (p.Q3929*) and *STAG2* (p.Q573*), and missense mutations in *KRAS* (p.G12V and p.A146T) and *ZFHX3* (p.P2282L).

All tumours from patient P0533 analysed by targeted capture also contained many unique non-synonymous variants (range 5-21). Tumour P0533-S02 was an interesting tumour as it shared a missense variant in *KLF5* (p.D418N) with P0533-S3A and a nonsense mutation in *RB1* (p.S576*) with tumour P0533-S04-PX, neither of which were identified in any other tumours. P0533-S02 also contained a unique mutation in *PIK3CA* (p.E452K). This suggests that tumour P0533-S02 contained at least 3 subclones, each of which went on to found separate tumour events.

Table 5.1: Summary of targeted sequencing results.
Variants were included if they were identified by at least two of the three variant callers. The number of unique variants are variants identified for that tumour that were not shared by all the other tumours from that patient. Potential driver mutations for these additional variants have been highlighted. [a]Variants in *ARID1A* are different for each tumour; P0533-S01 contains a missense mutation (p.I2282T) predicted to be deleterious and P0533-S3A contains a nonsense mutation (p.Q944*). [b]*UTY* has not been identified as a driver gene but it is the Y-chromosome homolog of *KDM6A*. [c]Nonsense mutation shared with tumour P0533-S04-PX. [d]Shared missense variant (p.D418N). [e]*KDM6A* (p.V558E fs*36) mutation is shared with tumours P1870-S2A-BX and P1870-S03-PX.

Patient	Tumour	No. variants identified	No. of unique variants	Additional driver events
P0418	P0418-S01	11	4	*PIK3CA*
P0533	P0533-S01	34	26	*ARID1A*[a], *ATM, KMT2C, UTY*[b]
	P0533-S02	18	10	*RB1*[c], *KLF5*[d], *NFE2l2, PIK3CA*
	P0533-S3A	27	18	*ARID1A*[a], *KMT2D, KLF5*[d], *MECOM*
P1870	P1870-S01	7	0	
	P1870-S05	14	7	*CDKN1A, KDM6A*[e]
P2329	P2329-S01	16	3	*FBXW7*

Patient P1870 had an additional 2 tumours available for targeted sequencing; tumour P1870-S01 resected 44 months prior to MMC treatment and tumour P1870-S05 resected 30 months after MMC treatment (25 months after tumour P1870-S03-PX). Analysis of the tumours identified 5 variants common to all four tumours. This included the hot-spot mutation in *FGFR3* (pS249C), a missense *PIK3CA* mutation (p.R88Q) as well as the missense mutation in *RHOB* (p.N94Y) predicted to be an ancestral driver event. Interestingly the mutation in *KDM6A* (p.V558Efs*36), predicted to be an ancestral event by exome sequencing, was not present in tumour P1870-S01. This suggests that tumour P1870-S01 branched off prior to this event. Neither of the other ancestral driver genes, *MTOR* or *PSIP1,* were included in the targeted gene panel. Two additional events were identified as being present in only tumours P1870-S01 and P1870-S05. These were mutations hitting the intronic regions of *NOTCH2* and *HMCN1*. However, investigations into these variants in the Interactive Genome Viewer (IGV) showed the presence of these mutations in all four tumours, yet the variants were not called in the exome samples. For the intronic *NOTCH2* mutation this was likely due to its presence in a region with low mapping quality whist for *HMCN1* this was likely due to strand bias present in the exome sequencing data.

Patient P0418 had a single additional tumour which underwent targeted sequencing; tumour P0418-S01 resected 4 months prior to MMC treatment. This identified 11 variants, of which 7 were shared with all other tumours including the frame-shift insertion mutation in *ARID1A* (p.Q802V fs*32). Of the four unique variants identified in tumour P0418-S01 only one affected a coding region. This was an additional *PIK3CA* mutation (p.E726K).

Patient P2329 also only had a single additional tumour; P02329-S01 resected 13 months prior to MMC treatment. Targeted exome sequencing identified 16 variants in this tumour. Of these, 13 were identified as shared with both the exome sequenced tumours and included the three ancestral driver mutations in *PIK3CA* (p.E545K), *FGFR3* (p.S249C) and *KDM6A* (p.S1114Ifs*37). Tumour P2329-S01 had 3 additional variants that were unique to that tumour only, one of which was a missense mutation in *FBXW7* (p.R479G), a potential driver gene.

5.2.5.2 Determination of clonality based on next generation sequencing data

In Chapter 3, recurrent tumours from 23 patients were assessed for copy number alterations using shallow-pass WGS. This data was supplemented with mutation

analysis of hot-spot regions in 6 key genes; *FGFR3, PIK3CA, HRAS, KRAS, NRAS* and the *TERT* promoter region. This identified a clear monoclonal origin for 15 of the patients. However, for 6 patients (P0468, P0536, P0933, P0990, P2161 and P2218) despite a suggestion of monoclonality, this could not be confirmed due to the shared events being very common in bladder cancer and therefore possibly independent. A further 2 patients (P2065 and P2291) showed no evidence for a monoclonal origin. In order to confirm or refute a monoclonal origin of tumours from these patients, tumours were further analysed by targeted sequencing.

Tumours from patients P2161 and P2218 underwent WES as part of the MMC treatment cohort. This identified numerous shared variants between tumours (Chapter 4, Figure 4.4) and mutation clustering and clonal ordering identified the presence of an ancestral cluster common to all tumours for each patient (Figure 5.4 and Figure 5.6), consistent with a monoclonal origin. Tumours from patients P0468, P0533, P0536, P0933, P0990 and P2291 underwent targeted sequencing using the bladder cancer targeted gene panel. Unfortunately, a lack of available DNA excluded patient P2065 from analysis.

Mutations shared by all tumours from an individual patient were found in patients P0468, P0536, P0933 and P0990 (Table 5.2) suggestive of a monoclonal origin for tumours from these patients. Two mutations were shared by all 3 tumours from patient P0468 including a frameshift insertion in *CDKN1A* (p.L37L fs) and an inframe insertion in *EGFR* (p.P772PH). Tumours P0468-S03 and P0468-S05 shared an additional 2 mutations both in *FAT1*: one frameshift insertion (p.F330F fs) and one missense (p.P329S). The missense mutation is likely to be a mapping error as it occurs close to the frameshift mutation and is always present on the same reads. Tumour P0468-S01 had 8 unique mutations, including three nonsense mutations in FAT1 (p.Q725*, p.S1908* and p.Q2304*) and a nonsense mutation in *CDKN1A* (p.R48*). All 3 *FAT1* mutations were different to those seen in tumours S03 and S05. P0468-S03 contained a unique missense mutation in *AHNAK2* (p.V1789L) predicted to be benign whilst P0468-S05 had 10 unique mutations including two tandem substitutions within *SPTAN1,* both of which hit separate codons (p.L9L & p.E10* and p.L1036L & p.E1037*), and a frameshift insertion in *FOXA1* (p.T46 fs).

Table 5.2: Targeted sequencing analysis for the determination of clonal origins.
The number of variants shared by all tumours was calculated and mutations in potential driver genes have been highlighted. Unique variants for each tumour (those not shared by all tumours) were identified. [a]*FAT1* mutations were identified in all 3 tumours. Tumours P0468-S03 and -S05 shared mutations but the mutations seen in P0468-S01 are unique. [b]Tumours P0536-S01 and -S03 share mutations in *ATM* and *KDM6A* whilst the mutations in these genes in tumour P0536-S02 are unique. [c]Tumours P0990-S04 and P0990-S05 shared an identical mutation spectrum.

Patient ID	Total shared variants	Shared Driver Genes	Tumour ID	No. variants identified	No. of unique variants	Additional driver events
P0468	3	*CDKN1A, EGFR, TERT promoter*	P0468-S01	11	8	*FAT1*[a], *CDKN1A*
			P0468-S03	6	3	*FAT1*[a]
			P0468-S05	13	10	*FOXA1, FAT1*[a], *SPTAN1*
P0536	2	*FGFR3, PIK3CA*	P0536-S01	10	8	*ATM*[b], *ZFHX3, KDM6A*[b], *STAG2*
			P0536-S02	6	4	*ATM*[b], *KDM6A*[b]
			P0536-S03	9	6	*ATM*[b], *KDM6A*[b], *STAG2, -146 TERT*
P0933	9	*ARID1A, ELF3, ZFP36L1, EP300, RHOA, FGFR3, KDM6A, TERT promoter*	P0933-S01	12	3	*ZFHX3, STAG2*
			P0933-S02	18	9	*EP300, PIK3CA*
P0990	6	*FGFR3, KDM6A, NOTCH1, PIK3CA, TERT promoter*	P0990-S01	7	1	*STAG2*
			P0990-S04	17	11[c]	*ARID2, KMT2D, FOXQ1, TSC1*
			P0990-S05	17	11[c]	
P2291	0		P2291-S01	6	6	*ARID1A, KMT2D, CREBBP, FGFR3*
			P2291-S02	8	8	*ELF3, ZFP36L1, KDM6A*

Targeted sequencing identified only 2 variants that were shared between all three tumours for patient P0536 and these were the hot-spot mutations in *FGFR3* (p.S249C) and *PIK3CA* (p.E545K) identified previously by SNaPshot analysis. Tumours P0536-S01 and P0536-S03 additionally shared 4 variants not seen in tumour P0536-S02 including a nonsense mutation in *STAG2* (p.W743*), a missense mutation in *ATM* (p.E2164Q), a frameshift deletion in *KDM6A* (p.D934 fs) and a missense mutation in *HMCN1* (p.W3746C). Tumour P0536-S02 also contained missense variants in *ATM* (p.D1548H and p.G2765S) and a frameshift deletion in *KDM6A* (p.QR117-118 fs) but these were alternative events to those seen in tumours P0536-S01 and P0536-S03. Tumours P0536-S01 and P0536-S02 both contained 4 unique mutations and P0536-S03 contained 2 unique mutations. Overall tumours P0536-S01 and P0536-S03 had a clear monoclonal origin. However, the origin of tumour P0536-S02 remains unclear.

Hot-spot mutation analysis had identified that the two tumours from patient P0933 shared *FGFR3* S249C and -124 G>T *TERT* promoter mutations, but P0933-S02 contained an additional *PIK3CA* E542K mutation not seen in P0933-S01. CN analysis identified a large loss of a region of 9q in tumour P0933-S01. This CN loss was not present in P0933-S02 which contained no detectable CNAs. A monoclonal origin could not be determined from this data alone so the tumours underwent targeted sequencing. This identified an additional 7 variants that were shared including a splice-donor mutation in *ELF3* (c.805+1 G>A), a missense mutation in *RHOA* (p.D67H) and nonsense mutations in *ARID1A* (p.Q585*), *EP300* (p.Q2141*) and *KDM6A* (p.S1154*). This would suggest a monoclonal origin for these tumours. Both tumours also contained unique mutations with P0933-S01 containing 3 unique mutations (*STAG2* p.R216*, *PHF* p.P506P and *ZFHX3* p.Q1857*) and P0933-S02 containing 9 (including: *RARG* p.R116Q and a splice donor variant in *EP300* (c.3671+1 G>A) as well as a *PIK3CA* hotspot mutation (p.E542K).

Analysis of the tumours from patient P2291 did not identify any variants shared between the two tumours (Table 5.2). Tumour P2291-S01 contained 7 variants: *ARID1A* (p.P1592 fs), *KMT2D* (p.L448 fs, p.C1103G and p.D3411N), *CREBBP* (p.R386*), *FGFR3* (p.S249C) and an intronic mutation in *SYNE2*. Tumour P2291-S02 contained 8 variants: *ELF3* (p.L342 fs, p.236-237 fs and p.254-255 fs), *ZFP36L1* (p.137-138 fs), *ACAN* (p.P59S), *RHOB* (p.E47K), *CACNA1D* (p.A66V) and *KDM6A* (p.1081-1082 fs). All sites were investigated in the Interactive Genome Viewer (IGV). One variant in P2291-S01 (*SYNE2*) was found to be false as all SNVs occurred at the end of reads only, with very few reads covering the variant. The two *ELF3* frame-shift mutations in close proximity (p.236-237 fs and p.254-255 fs) were never present on the

same read, suggestive of two separate events and inactivation of both alleles of the protein. All other variants were verified to be unique to their respective tumour but two variants did show a single read in the other tumour; *FGFR3* p.S249C*,* unique to S01, contained 1 read with the variant allele in tumour S02 and *CACNA1D* (p.A66V)*,* unique to S02, contained 1 read with the variant allele in S01. For *FGFR3* this was one read out of a total of 345 reads covering the base and for *CACNA1D* this was one read out of a total of 789. It is possible that these were detected in very minor subclones within the two tumours, however the presence of only a single read despite high coverage may also suggest that cross contamination or error could be the likely cause. Overall, the evidence would suggest that the tumours from patient P2291 were oligoclonal in origin.

5.3 Discussion

Detailed knowledge of the clonal composition and mutational heterogeneity of NMIBC and the changes associated with the use of therapy would aid in the design of new therapies or therapeutic regimes. By gathering a detailed understanding of the molecular alterations present within tumours and identifying variants that frequently occur on the trunk of the evolutionary tree, new therapeutic targets that are present in all cancer cells from a patient can be identified. Following clonal dynamics and the gain and loss of mutations post-therapy could identify constraints in tumour evolution that can be exploited. This Chapter focused on deconvoluting the clonal structure of the individual tumours used in this study and inferring the order of mutation gain.

Accurate clonal ordering requires the input of allele-specific CN data. The shallow-pass WGS data does not provide enough depth for the estimation of B-allele frequencies and therefore cannot be used to generate allele-specific CN data. Due to the popularity of WES as a method to study genomic variation, there are now several packages that can identify CNAs from WES data[245,339,340].

The FACETS[245] package was used to extract CN data from the WGS data. FACETS was able to detect substantially more CNAs than shallow-pass WGS, including subclonal CNAs as well as clonal CNAs and regions of CN neutral LOH. Shallow-pass WGS gave an average coverage of 0.7x. Whilst this gives enough coverage to identify larger CNAs present in the majority of cells, and can give a good indication of the level of genome instability within a tumour, it is not enough for the identification of smaller

CNAs, CN neutral LOH or CNAs that are subclonal. Subdividing the unique CNAs identified by FACETS into clonal and subclonal CNAs showed that the majority of clonal CNAs, which hypothetically can be detected by shallow-pass WGS, were below the size limit of detection for shallow-pass WGS. Those that were larger tended to be regions of CN neutral LOH, again undetectable by shallow-pass WGS.

The ability of FACETS to identify subclonal CNAs helped disambiguate the SNaPshot analysis result for the *FGFR3* mutation in patient P1175. In Chapter 3, tumour P1175-S01-BX displayed a homozygous SNaPshot profile for *FGFR3* p.G372C, whilst tumour P1175-S02-PX displayed a heterozygous profile. WES identified a VAF of 0.89 for tumour P1175-S01-BX and 0.84 for tumour P1175-S02-PX, suggesting that both tumours are homozygous for this mutation. Analysis of the CN for this mutation with FACETS identified that it was in a region of subclonal LOH for both tumours. For tumour P1175-S01-BX, this was a region of CN neutral LOH with an estimated CCF of 0.65, whilst for tumour P1175-S01-PX this region of LOH had an estimated CCF of 0.3 but it was also precited to contain a single copy gain, producing 3 copies of the mutant allele. Unfortunately, FACETS only provides one estimate for a region, so it is not possible to identify if all cells with LOH of *FGFR3* in tumour P1175-S02-PX had the additional gain, or if this was a subpopulation of cells with LOH. Nevertheless, this subclonal nature explains why the wild-type allele could be identified in P1175-S02. It is interesting that no wild-type allele was identified in tumour P1175-S01 during SNaPshot analysis, however this could have been at the lower limit of detection or sampling of the DNA may have been uneven.

Prior to running clonal clustering algorithms, it was desirable to first identify if there were likely to be any subclones present. In CN neutral regions of the genome the variant allele fraction (VAF) is a readout of the proportion of cells containing the mutation[91]. For example, heterozygous mutations present in all cells of a pure tumour would have a VAF around 0.5 in a CN neutral region. If a tumour contains clusters of mutations with dissimilar VAFs this is evidence of intratumour heterogeneity and is suggestive of the presence of clusters of cells with different genotypes[53,341]. Kernel-density analysis was used to quickly visualise the spread of the VAFs[341] to see if it was worth proceeding with clonal clustering and ordering. At least two peaks of different VAFs within the CN neutral regions of each tumour was identified providing evidence for the presence of subclones. This suggested that further analysis of the tumours to identify the number and contents of the subclones was worthwhile.

The subclone composition of the tumours was assessed using PyClone, a CN-aware Bayesian clustering method for grouping sets of mutations[246]. PyClone uses information on allele imbalances resulting from changes in copy number and the normal cell contamination along with the observed VAF to provide an estimate of the cellular presence of each mutation. Mutations with similar cellular prevalence are grouped together to form clusters. Analysis of multiple tumours from the same patient allows the identification of mutations that change in cellular prevalence together and this helps to refine clusters.

Correlating with the kernel density analysis, PyClone analysis identified all tumours to be heterogeneous. In all 8 patients an ancestral origin was identified consisting of a set of shared mutations present at a high cellular prevalence. These were designated the founding clones. Each tumour then contained subclones with a median of 2 subclones observed per tumour (range 1-4). Previous studies identified a similar small number of subclones for NMIBC with 1 to 2 private subclones identified per tumour[61,164].

Tumours from patients P1175 and P2161 showed very simple clonal dynamics with only a single subclone identified for each tumour producing a single possible evolutionary tree for each patient. For all other tumours at least two evolutionary trees were possible due to the prospect of either linear or branched evolution of the minor subclones. Sequencing of bulk tumour tissue makes these types of events very difficult to disambiguate as there will always be an amount of noise and error in the data[342]. To be able to explicitly determine the type of evolution occurring and refine the evolutionary tree, single cell sequencing would need to be implemented. This would provide the exact genotype of a cell, from which the true evolutionary path could be deducted through the analysis of multiple cells[343]. Single cell sequencing has been applied to many cancers[54,344,345] including MIBC[169], but as yet it has not been used in the sequencing of NMIBC.

Patient P0960 showed a complicated clustering of mutations from which an evolutionary tree could not be generated using the ClonEvol package. This was due to the incompatible cellular prevalence of cluster 3 across the tumours. The similar presence of this mutation in the two post-MMC tumours suggests that this is a real event. This would also make contamination of the post-MMC tumours unlikely as it would have to happen twice. Targeted deep sequencing would need to be employed to try to disambiguate the clusters further. This highlights that a limitation of this study is the relatively low level of coverage. Tumours were sequenced to an average depth of 80x; a heterozygous mutation in a subclone consisting of 10% of the total tumour mass

would therefore only have 4 reads covering the mutation in a pure tumour. This may mean that important subclones present at very low VAFs may not have been detected or biases in the sequencing may have distorted the VAF of certain variants.

This low coverage could mean that it is possible that some of the post-MMC unique variants may have been outgrowths of undetected subclones in the pre-MMC tumours. As the identified MMC-related signature is very specific and correlates highly with the *in vitro* literature, it is unlikely that these mutations were present in the pre-MMC tumour. The majority of these mutations were subclonal. It is possible that some of the other mutations, such as those derived from APOBEC mutagenesis, that became highly prevalent post-MMC may have been present at very low VAFs pre-treatment. This can be evidenced by the observation that the nonsense mutation of *RB1* identified uniquely in tumour P0533-S04-PX was also identified in P0533-S02 by targeted sequencing. Again, ultradeep sequencing of selected variants would provide validation of the MMC-associated events and allow for better separation of subclones. Due to time and resource constraints, targeted re-sequencing could not be performed as part of the current project but would be a key follow up experiment in later projects.

PyClone makes several assumptions during analysis. It assumes that all cells of a clonal population have the same CN status, that no site mutates more than once in its evolutionary history and that mutations do not disappear[246]. These assumptions are easily violated especially as CNAs or LOH events could occur before and/or after the mutational event resulting in loss of variants. These limitations are highlighted with the results from patient P0418 for whom cluster 1 produced a result that was incompatible with the evolutionary model (Figure 5.8A). Cluster 1 increased from subclonal to clonal after MMC treatment, whereas other clusters that were clonal pre-MMC treatment became subclonal. This made clonal ordering impossible due to incompatible CCFs between the two tumours. Cluster 1 contained 6 mutations including the *PIK3CA* E545K hotspot mutation. Further analysis of these variants showed that they were in regions of subclonal CN deletions in P0418-S02-BX yet at a normal CN in P0418-S03-PX. It was thought that the subclonal nature of the CNAs may be confounding the clustering. Removing the subclonal CN estimation by returning these variants to a normal CN estimation did indeed make these variants cluster with the ancestral variants as expected. The use of a clustering algorithm that considers subclonal CNAs and incorporates CNAs as events into the subclonal deconvolution, such as Canopy[336], would likely improve the clustering and delineate more subclonal populations. This could provide interesting results as highlighted by the *FGFR3* mutation in patient

P1175, which clearly shows subclonal diversity with respect to CNAs. Unfortunately there was not enough time to perform this analysis during the project.

In Chapter 3, phylogenetic trees were drawn for each patient using CN alterations whilst in this chapter trees have been created using the subclonal clusters of mutations. Comparing the trees from the two methods shows similar results in patients that have tumours with several CNAs. For example, phylogenetic ordering of CNAs for patient P0418 identified that tumours P0418-S02-BX and P0418-S03-PX shared many alterations before branching occurred. After branching both tumours gained private CNAs, but tumour P0418-S03 gained more CNAs than tumour P0418-S02. This was reflected in the mutation data with tumour P0418-S03-PX containing more private SNVs than tumour P0418-S02-BX. Similarly, for patient P2218, all tumours contained some CNAs that were shared. Tumours P2218-S1A-BX and P2218-S1B-BX then branched off together with an additional 2 shared CNAs between them before diverging (see Appendix G). This was also seen in the mutation data where these tumours shared additional mutations that were not present in tumour P2218-S02. Tree building using CNAs in tumours containing few CNAs was uninformative or showed very simple linear evolution of tumours. For example, tumours from patient P2161 contained only one CNA that was shared making tree building difficult. The addition of mutation data identified a simple evolution pattern for patient P2161, with a single branching event. This data shows that phylogenetic tress predicted from CN data appear to accurately reflect the evolution of tumours as determined by mutation data, but only if many CN events are present. Unsurprisingly, mutation data provides a deeper analysis of the evolution of the tumours, especially in tumours with few-to-zero CNAs. In all cases, the use of a clustering algorithim that can order both SNVs and CNVs would help subclonal reconstruction and identify the order of mutational gain.

As sequencing was performed on bulk tumour tissue from a single region of each tumour it is possible that some subclones may have been missed due to spatial positioning. Lamy *et al.*[61] performed multiregional exome sequencing on a single muscle-invasive tumour and found no mutations or subclones unique to any one region. However, the mean sequencing depth was 61.1x, so it is entirely possible that rare mutations were missed. Further studies investigating spatial heterogeneity would be of use, however the low numbers of subclones identified in NMIBC would suggest that spatial heterogeneity may not be highly prevalent in this disease.

An ancestral clone was identified for all patients and this tended to contain the majority of driver mutations identified. Overall, 26 driver genes were identified as ancestral

events present in the trunk of the tumour evolutionary tree from both the WES and targeted sequencing panels (from a total of 12 patients with monoclonal disease). Of these, 6 were present on the trunk of more than one patient. These included *FGFR3* (7 patients), *PIK3CA* (5 patients), *KDM6A* (4 patients), *ARID1A* (2 patients), *RHOA* (2 patients) and *ELF3* (2 patients). Targeting of truncal driver mutations would ensure targeting of all tumour cells. In the case of activating mutations, this could be done through small molecule inhibitors. Inhibition of FGFR3 is being investigated for the treatment of bladder cancer[346] as well as other cancers[347]. However, a common problem with small molecule inhibitors is the development of resistance through additional mutations[348] or the activation of equivalent or downstream pathways[349]. Additionally, targeted treatment would need to be localized as systemic administration would likely cause unacceptable toxicity in these lower-risk patients that have a long-life expectancy.

Interestingly, targeted sequencing for patient P1870 identified that the driver mutation in *KDM6A* (p.V558Efs*36), predicted to be an ancestral event by exome sequencing, was actually a later event not present in tumour P1870-S01. This became clonal and remained dominant in tumours resected at a later date. These results identify the need to sequence all tumours from a patient for the accurate identification of the early truncal events before applying a personalised targeted therapy regime.

Chromatin modifiers are frequently mutated in bladder cancer[100] and these can be early events as evidenced by truncal mutations of *ARID1A* and *KDM6A.* Epigenetic changes have the ability to be manipulated using pharmaceuticals and therefore could represent a therapeutic option for bladder cancer[350]. However, the mutations in *ARID1A* and *KDM6A* are frequently inactivating, meaning that directly targeting the protein product of these genes is not possible. Recently, *ARID1A* deficiencies have been exploited through synthetic-lethal interactions[351,352]. Inhibition of aurora kinase A was shown to confer selective vulnerability in *ARID1A* deficient colorectal cancer cells[351] and inhibition of EZH2, a methyltransferase, was shown to be synthetically lethal in ovarian clear cell carcinoma cells and caused regression of *ARID1A* mutated tumours in a mouse model[352]. EZH2 inhibition has also been shown to be effective in *KDM6A*-null bladder cancer cell lines. Ler *et al.*[177] showed that cell lines with loss of *KDM6A* were sensitive to EZH2 inhibition whilst a cell line with wild-type *KDM6A* did not respond to EZH2 inhibition at all. Additionally, EZH2 inhibition inhibited growth of *KDM6A* mutant tumours in cell line and patient derived xenograft models[177]. These studies highlight that when a mutant gene product is not directly targetable, knowledge

of the signaling pathways and interactors of the protein may identify interactions that can be targeted instead.

Whilst the majority of driver events were identified in the ancestral clusters, some subclones contained additional potential driver mutations whilst others did not. The lack of drivers in some subclonal clusters could mean that the event driving this outgrowth is yet to be identified. However, it is also possible that the driver events in the ancestral clone were sufficient to drive recurrent tumour outgrowth and the subclonal mutations identified in these clusters are simply passenger events.

The subclone composition of bladder tumours will be influenced by tumour resection[164]. It is likely that this physically removes some subclones with private mutations and may leave behind cells containing the ancestral mutations due to a large area of the urothelium being altered[4]. Indeed, the observed presence of an ancestral clone in all tumours is suggestive of a “field defect” in the urothelium. This may explain the lack of shared subclones identified between tumours from the same patient as these have been removed. Only three patients, P0418, P0960 and P2161, contained subclones that were shared across tumours. This could be indicative of incomplete resection and regrowth, undetected microscopic growths or could be due to the shedding and reimplantation of tumour cells. Whole organ mapping studies of cystectomy specimens have shown that geographically distant regions of altered mucosa that may appear phenotypically normal are clonally related[22,168]. Sequencing of adjacent normal samples in patients with multifocal or recurrent disease identified some of the mutations present in the tumours[55]. Mutations that were shared between multifocal tumours, or spatially distinct regions within a tumour, were much more likely to be detected than private mutations suggestive of spread of the ancestral mutations forming the field disease. By combining whole-organ mapping with next generation sequencing the extent of this widespread field disease can be fully realized, a full understanding of which could help inform therapeutic design and surgical procedures.

An interesting occurrence is the observation of convergent evolution in tumours from many of the patients. Private mutations hitting the same driver genes within a patient were identified in 3 of the patients analysed by exome sequencing and two of the targeted sequencing panel. In patient P0533 both exome-sequenced tumours contained disruptive mutations in the *RB1* gene; a splice-donor mutation in P0533-S3B-BX and a nonsense mutation in P0533-S04-PX. These mutations were both present in clusters that had undergone a clonal sweep and were likely present in almost all cells of each tumour (Figure 5.5). A deletion of 13q13.3 - 13q14.3, covering

the genomic position of *RB1*, was identified in all tumours from patient P0533, suggesting that almost all pRB activity was eradicated in both tumours. The convergence on complete loss of pRB activity could suggest that loss of this pathway was important for tumour outgrowth in this patient and may represent a therapeutic opportunity[353], either through targeting the E2F pathway[354-356] or through the conferred increased sensitivity to radiation therapy[357] and cisplatin-based therapies. Separate mutations in *PIK3CA* were also identified in P0533-S02 and P0533-S04. These are known APOBEC targets[83] and likely reflect the activity of the APOBEC enzymes rather than convergent evolution.

Other driver genes seen to be privately mutated in tumours from the same patient include *KMT2C* in patient P2161, *TSC1* in patient P2218, *FAT1* in patient P0468 and *ATM* and *KDM6A* in patient P0536. Interestingly patient P2218 had loss of chromosome 9, on which *TSC1* is encoded, and tumour P0418-S01 had loss of a region of chromosome 4 including the region encoding *FAT1.* Like *RB1*, both of these are tumour suppressor genes thus function of the proteins encoded by these genes are likely eradicated. Convergent evolution of mutations affecting the same gene in different subclones, often separated by space or time, has been reported in many cancers including renal cell carcinoma[53], chronic lymphocytic leukaemia[358] and brain metastases[64]. This may signify that tumour evolution has constraints. Further analysis of multiple regions and tumours from individuals is required to shed increased light on the convergent nature of cancer. Perhaps this will identify predictable evolutionary paths that could be targeted[359].

Investigations into the mitomycin C-associated variants identified that these were present in the subclonal clusters. Analysis of the mutation spectrum showed two very different substitution patterns in the clonal vs subclonal variants and statistical analysis of these differences confirmed a significant difference in the number of C > A, C > T and C > G mutations. The subclonal mutations contained predominantly C > A mutations, consistent with those identified in Chapter 4 and correlating with the increase in C > A mutations identified in *in vitro* studies[201,202,204]. This resulted in a much-reduced transition:transversion ratio for the subclonal variants. It was reasoned that the unique tandem substitutions identified in the post-MMC tumours could be used to represent the MMC-induced event as this was an event specific to these tumours. Only 2 tandem substitutions were identified in post-MMC unique clonal clusters. The rest were present in subclonal clusters, providing further evidence that the MMC-associated variants tend to be subclonal in nature.

The observation of therapy-induced subclonal mutations is not unique. A high proportion of mutations associated with a cisplatin signature in MIBC were found to be subclonal[176] and treatment with temozolomide has been shown to induce subclonal hypermutation in glioblastoma[360]. In MIBC it was suggested that the short interval between treatment and recurrence may not have provided enough time for mutations to undergo a clonal sweep, thus the subclonal nature of the chemotherapy related events was suggested to be in relation to time[176]. In our cohort the median time from the start of the MMC treatment course until the sequenced post-MMC tumour was 232 days (range 97-486 days). This may not have provided enough time for clones containing MMC-associated mutations to grow to clonal levels. It would be interesting to look at tumours from patients that recurred later to see if MMC-induced events become clonal.

Copy number and hotspot gene mutation analysis failed to determine the clonal origins of tumours from 8 patients. The addition of variant identification, either by WES or by targeted sequencing of the bladder cancer panel, identified a clear monoclonal origin for a further 5 patients (P0468, P0933, P0990, P2161 and P2218). No additional mutations were identified that were shared by all three tumours for patient P0536 but tumours P0536-S01 and P0536-S03 shared an additional 4 mutations. These two tumours therefore arose from a monoclonal origin. The origin of tumour P0536-S02 could not be explicitly determined. However, the presence of the two hotspot mutations would suggest a monoclonal origin. Additionally, convergent evolution was identified in this patient with tumour P0536-S02 also containing mutations in *ATM* and *KDM6A* that were distinct from the mutations in these genes identified in the other two tumours. This may provide further evidence that tumour P0536-S02 is likely clonally related to the other tumours from this patient.

Tumours from patient P2291 did not share any mutations or CN alterations suggestive of an oligoclonal origin. Whilst the majority of bladder cancers show a monoclonal origin[26,29,159,160] a small fraction of patients have been reported to show evidence of oligoclonal tumours[161,361,362]. These early investigations into the clonality of bladder cancer utilized methods such as X-chromosome inactivation[160,362], LOH analysis[26] and *TP53* mutational status[161,361]. Given that cancer is genetically unstable and may have subclonal diversity, the use of small numbers of genetic markers may not provide a full picture of the relatedness of the tumours. Next-generation sequencing provides 10's to 1000's of potential genomic markers that can be used to assess clonality. NGS studies involving paired tumour samples from patients with recurrences[61,164], matched pre- and

post-chemotherapy samples[176] including metastases[62] as well as multifocal tumours[174,180] have all demonstrated clonal relationships between tumours from the same patient. It is possible that shared mutations have not been identified in patient P2291 due to the limited number of genes targeted in the bladder cancer panel utilized. However, as these genes represent the most frequently mutated genes in bladder cancer one would assume that if any variants were to be detected then analysis of this panel of genes would greatly facilitate this. Tumours were assessed by short-tandem-repeat (STR) profiling which confirmed that they came from the same patient. Notes taken by the clinician at the time of resection suggest that the recurrent tumour was resected from around the original resection site making the oligoclonal finding even more interesting. These results are suggestive of a potential “field cancerization” where multiple cells have become transformed. This could come from exposure to mutagens, however no occupational hazards or history of smoking were noted at consent, or could be due to a genetic predisposition. However, these results must be viewed with caution; the single shared reads identified in two mutations may be suggestive of a shared minor subclone. Targeted deep sequencing would be required to confirm or refute this. Sequencing of any additional tumours from this patient would be interesting but as yet this patient has had no recurrences since tumour P2291-S02.

5.4 Summary

Clonal clustering identified the presence of subclones in all tumours. However, the number of subclones identified per tumour was quite low (average of 2, range of 1-4). Analysis of the different clusters identified that mitomycin C-related mutations tended to be subclonal in nature. Targeted sequencing of a panel of genes commonly mutated in bladder cancer was used to improve analysis of clonality for 5 patients. This identified 4 patients with tumours of a monoclonal origin and confirmed a likely oligoclonal origin for tumours from patient P2291.

Chapter 6
Final discussion

Non-muscle-invasive bladder cancer (NMIBC) is a clinically challenging disease. The high rate of recurrence and the possibility of progression to muscle invasive disease in these patients necessitates continued surveillance for many years after diagnosis[6]. Despite having a good prognosis, patients are treated with DNA damaging chemotherapy. Some chemotherapy agents have been shown to create new mutations in cancers[114,177], potentially driving evolution that could lead to treatment resistance and progression. As yet, the mutational effect of mitomycin C (MMC) chemotherapy treatment has not been investigated in bladder cancer. This project aimed to use next generation sequencing (NGS) to characterise the copy number alterations (CNAs) in a cohort of 67 tumours from 23 patients with recurrent disease, and identify any potential genomic alterations associated with MMC chemotherapy in NMIBC.

Initially, a cohort of patients with recurrent non-muscle-invasive disease was analysed for changes in copy number using shallow-pass whole-genome sequencing. This identified that the copy number (CN) profiles of the tumours from each patient were predominantly similar, even after treatment with MMC. For the majority of patients, tumours shared a subset of copy number alterations considered to be likely early events, after which some tumours evolved additional alterations. Hierarchical clustering of the cohort identified that tumours from the same patient tended to cluster together, suggesting that tumours are relatively stable at the CN level. This observation of conservation of CNAs across bladder tumours from an individual patient has been identified in other studies investigating multifocal and recurrent tumours[29,167]. Studies performing hierarchical clustering of multiple non-muscle-invasive tumours from the same patient have also reported that tumours tend to cluster together[163,167]. This stability in CNAs is not limited to NMIBC; investigations in muscle-invasive bladder cancer (MIBC) have also identified that tumours from the same patient cluster close together[62]. Taken together, these studies suggest that some CNAs are early events in bladder cancer and may imply that bladder cancer is relatively stable during evolution at the CN level.

Early *in vitro* studies identified several structural rearrangements in cells after treatment with MMC[194,199,281] leading us to hypothesize that MMC treatment in bladder cancer patients could lead to the formation of CNAs in post-treatment tumours. Analysis of tumours from patients that underwent 6 weeks of MMC treatment demonstrated that

MMC did not appear to alter the CN landscape, with no MMC-specific CN events and no significant change in the fraction of genome altered (FGA) being identified after therapy. For individual patients, some tumours showed an increase in the FGA after therapy, some a decrease, whilst others showed no change at all. This lack of change in the CNA landscape after treatment has also been reported in MIBC treated with cisplatin[176], another DNA alkylating agent that can form intrastrand and interstrand crosslinks[363]. This could suggest that the generation of CNAs by these agents is uncommon. Alternatively, the lesions induced by MMC could be highly cytotoxic and therefore are not represented at a detectable level in the post-treatment samples. Investigations using model cell lines, or tumour organoids, combined with single cell sequencing directly after treatment could be used in the future to identify if MMC does induce CNAs. The maintenance of such alterations in cells could be examined by next-generation sequencing at different follow-up times after drug treatment.

It is also possible that MMC induces structural rearrangements that do not involve a change in copy number. The presence of such rearrangements could be investigated using cytogenetic methods such as FISH and chromosome painting or a different next-generation sequencing strategy. Mate-pair sequencing has been used to resolve structural rearrangements[364,365] and could be used to resolve any potential structural rearrangements not identified by shallow-pass whole genome sequencing (WGS) or whole exome sequencing (WES) in the current study. The technique leverages the known distance between the two read ends to identify any pairs that show discordant mapping. This information can then be used to disambiguate structural rearrangements such as insertions, deletions, inversions and translocations[257]. Alternatively, long-read methods, such as Oxford Nanopore sequencing, could be used to investigate structural rearrangements. The length of a read for Nanopore technologies is limited only by the size of the input DNA[366] and sequencing reads in excess of 2 Mb have been reported[367]. One of the benefits of these long reads is that regions of repeats can be covered in a single read, making alignment easier and therefore making it possible to resolve regions of the genome that are inaccessible with short-read sequencing[368]. Nanopore technology has even been used to sequence through the Y-chromosome centromere[369]. As the structural rearrangements induced by MMC *in vitro* occurred primarily within the repeat rich, heterochromatic, peri-centromeric regions of chromosomes 1, 9 and 16[196-198], Nanopore sequencing technology could allow the identification of breakpoints in these repetitive regions.

Previous studies in model organisms identified deletions ranging in size from 110 bp to 8 kb in mouse models[203] and from 2 bp to 318.8 kb in *C.elegans*[205]. These deletions are small, focal events that are below the limit of detection of our shallow-pass WGS and would therefore have been missed. Higher coverage of around 5x[364] would allow the identification of these regions but would come with an increased overall cost as currently samples are multiplexed in pools of 50 and this generates an average raw coverage depth of 0.7x.

As the majority of MMC-associated mutations were subclonal, it could be hypothesized that any copy number alterations caused by treatment may also be subclonal. These would be missed by shallow-pass whole genome sequencing as the coverage is not deep enough for this type of analysis. Analysis of the exome sequencing data using the FACETS package enabled the identification of subclonal CNAs, yet no difference in the number of subclonal CNAs was identified between the two treatment groups (data not shown). It is possible that CNAs present in the non-coding regions of the genome could be missed by WES, especially in long intergenic regions, as exons are not evenly placed within the genome[370].

To investigate the effect of MMC at the single nucleotide level, pre-MMC and post-MMC treatment tumours from 8 patients underwent WES. Patients were selected from the copy number cohort if their pre- and post-treatment tumours occurred within a short timeframe (less than 2.5 years) and no instillations of BCG therapy had been given between the tumours. Tumours were assessed for changes in mutational burden, types of base substitution and mutational signatures. This identified an increase in the number of C > A transversions and an increase in tandem substitutions, specifically at CC or GG bases in the MMC treated tumours. These alterations are concordant with mutations caused by MMC *in vitro*[201,202] and *in vivo*[203].

Deconvolution of the mutational signatures in this study associated these MMC-induced changes with COSMIC mutational signature 24, the aflatoxin signature. It is likely that this reflects the preference of both MMC and aflatoxin for the generation of C > A mutations. This highlights a limitation of mutational signatures at present, as many signatures may have a similar sequence context preference. The recent analysis by the PanCancer Analysis of Whole Genomes (PCAWG)[312] will likely improve delineation of signatures. However, to generate a true signature of MMC mutagenesis, more tumours need to be sequenced.

Alongside the increase in C:G > A:T mutations, an increase in tandem substitutions, specifically at CpC or GpG dinucleotides, was identified in the MMC treated tumours. These results reflect those of previous *in vitro* studies[201,202,204], however *in vivo* studies did not identify any increase in single base mutations and instead identified an increase in deletions[203,205]. To be activated, MMC first requires reduction. Reduction of the drug may occur differently within a tumour environment compared to within model organisms, and this could explain the lack of correlation between the two. In the *in vitro* models, Srikanth *et al.*[201] used $NaBH_4$, which has the potential for both mono- and bi-functional activation, whilst Maccubbin *et al.*[202] used only mono-functionally activated MMC. Both identified single base changes associated with monofunctional activation. Additionally, Srikanth *et al.* detected tandem substitutions, associated with intra-strand crosslinks from bifunctional activation. The deletions identified in the *in vivo* studies are likely induced by interstrand crosslinks which may have been missed in the study by Srikanth *et al.* as they may have interfered with the packaging of the lambda phage. This could suggest that in the model organism studies, MMC was always bifunctionally activated, and therefore not able to generate SNVs. The presence of both SNVs and tandem substitutions in our study suggests that both mono- and bifunctional activation occurred within the tumours.

Overall, there was no difference in the total mutational load between pre- and post-MMC tumours. However, each post-MMC tumour contained many private mutations, with an average of 29% of the total mutations per patient being unique to the post-MMC tumour. Across all patients, 31% of the post-MMC unique mutations were observed in a signature that we associated with MMC therapy (signature 24). This suggests that whilst there was no change overall in mutation load, MMC was a significant contributor to the mutational burden post-treatment. A similar picture was painted in MIBC treated with cisplatin-based chemotherapy, where no increase in mutational burden was identified post-chemotherapy despite the generation of new, treatment-associated mutations[176].

In the case of the muscle-invasive study, the pre-treatment samples were from biopsies, not tumour resections, and patients were treated with a neo-adjuvant regime[176]. In our study, each patient underwent tumour resection prior to chemotherapy, from which samples were taken for sequencing. Patients then underwent a course of MMC treatment (which started between 1 day and 10 months later) followed by resection of the post-MMC tumour (3 to 16 months after treatment initiation). Analysis of tumours at the subclone level identified that between pre-MMC

and post-MMC tumours, a median of 1.5 subclones were lost (range: 1-3) and 2 gained (range: 1-3). Contrasting with observations in the MIBC study[176], no minor subclones pre-treatment were identified as expanded post-treatment as only highly prevalent subclones or the ancestral clone was maintained post-MMC treatment. Due to tumour resection, it is impossible to say if MMC treatment eradicated any clones or if these were removed by resection alone. This data could suggest that resection removes the most recently evolved clones that are likely present in the tumour outgrowth. The ancestral mutations and highly prevalent, early subclones could possibly be identified in the tissue surrounding the tumour and it is possible these could lead to tumour outgrowth. Indeed, in the analysis of the adjacent normal urothelium in recurrent and multifocal patients, mutations that were shared by multiple tumours were more likely to be detected in the "normal" sample[55]. Full organ mapping of a cystectomy specimen using next-generation sequencing would provide valuable information on the spreading of these clones.

MMC is a DNA damaging agent and overall a trend towards an increase in mutations was seen in post-MMC tumours. A significant increase in the number of non-silent mutations affecting coding regions of genes was identified. This could possibly be related to the sequence specificity of MMC for CpG dinucleotides which are enriched in exons and promoter regions relative to the rest of the genome[206]. It would be interesting to investigate the distribution of MMC induced mutations using whole genome sequencing to identify if these are indeed enriched in the promoter and exonic regions. Additionally, if mutations in the promoter regions are identified, it would be interesting to see if they have any effect on gene expression levels. Alternatively, the increase in non-silent coding mutations could reflect positive selection for potentially functional mutations. Investigations into the synonymous and non-synonymous mutation rates could shed some light on this. However, it was not possible to carry out such analysis in the current study due to time limitations.

The increase in the number of non-synonymous mutations observed in MMC-treated tumours might lead to some subclones expressing an increased number of neoantigens, thus increasing immunogenicity[371]. As identified in other cancers, such as lung[372] and melanoma [373], a high mutation burden has been associated with an increased response to immunotherapy in MIBC[374,375] and this is thought to be related to the increase in novel neoantigens. BCG is an immunotherapy and is preferentially used for high-risk NMIBC where outcomes are better than for chemotherapy. Stage T1 grade 3 tumours have been shown to have a higher level of genome instability through

CN analysis[125,126] and a higher mutational burden in high-risk NMIBC has been associated with reduced progression and better response to intravesical BGC therapy[173]. If an increased number of non-synonymous mutations post-MMC treatment is a feature of such treatment it could be hypothesized that these patients will respond better to BCG therapy and may suggest that a combination of the two therapies may improve response rates. Indeed, several clinical studies have investigated the efficacy of combining mitomycin C and BCG for the treatment of NMIC[376-380]. A study investigating the use of sequential BCG and MMC vs BCG alone for the treatment of patients with carcinoma *in situ* (CIS) did not support the use of the combination therapy for the treatment of CIS[379], a conclusion supported by other studies[221,376]. However, a study investigating the combination in intermediate and high-risk NMIBC patients demonstrated an improvement in the disease-free interval and recurrence rate, yet this was at the cost of a greater toxicity compared to BCG alone[378].

Many of the studies investigating combination therapy have used different treatment schedules and compared to different control groups: some compared to BCG alone, others to MMC alone. Deng *et al.*[381] performed a systematic review of 25 such studies. Overall, they identified that a combination of both MMC and BCG showed improved responses. They postulated that the induced disruption of the urothelium by MMC treatment could improve the attachment of BCG to the bladder wall and improve anti-tumour activity. The results from our study could suggest that an increase in non-synonymous mutations might also contribute to an improved response. The prediction of neoantigen generation would have provided further evidence for this hypothesis. It is possible to predict the presence of neoantigens from WES data[382], however there was not enough time to carry out such an analysis during the timeframe of the current project. Follow-up work for publication should include this analysis.

It would be interesting to analyse mutation data from patients undergoing the BCG-MMC combination therapy as well as mutation data from patients undergoing BCG monotherapy and compare this to the data generated by our study. This could help explain the improved response identified in patients undergoing the combination therapy and may help identify why BCG is better than MMC in the treatment of high-risk patients.

In this cohort the total mutation rate is higher than that reported previously for NMIBC by Hurst *et al.*[82] (5.90 vs 2.41 mutations per megabase (Mb) respectively) and this may be due to the different stages and grades used in the two cohorts. Other studies have

used NGS to investigate higher-risk NMIBC; Meeks *et al.*[173] investigated high-risk NMIBC whilst Pietzak *et al.*[157] investigated both low- and high-grade stage Ta tumours as well has high-grade stage T1 tumours. Both of these studies report a much higher mutation rate with 10-15 mutations per Mb identified by Meeks *et al.* and a mutation rate of 9 mutations per Mb in high-grade tumours reported by Pietzak *et al.* Both of these studies used targeted sequencing panels which inflates the mutation rate and makes them less easy to compare to other studies without additional bioinformatic steps. Additionally, the use of different variant calling methodologies could result in inconsistent numbers of variants being identified. For example, Hurst *et al.*[82] used VarScan2 to call their variants and this variant caller was identified to be the most stringent variant caller in our analysis. This means that their analysis may have missed some lower-frequency mutations. In comparison, our pipeline was developed to identify shared variants, including those at low frequencies, whilst maintaining a low false-positive rate. This highlights the benefits of large sequencing schemes, such as the TCGA and ICGC, as these provide a standard approach to analysing samples, making results more comparable.

Clonal evolution analysis identified mostly simple subclonal evolution patterns. All tumours had a set of ancestral mutations that remained present in all cells throughout tumour development. After accounting for these ancestral clones, pre-MMC and post-MMC tumours differed in the subclones they contained; subclones were lost from the pre-MMC tumours and different subclones gained in the post-MMC tumours. No subclones were identified that were present at a low level in one tumour yet expanded in another tumour. It is possible that minor subclones were missed in this analysis due to the relatively low depth of sequencing for subclone delineation. Alternatively, these subclones may have been removed by resection as discussed previously. Deeper targeted sequencing of selected mutations could be used to confirm shared and unique mutations. Time and resources did not allow for this to occur during the project, but this kind of validation would be important for future work.

The observation of Sylvester *et al.*[213] that a single immediate instillation of chemotherapy resulted in an increased risk of death in higher risk patients is suggestive that chemotherapy influences tumour evolution in this patient subgroup. Our data shows that MMC treatment can certainly induce new mutations and there is the possibility that these could contribute to disease progression. In our cohort, no patients had progression to muscle-invasive disease. As a follow-up study, it would be interesting to analyse tumours from patients who progressed after receiving multiple

instillations of MMC to examine whether the progressed tumours contain MMC-induced alterations, what the frequency of those alterations are, and if treatment alters subclonal dynamics. This represents a feasible study as patients undergoing a course of MMC chemotherapy are of intermediate to high-risk of progression, therefore such samples should be available for analysis.

One of the objectives for this study was to investigate the relationships between tumours from the same patient. Overall, a monoclonal origin was identified for 21 out of 23 of the patients in the study using a combination of copy number analysis and hotspot mutation analysis. This was supplemented by sequencing of a targeted gene panel for tumours where necessary. Due to a lack of available sample, it was not possible to carry out further analysis using targeted sequencing of tumours from one patient (P2065) for whom monoclonality could not be established. Two tumours from this patient both contained loss of chromosome Y but only tumour P2065-S01 also contained hot-spot mutations in *FGFR3* and *PIK3CA*. As there are only two loci available for comparison in these tumours (i.e. the two mutations) neither a monoclonal or oligoclonal origin can be defined for this tumour. Targeted sequencing of whole - genome amplified DNA could help define the origins of these tumours but there was not enough time available to optimise and implement this.

An oligoclonal origin was suggested for tumours from patient P2291 with 6 unique mutations identified in tumour P2291-S01 and 8 unique mutations identified in tumour P2291-S02. To our knowledge, this is the first study using NGS-based approaches to identify a possible oligoclonal origin for recurrent tumours from a bladder cancer patient. It is possible that limiting analysis of these tumours to 140 genes may have missed shared mutations elsewhere. However, this gene panel is representative of the most frequent mutations in bladder cancer. Whole-genome sequencing of this patient would be interesting. It would unambiguously define the relationship between the two tumours and, if monoclonality was demonstrated, could identify alterations that occur very early on in bladder cancer. Additionally, it would be interesting to look for possible pathogenic germline variants within this patient as these could provide evidence for a predisposition to bladder cancer.

Overall, the combination of these results and others from the literature[26,29,61,62,164] suggest that the majority of recurrent and multifocal bladder cancers are clonal in origin. The analysis of multiple related tumours from an individual generates a higher resolution for the identification of early events within bladder carcinogenesis, as alterations need to be present in all these tumours. In this study, *TERT* promoter

mutations, chromosome 9 loss and *FGFR3* mutations were identified to be early events, present on the trunk of the tumour evolutionary trees. These results are consistent with previous data[154,163,383].Through the assessment of more paired data it may be possible to identify additional common early events, especially if this is done using deep sequencing data. It is essential that this analysis uses multiple tumours from the same patient as this is the only way to disambiguate highly prevalent clones as evidenced by patient P1870, for whom analysis of an additional tumour by targeted sequencing identified that the driver mutation in *KDM6A* predicted to be ancestral by exome sequencing of two tumours was actually a later event.

There are several limitations to this study. The primary issue is the relatively low number of samples analysed. Pre-MMC and post-MMC tumours from 8 patients were analysed by WES with a total of 18 tumours being assessed. The data generated did, however, clearly demonstrate MMC-associated alterations in 7 out of the 8 patients. The similarities between our observations and mutations generated by MMC experimentally suggests that these findings are truly reflective of the effects of treatment with a course of MMC chemotherapy. Sequencing of an additional cohort of tumours from patients that have undergone such a course of therapy would validate the results from this project.

There was no control group used within this study, which is a limitation. It would be difficult to define a control group, as the majority of patients with recurrent disease will undergo some form of treatment. For the identification of the MMC signature, the lack of a control group is mitigated in part by the heterogeneity of the treatment timings of the patient cohort. For some patients, the pre-treatment tumour was the initial primary tumour whilst for others it was a later recurrent tumour, yet the same signature was found in all cases. This is highly suggestive that the signature is indeed a MMC-related signature and not related to recurrent tumours. Additionally, using matched pre- and post-treatment samples is a form of control within itself. However, sequencing of a panel of paired, treatment naïve, recurrent tumours would allow us to determine if treatment with MMC increases the mutation rate in the recurrent tumours compared to resection alone.

The sequencing depth of the current study limits the ability to investigate the clonal dynamics. Despite achieving an average coverage of 80x, only 71% of targeted bases were sequenced to 50x or higher, and many mutations had coverage of only 30x. Targeted re-sequencing of variants with an orthogonal method, such as Ampliseq, would ensure accuracy of the sequencing results. Important mutations in terms of

clonal dynamics could be targeted to ensure that these were not present at low levels in the pre-treatment tumours.

As patients underwent resection prior to treatment, this study was limited in its power to detect any mutations present in subclones that may have conferred sensitivity to MMC treatment. Additionally, the results from this study suggest that resection may limit the detection of subclones that may be resistant to the therapy. A recent clinical trial investigating the use of MMC in the neo-adjuvant setting could shed some light on possible markers of sensitivity or resistance. The trial, chemoresection and surgical management in low-risk non-muscle-invasive bladder cancer (CALIBER - clinical trial number: NCT02070120), aimed to identify if MMC could be used to chemoresect tumours by treating patients with 4 once weekly instillations of MMC, rather than using surgery. Unfortunately, the trial ended early as the complete response rate at 3 months was low in the chemoresection group (estimated rate of 37.3% vs 80.8% in the surgery group)[384]. Sequencing of samples from patients that underwent chemoresection in this trial would provide a greater insight into the role of MMC in clonal dynamics. Comparing tumours from patients that failed chemoresection with those who had a response could identify markers of recurrence or sensitivity.

Whilst we have investigated the alterations present in DNA after MMC treatment, we have not examined genome-wide mRNA expression data for these samples. mRNA subtyping of bladder cancer has identified distinct subtypes, and these display differential survival outcomes and sensitivity to chemo- and immuno-therapies (reviewed in [385]). It would be interesting to use mRNA profiling to subtype the tumours and identify if MMC treatment changes the subtype of the tumour at all.

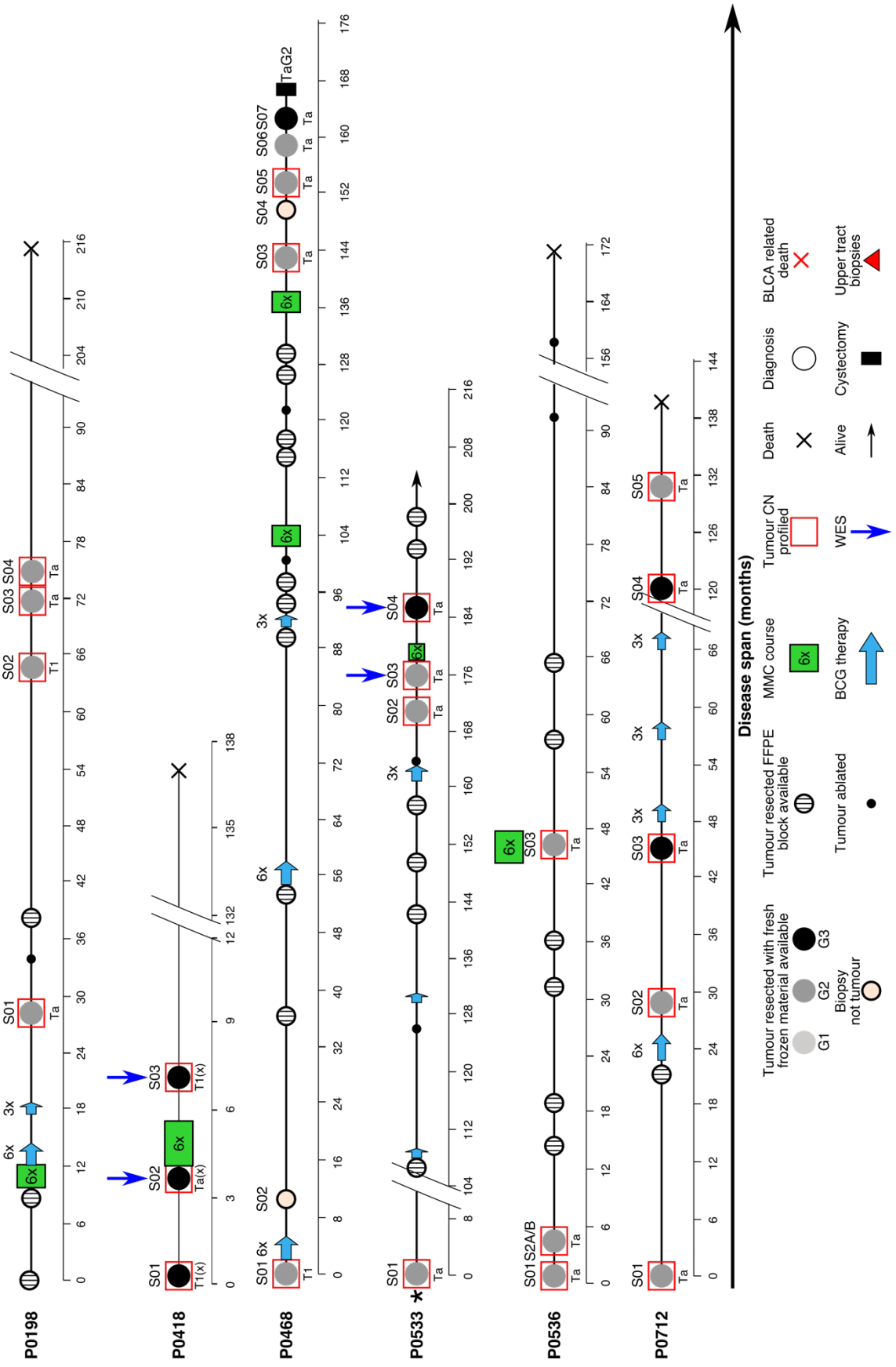
This study may have underestimated the extent of intratumour heterogeneity. Biopsies were taken from the resected samples and this may mean that we have missed some subclones due to spatial separation of the subclones. In MIBC, multi-region sequencing of primary and metastatic bladder suggested a complex mixture of clones, however these were not spatially distinct[63]. An additional study that analysed 8 regions from a single muscle-invasive bladder tumour also identified intermixing of the identified subclones in all regions[61]. This may suggest that spatial heterogeneity is low within bladder tumours but, as yet, this not been systematically investigated in NMIBC.

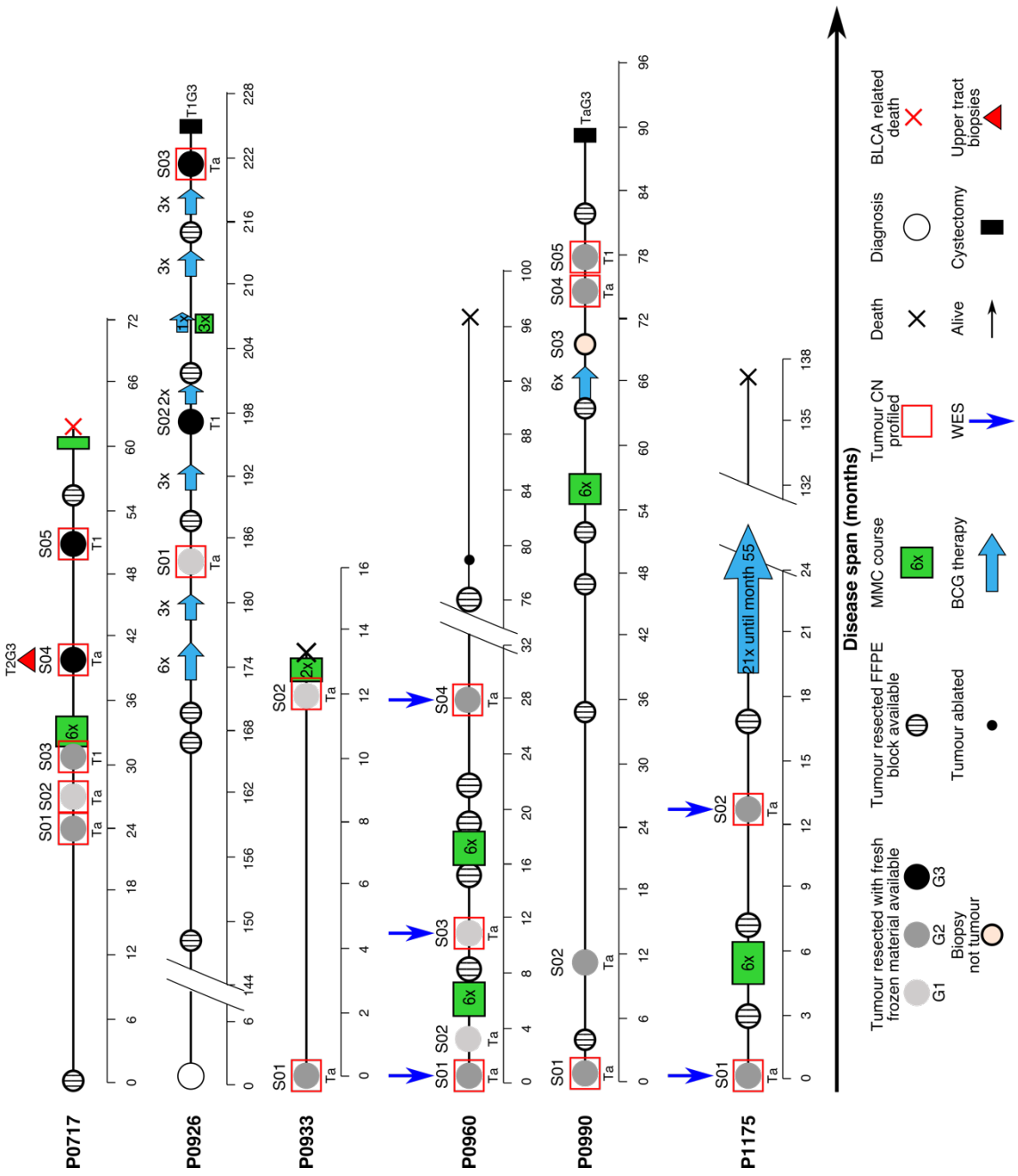
In conclusion, our study has shown that MMC has a clear mutational effect on tumour DNA that can be identified in post-MMC treated tumours. The clinical implications of these findings are as yet unknown. The ability of MMC to generate new mutations

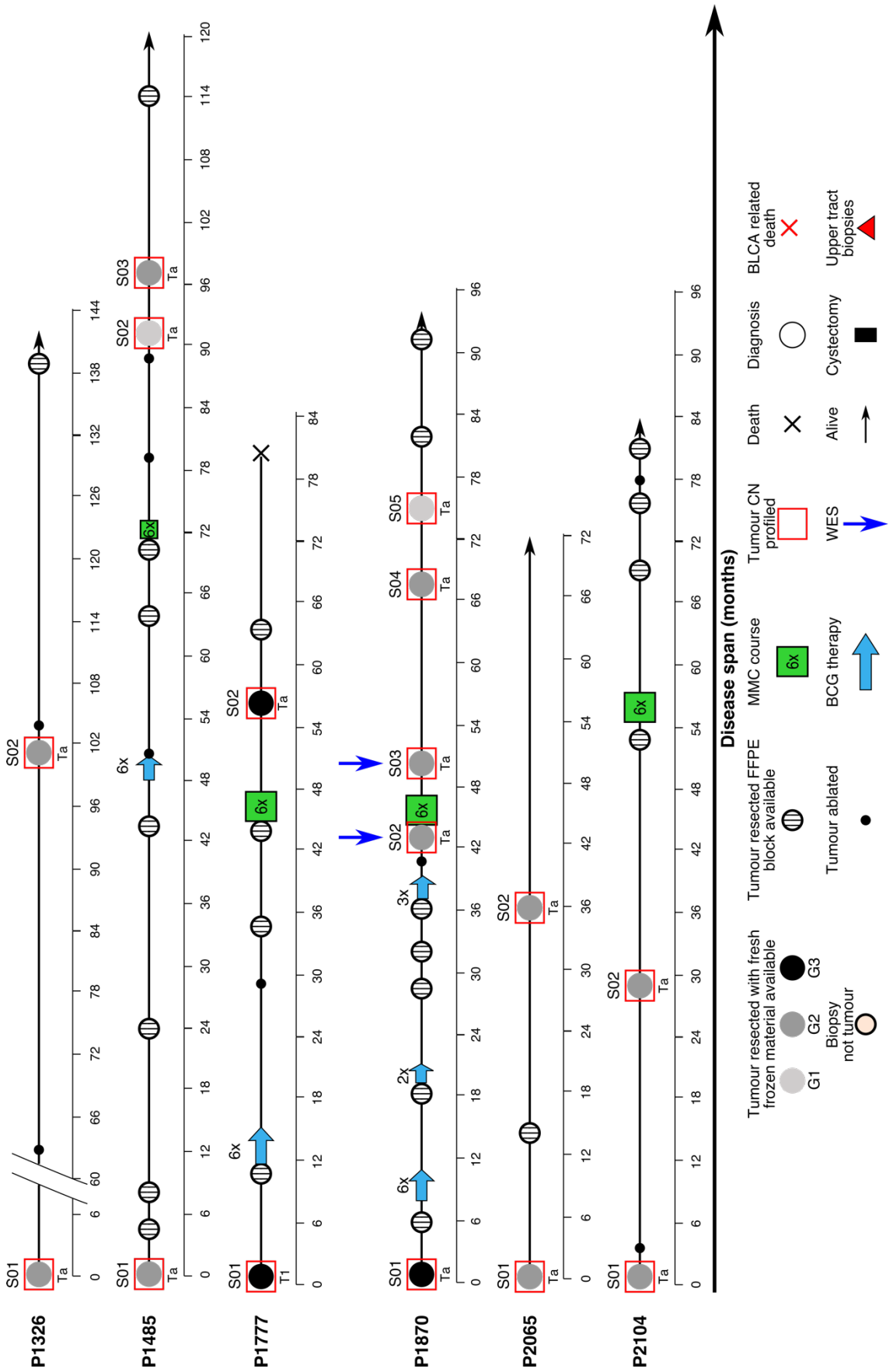
provides an opportunity for these mutations to provide a growth advantage to a cell with an otherwise low malignant potential, which could promote recurrence and progression. On the other hand, the generation of non-synonymous mutations could increase the number of neoantigens and consequently the immunogenicity of these tumours thus providing a potential opportunity for the use of a combination of MMC and BCG treatment. Ultimately, more studies are required to investigate the mutagenic landscape of MMC treatment and to discern the effect of MMC treatment on the clonal dynamics of NMIBC. The incorporation of genomic data, such as WGS, WES or RNA-seq data into clinical trials would help identify possible correlations between cancer genome characteristics and patient outcome. Additionally, a large-scale systematic review of the long-term outcomes of patients receiving a course of mitomycin C is warranted to ensure that the treatment is of benefit to patients in the long term.

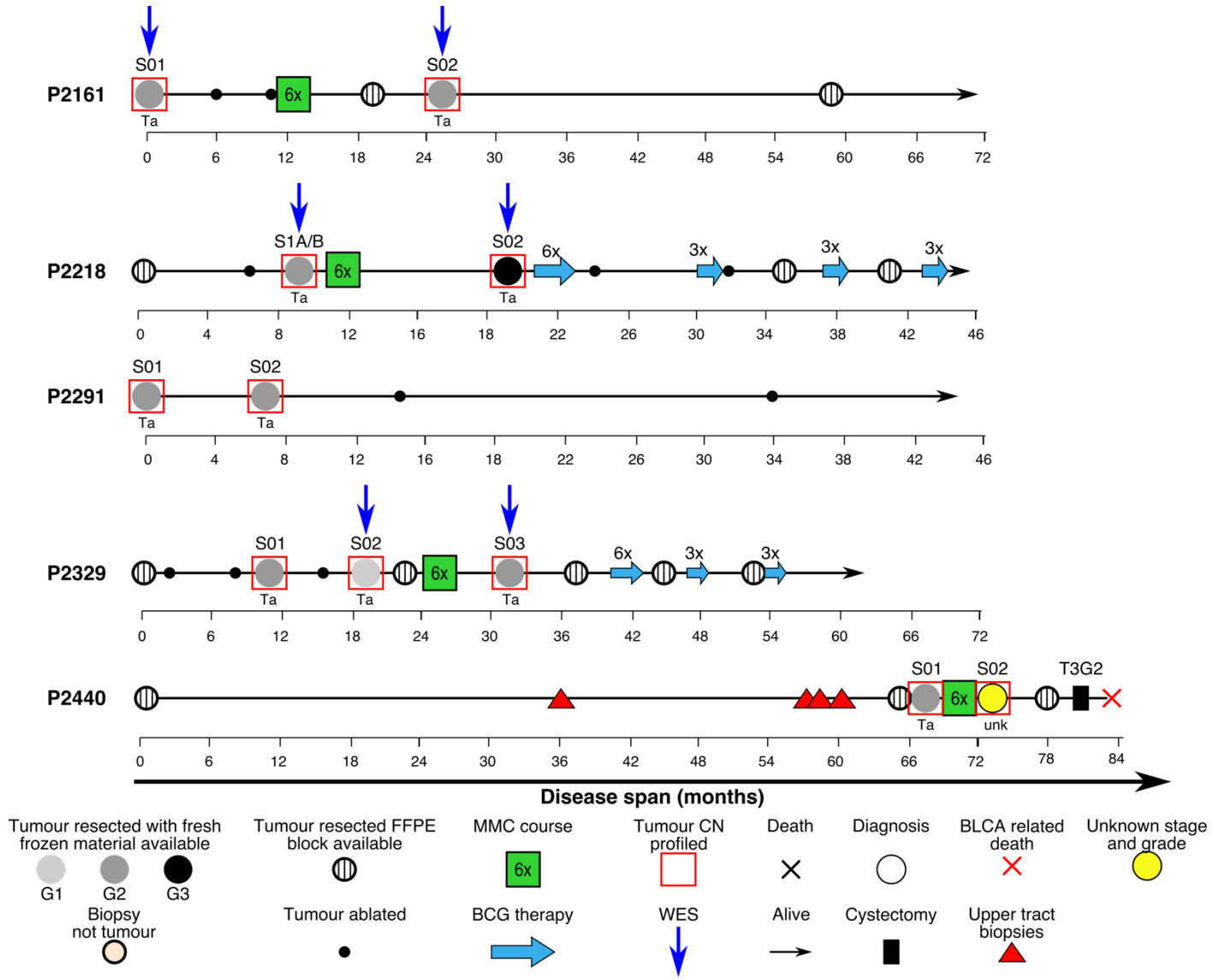
Appendix A
Clinical timelines for all 23 patients

The disease history for each patient is depicted over the next four figures. Tumour events for which fresh frozen biopsy material was available are depicted by large solid circles. Tumours that were resected but not sampled and have formalin fixed paraffin embedded (FFPE) tumour material available are indicated by dashed circles. Other events that may have been removed by diathermy are indicated by small black circles. Mitomycin C (MMC) treatment courses are indicated by green boxes. BCG treatment is indicated by light blue arrows. Tumours that underwent copy number (CN) profiling are indicated by a red square. Tumours that underwent whole exome sequencing (WES) are indicated by dark blue arrows. Red triangles represent resection of upper tract disease. * P0533 was diagnosed 3 years previous but there is no information available for this time point


P0198
P0418
P0468
P0533
P0536
P0712
Disease span (months)
Tumour resected with fresh frozen material available
G1
G2
G3
Biopsy not tumour
Tumour resected FFPE block available
Tumour ablated
MMC course
BCG therapy
Tumour CN profiled
WES
Death
Alive
Diagnosis
Cystectomy
BLCA related death
Upper tract biopsies


P0717
P0926
P0933
P0960
P0990
P1175
Disease span (months)
Tumour resected with fresh frozen material available
G1
G2
G3
Biopsy not tumour
Tumour resected FFPE block available
Tumour ablated
MMC course
BCG therapy
Tumour CN profiled
WES
Death
Alive
Diagnosis
Cystectomy
BLCA related death
Upper tract biopsies
21x until month 55


P1326
P1485
P1777
P1870
P2065
P2104
S01
S02
S03
S04
S05
Ta
T1
6x
3x
2x
Disease span (months)
Tumour resected with fresh frozen material available
G1
G2
G3
Biopsy not tumour
Tumour resected FFPE block available
Tumour ablated
MMC course
BCG therapy
Tumour CN profiled
WES
Death
Alive
Diagnosis
Cystectomy
BLCA related death
Upper tract biopsies


P2161
S01
S02
6x
Ta
Ta
0 6 12 18 24 30 36 42 48 54 60 66 72
P2218
S1A/B
S02
6x
3x
3x
3x
Ta
Ta
0 4 8 12 16 18 22 26 30 34 38 42 46
P2291
S01
S02
Ta
Ta
0 4 8 12 16 18 22 26 30 34 38 42 46
P2329
S01
S02
S03
6x
3x
3x
Ta
Ta
Ta
0 6 12 18 24 30 36 42 48 54 60 66 72
P2440
S01
S02
T3G2
6x
Ta
unk
0 6 12 18 24 30 36 42 48 54 60 66 72 78 84
Disease span (months)
Tumour resected with fresh frozen material available
G1
G2
G3
Biopsy not tumour
Tumour resected FFPE block available
Tumour ablated
MMC course
6x
BCG therapy
Tumour CN profiled
WES
Death
Alive
Diagnosis
Cystectomy
BLCA related death
Upper tract biopsies
Unknown stage and grade

Appendix B

Black-listed regions for CNA analysis

Chromosome	Start	End
chr1	123400000	144323271
chr2	89346407	95501179
chr3	90112909	93794219
chr4	48936766	51851287
chr5	45617513	50369323
chr7	56951145	63142518
chr8	43176239	47320014
chr9	39034407	68120289
chr10	37984863	42524659
chr11	50163315	55410244
chr12	34352090	38227527
chr13	17700000	18753153
chr14	17200000	19696838
chr15	17000508	23409383
chr16	32470325	46572884
chr17	21686799	27369989
chr18	0	117149
chr18	14397773	21233605
chr19	24065361	27794772
chr20	26300552	31408829
chr21	0	12000000
chr22	0	16927042
chrX	57919249	61000000
chrY	10400000	11919235
chrY	56677925	57227415

Appendix C

Files used for phylogenetic tree generation

C.1 ngCGH output example

Name	Chromosome	Start	End	Value
chrY_2781914	chrY	2781914	3533334	-2.190506
chrY_3534345	chrY	3534345	4439886	-2.141596
chrY_4439889	chrY	4439889	5357964	-2.228308
chrY_5357964	chrY	5357964	6287018	-2.153669
chrY_6290705	chrY	6290705	7094954	-2.215597
chrY_7095527	chrY	7095527	7598757	-1.726559

C.2 TuMult probe file example

Name	Chr	StartPosition	EndPosition	StartCytoband	EndCytoband
Chr-1-10002	1	10002	844162	1p36.33	1p36.33
Chr-1-844743	1	844743	944382	1p36.33	1p36.33
Chr-1-944657	1	944657	1013776	1p36.33	1p36.33
Chr-1-1013787	1	1013787	1073139	1p36.33	1p36.33
Chr-1-1073197	1	1073197	1155631	1p36.33	1p36.33

C.3 TuMult profile file example

P0198_S1.value	P0198_S1.status	P0198_S2.value	P0198_S2.status	P0198_S3.value	P0198_S3.status	P0198_S4.value	P0198_S4.status
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
-0.73757	-1	-0.8885	-1	-0.82799	-1	-0.92319	-1
-0.73757	-1	-0.8885	-1	-0.82799	-1	-0.92319	-1

C.4 TuMult reference data set example

	Chr-1-10002	Chr-1-844743	Chr-1-944657	Chr-1-1013787	Chr-1-1073197
P0198B	0	0	0	0	0

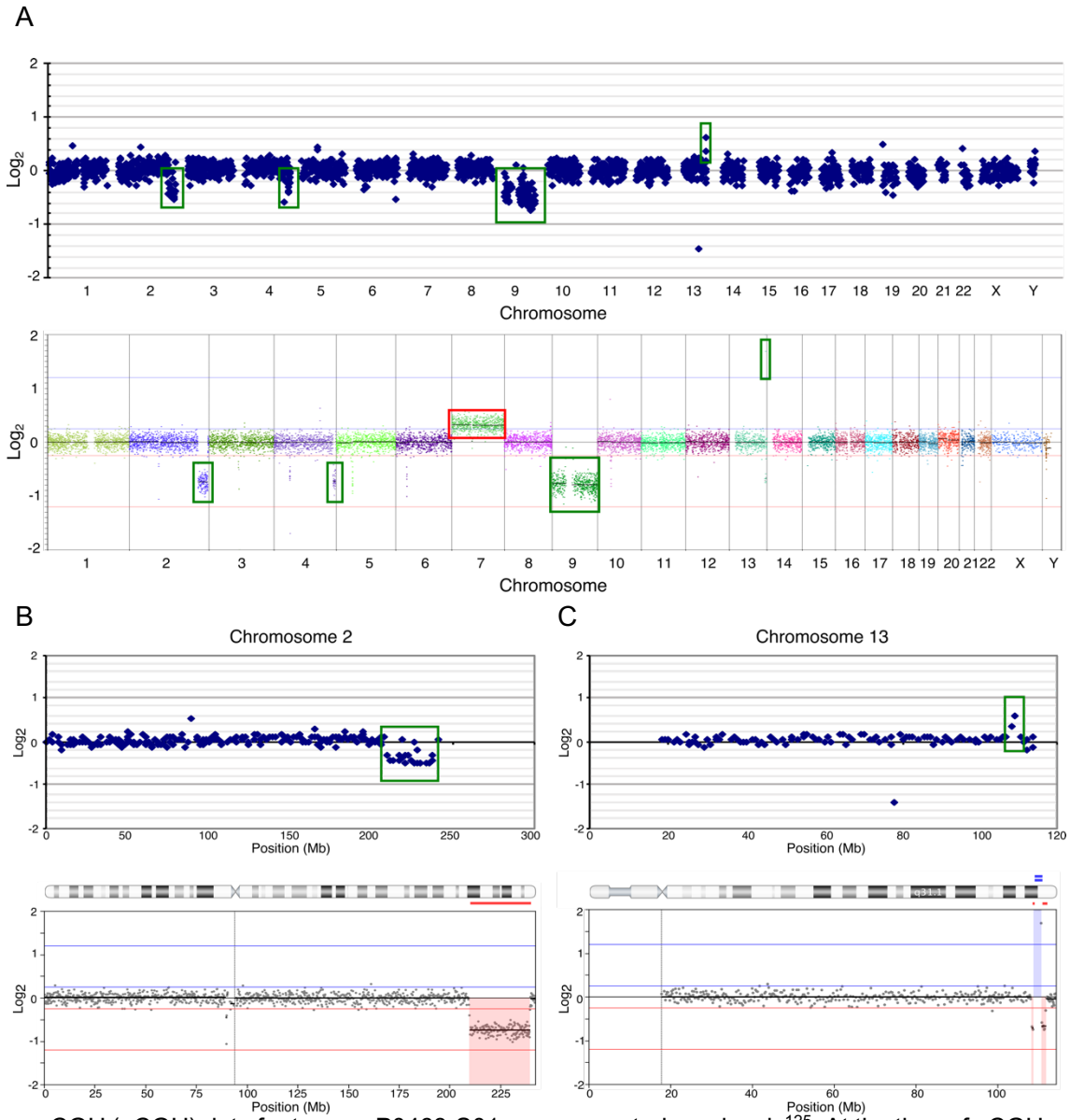
Appendix D

Genes in bladder cancer targeted sequencing panel

Symbol	Symbol	Symbol	Symbol
ACAN	EGFR	LARP1B	RAD21
AHNAK2	ELF3	LGALS8	RARG
AKT1	EP300	LPHN3	RB1
ARHGAP18	EPG5	LRRC7	RBM10
ARHGEF10	ERBB2	MAGI3	RBM6
ARHGEF3	ERBB3	MAML1	RHOA
ARID1A	ERCC2	MAPK8IP3	RHOB
ARID2	ESPL1	MECOM	RREB1
ARID4A	FANCA	MYCBP2	RXRA
ASH1L	FAT1	MYO5B	RYR2
ASXL2	FAT2	NAT10	SCN1A
ATM	FAT3	NCOR1	SLC25A48
ATP6V1B2	FAT4	NCSTN	SPTAN1
ATP7B	FBXW7	NF1	STAG1
B3GNT9	FGFR3	NFE2L2	STAG2
BRAF	FMN1	NFE2L3	STK38
BRCA2	FOXA1	NOTCH1	SYNE1
BTG2	FOXQ1	NOTCH2	SYNE2
C1ORF173	FREM2	NOTCH3	TET3
CACNA1D	HAUS6	NRAS	TEX15
CCND1	HEPACAM	OSMR	TNC
CCND3	HERC1	PAIP1	TP53
CDKN1A	HMCN1	PALM3	TRAK1
CDKN2A	HRAS	PCDHA9	TSC1
CDKN2B	HRNR	PDZD2	TSC2
CEP290	INADL	PGS1	TXNIP
CHD6	ITK	PHF3	UEVLD
CLTC	KDM3A	PIK3CA	USP47
CLU	KDM6A	PIK3R1	UTY
COL11A1	KIF16B	PIK3R4	VCAN
CPAMD8	KLF5	POLE	WHSC1L1
CREBBP	KMT2A	POLE2	WNK1
DLG4	KMT2C	POTEF	ZFHX3
DOPEY1	KMT2D	PTEN	ZFP36L1
DUX4L4	KRAS	RAB11FIP1	ZFYVE26

Appendix E

Comparison of copy number results from array CGH and shallow-pass WGS for tumour P0468-S01



Array CGH (aCGH) data for tumour P0468-S01 was generated previously[125]. At the time of aCGH analysis, a matched blood sample was unavailable therefore an unmatched lymphoblastoid cell line was used. For the shallow-pass WGS analysis, matched blood was available and used as the paired normal. A) Whole genome copy number plots for tumour P0468-S01 generated using aCGH (top) or shallow pass WGS (bottom). Green boxes highlight shared alterations whilst the red box indicates an alteration that is not shared by the two methods. B) Copy number plots for chromosome 2 generated using aCGH (top) or shallow-pass WGS (bottom). C) Copy number plots for chromosome 13 generated using aCGH (top) or shallow-pass WGS (bottom). For B) and C); green boxes have been drawn around CNAs in the aCGH data that were also identified using shallow-pass WGS. These CNAs are highlighted in blue (gain) or red (loss) in the shallow-pass WGS panels.

The Figure above depicts CNAs identified using aCGH or shallow-pass WGS for the same tumour sample (P0468-S01). The basic principle underlying both aCGH and NGS based CN methods is the same: both methodologies compare a test sample with a reference sample to identify differences. In aCGH this is by competitive hybridization to an array of known sequences whilst for NGS the read depth between the two samples is compared. For both methods the data needs to be corrected prior to analysis; for aCGH this involves correcting for the total signal intensity difference between the two fluorescent channels (as there may have been an imbalance in the amount of DNA loaded) as well as other biases such as differences between the signal from different blocks on the array[386]. For our NGS data there are slightly more challenges to overcome. Each probe is calculated as the log2 transformation of the ratio between the number of reads in the normal and tumour for each window. This can be confounded by overall coverage as one sample may have had more reads in total, potentially due to slightly more of that library entering the sequencing pool. This was corrected by median centering where the median log2 value was subtracted from the entire vector of results, eliminating any total read-depth bias. This also centres the data around 0. The GC content of the region can also affect coverage, with GC-poor or GC-rich regions demonstrating lower coverage than balanced regions of the genome[387]. This is corrected for within the Nexus software using a quadratic model.

The results produced by both methods are largely concordant with the same CN losses being identified on 2q and 4q and chromosome 9 as well as the same gain on chromosome 13 (A). However, there is one difference between the results generated by the two approaches; a low level CN gain on chromosome 7 was identified in the shallow-pass WGS that was not detected in the aCGH data. This could be due to the sensitivity of the two methods; the WGS method appears to show better differentiation between regions of CNA and CN neutral regions as the log2 ratios for the regions of CNA are larger or smaller than those seen by aCGH for gains or losses respectively. This may be due to the inherent difficulty in quantifying small changes in fluorescent signal in aCGH analysis compared to quantifying changes in read depth in the analysis of the WGS data[386]. Alternatively this difference could be due to the use of different sections of the tumour for DNA extraction. As the tumour is stage T1 grade 2 it is possible that the tumour could be heterogenous. Indeed, the low level of the gain suggests that it is likely to be a subclonal alteration. It could also be possible that the gain is an artifact in the WGS method potentially due to an increase in overall coverage in the tumour compared to the matched normal. However, the median centering of the

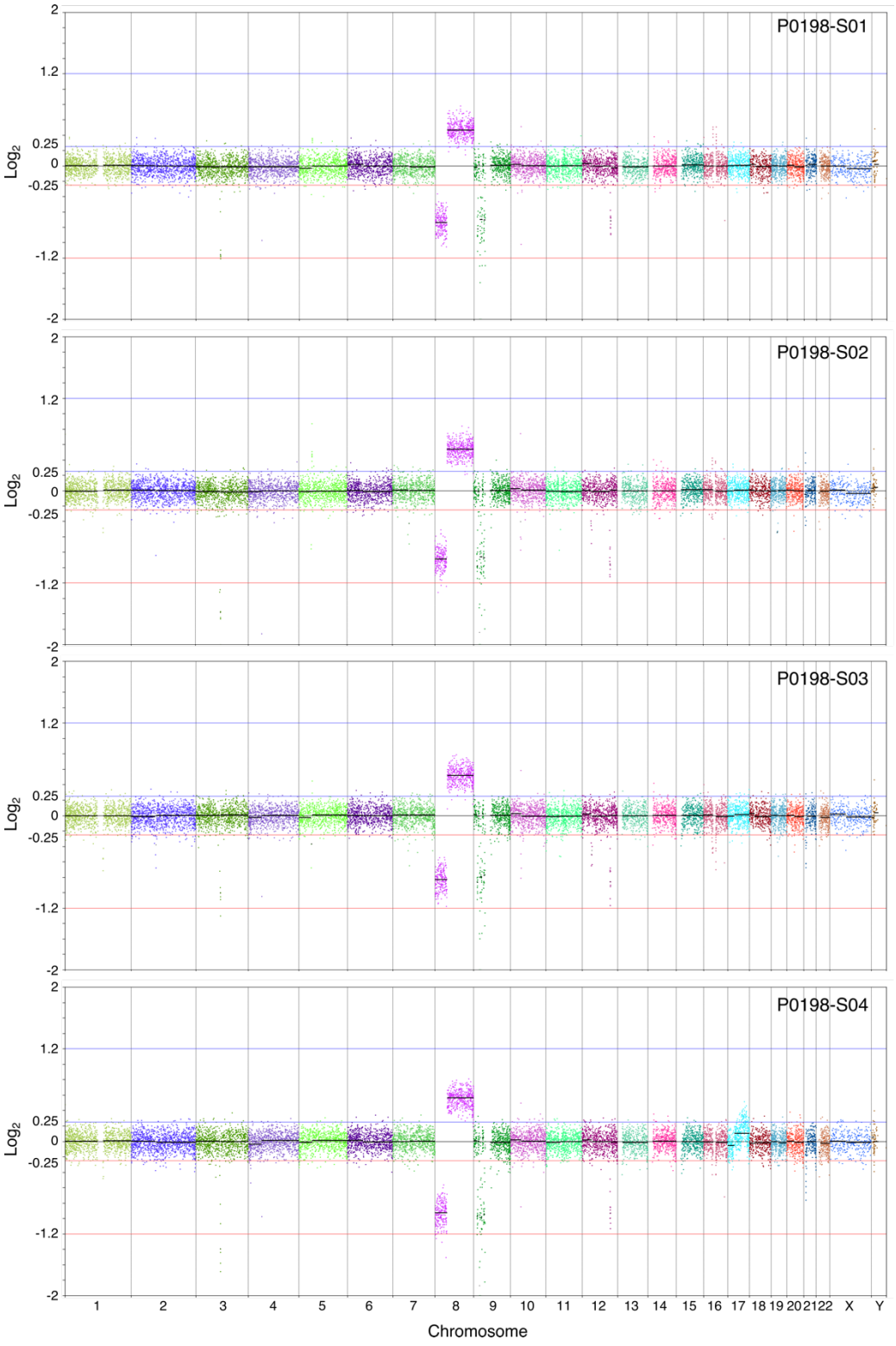
data performed should eliminate this kind of error. Additionally, this low level gain is also seen in another tumour from this patient, validating its presence.

Looking at the individual chromosomes highlights the same CN loss on chromosome 2 (B) but a slightly different profile for the CNAs detected by the two methods on 13q (C). The CN gain on 13q is present in the aCGH data, albeit at a lower amplitude, but the regions of CN loss surrounding the region are not present. This is likely due to differences in the resolution of the two approaches: the array contained approx. 4000 probes spaced just under 1 Mb apart whilst the shallow-pass WGS contained an equivalent of ~12800 probes. In the aCGH data, only 3 probes spanned the entire region of CNA on this chromosome, and with the large size of the probes (average 162 kb) some of the regions of loss and gain were present within the same probe region, reducing the sensitivity of the probe to detect either alteration. In the shallow-pass WGS data, 18 probes spanned the region, improving the sensitivity of breakpoint detection.

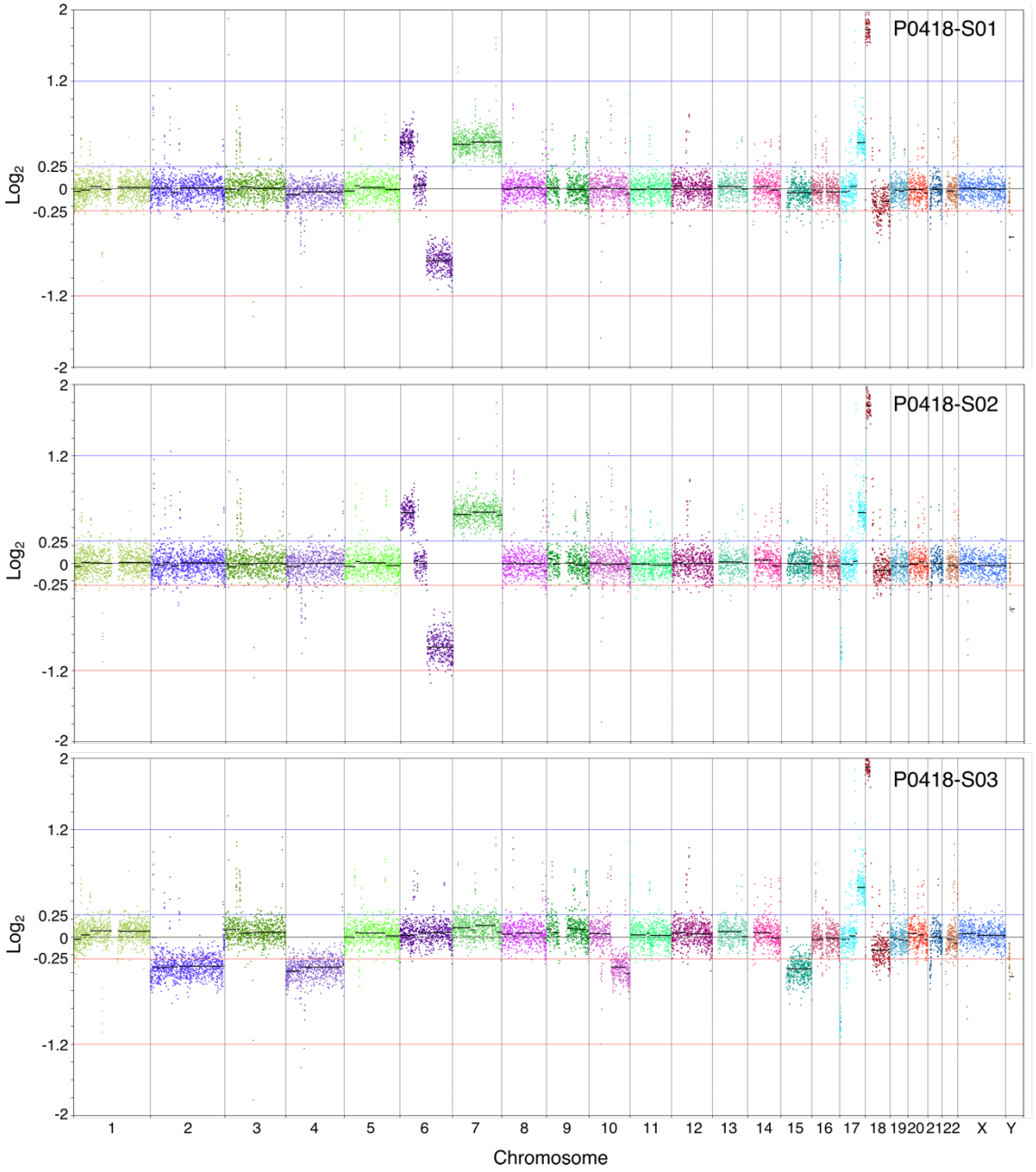
There are some individual probes in the aCGH data that do not align with the rest of the data or with results from WGS. As these are individual probes it is likely that these events are just noise. This may be a consequence of not using a matched normal for comparison in this data.

Appendix F

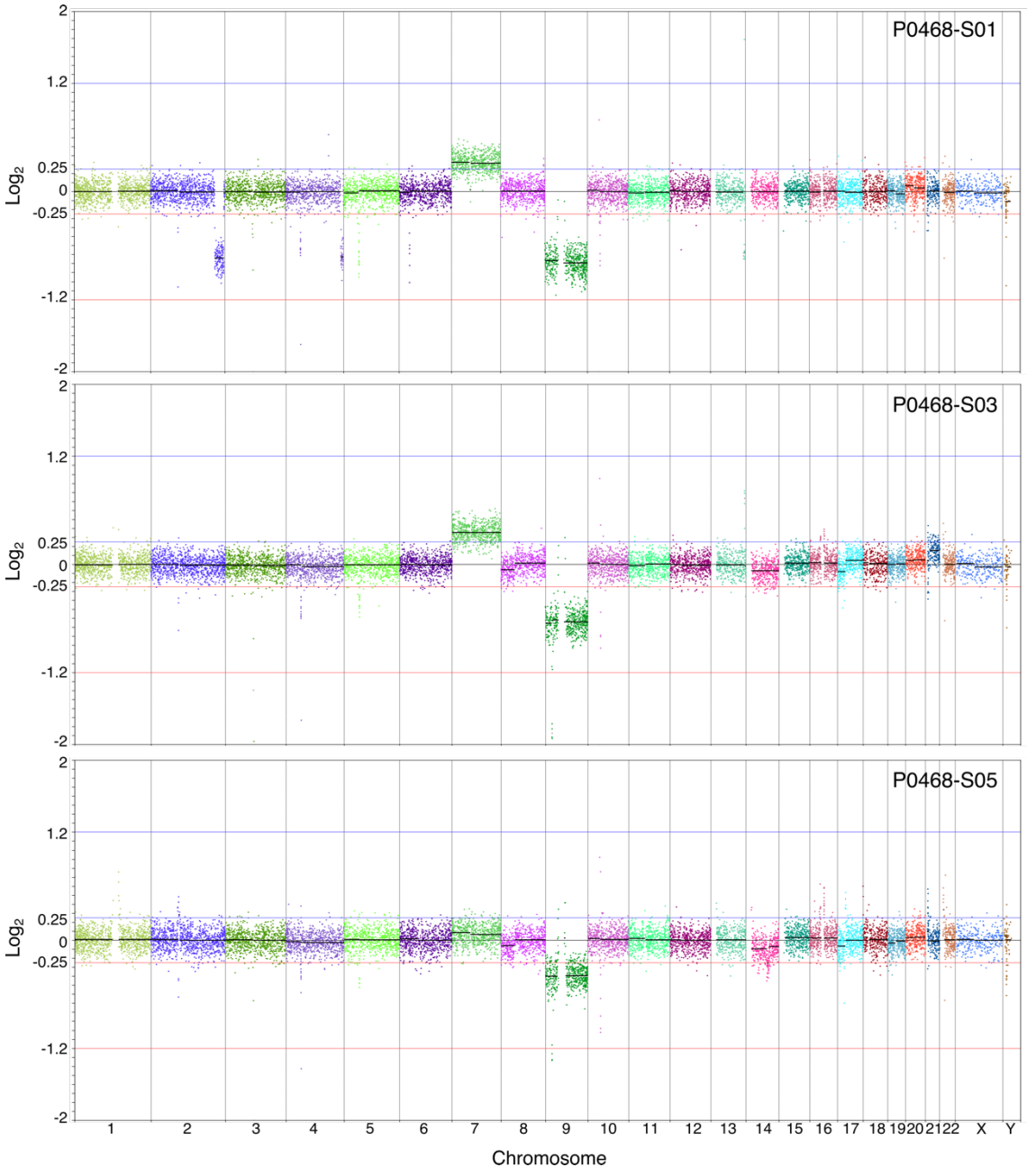
Whole genome copy number plots



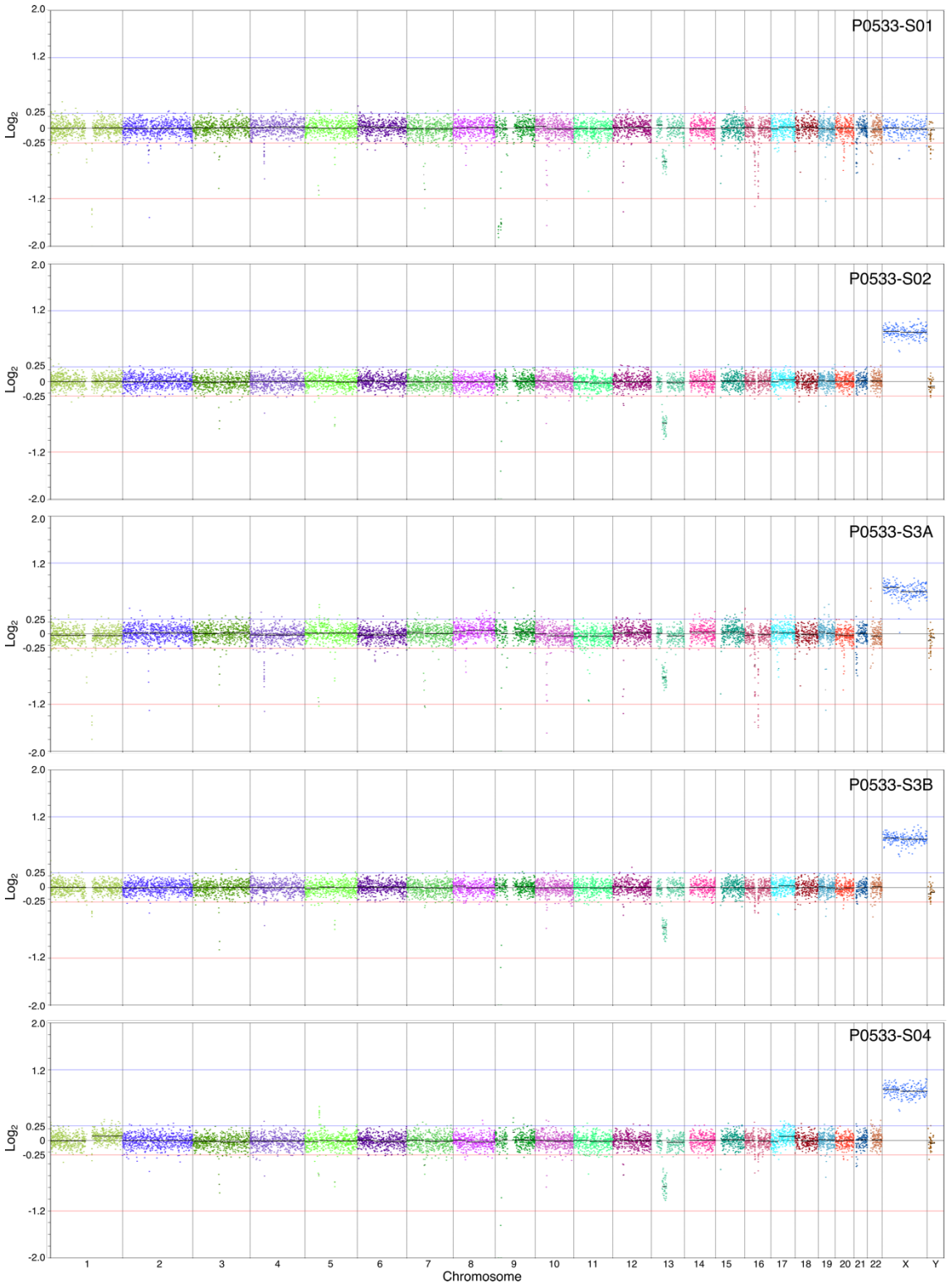
Whole genome copy number plots for tumours from patient P0198.



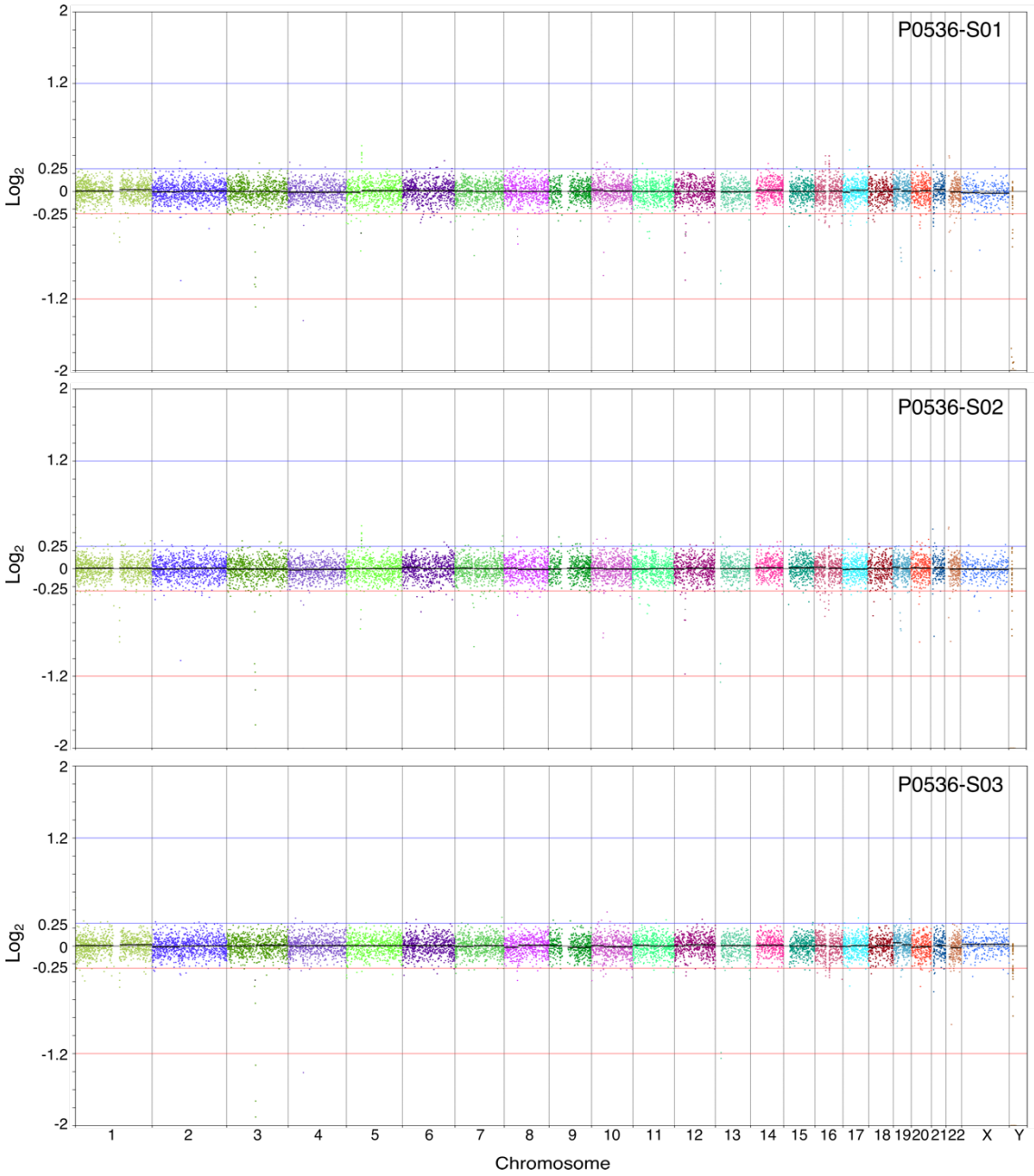
Whole genome copy number plots for tumours from patient P0418.



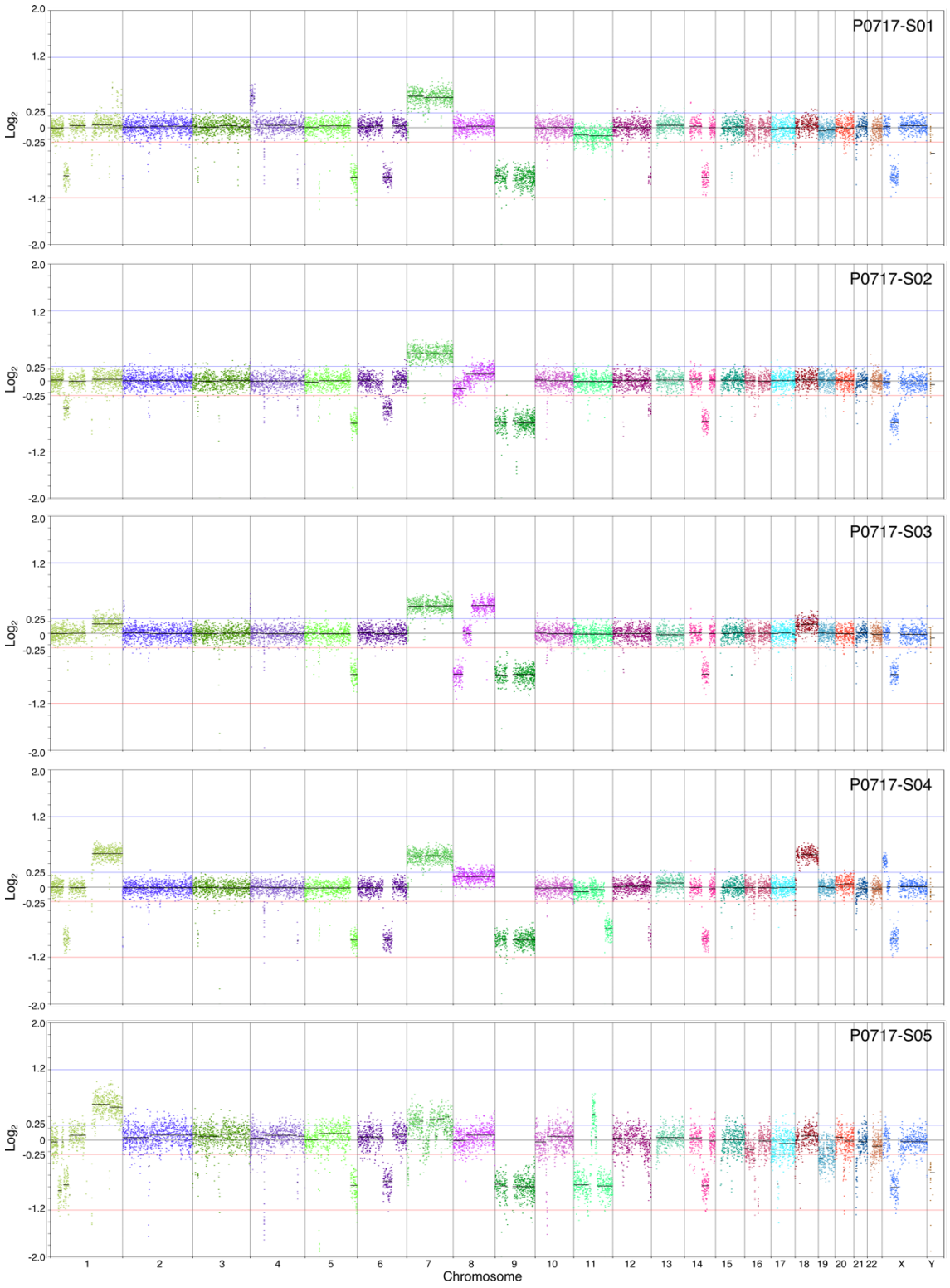
Whole genome copy number plots for tumours from patient P0468.



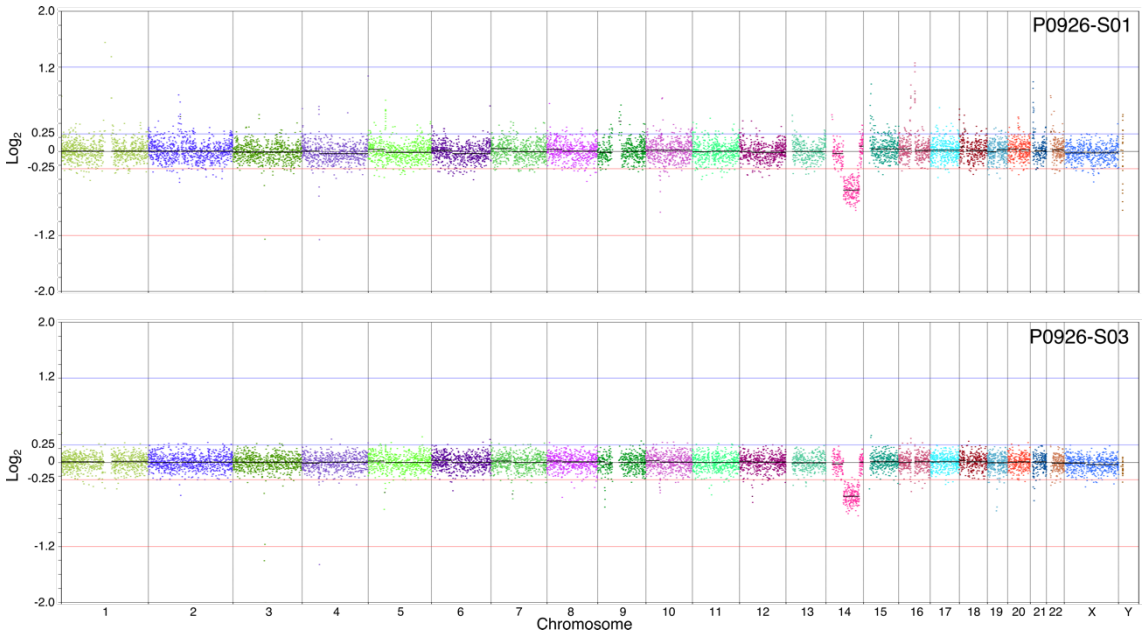
Whole genome copy number plots for tumours from patient P0533.



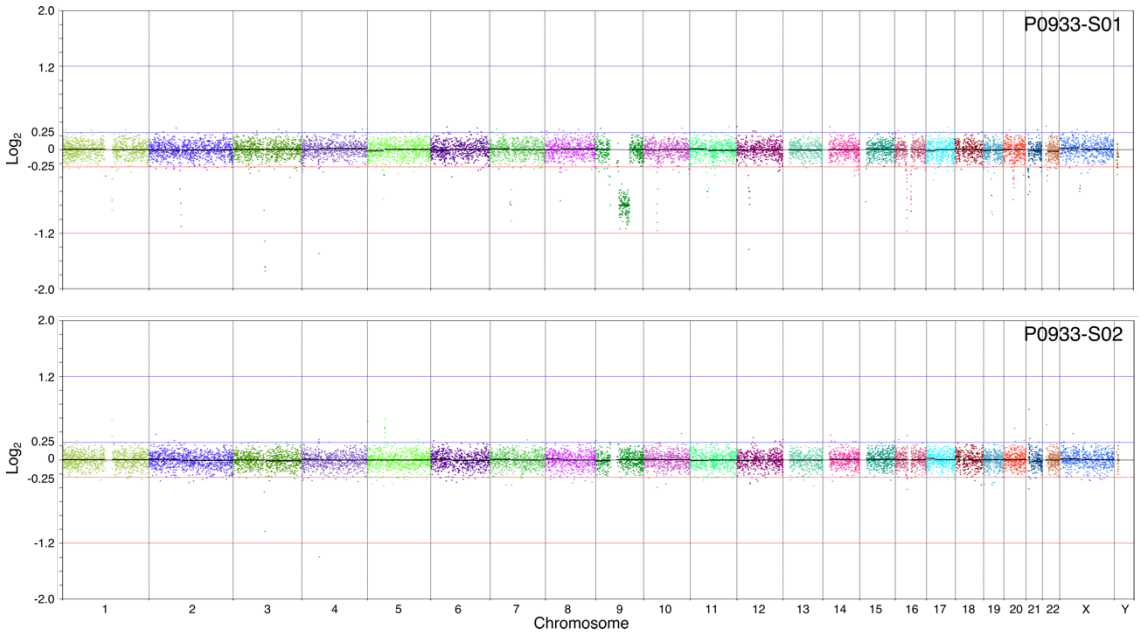
Whole genome copy number plots for tumours from patient P0536.



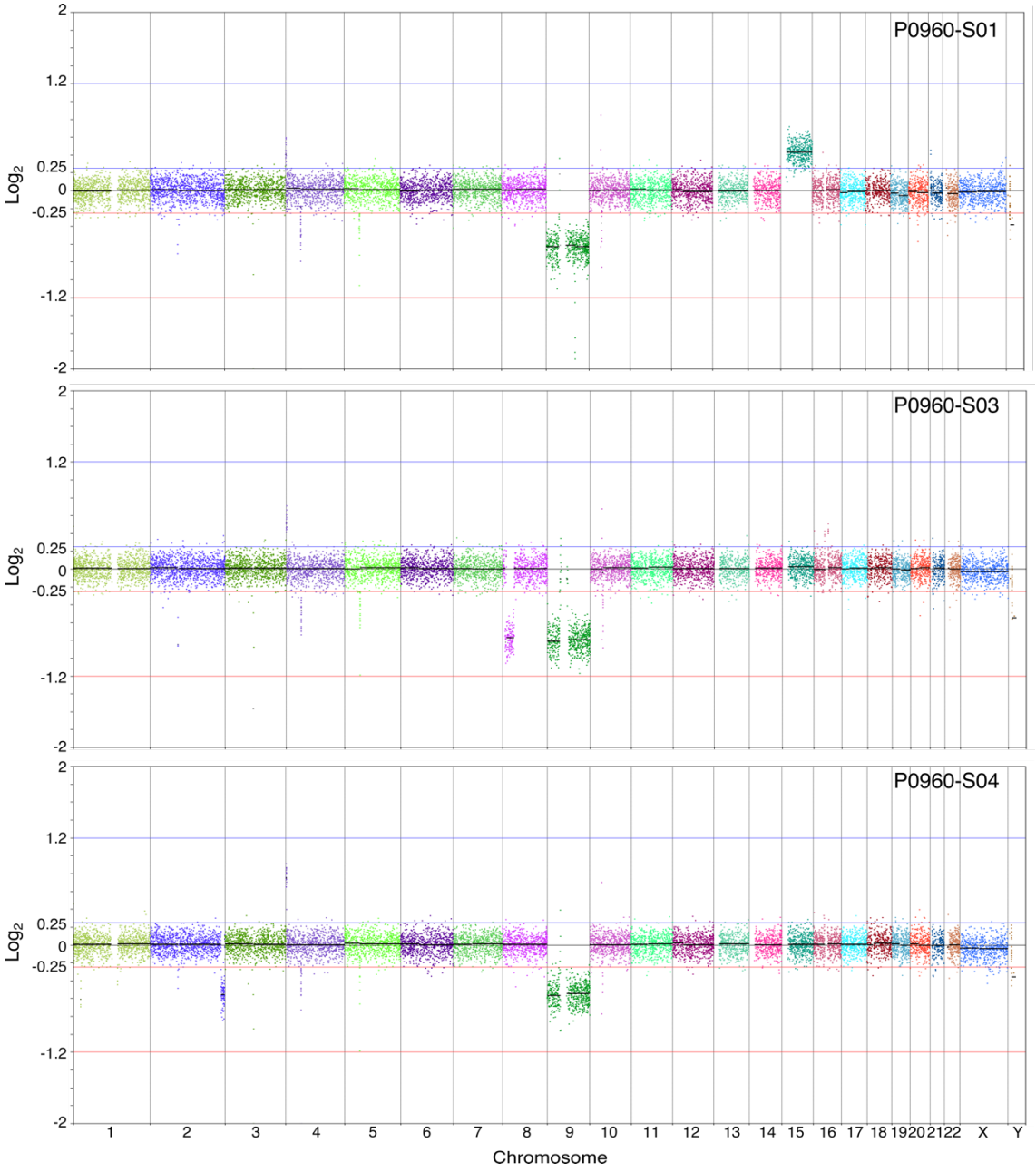
Whole genome copy number plots for tumours from patient P0717.



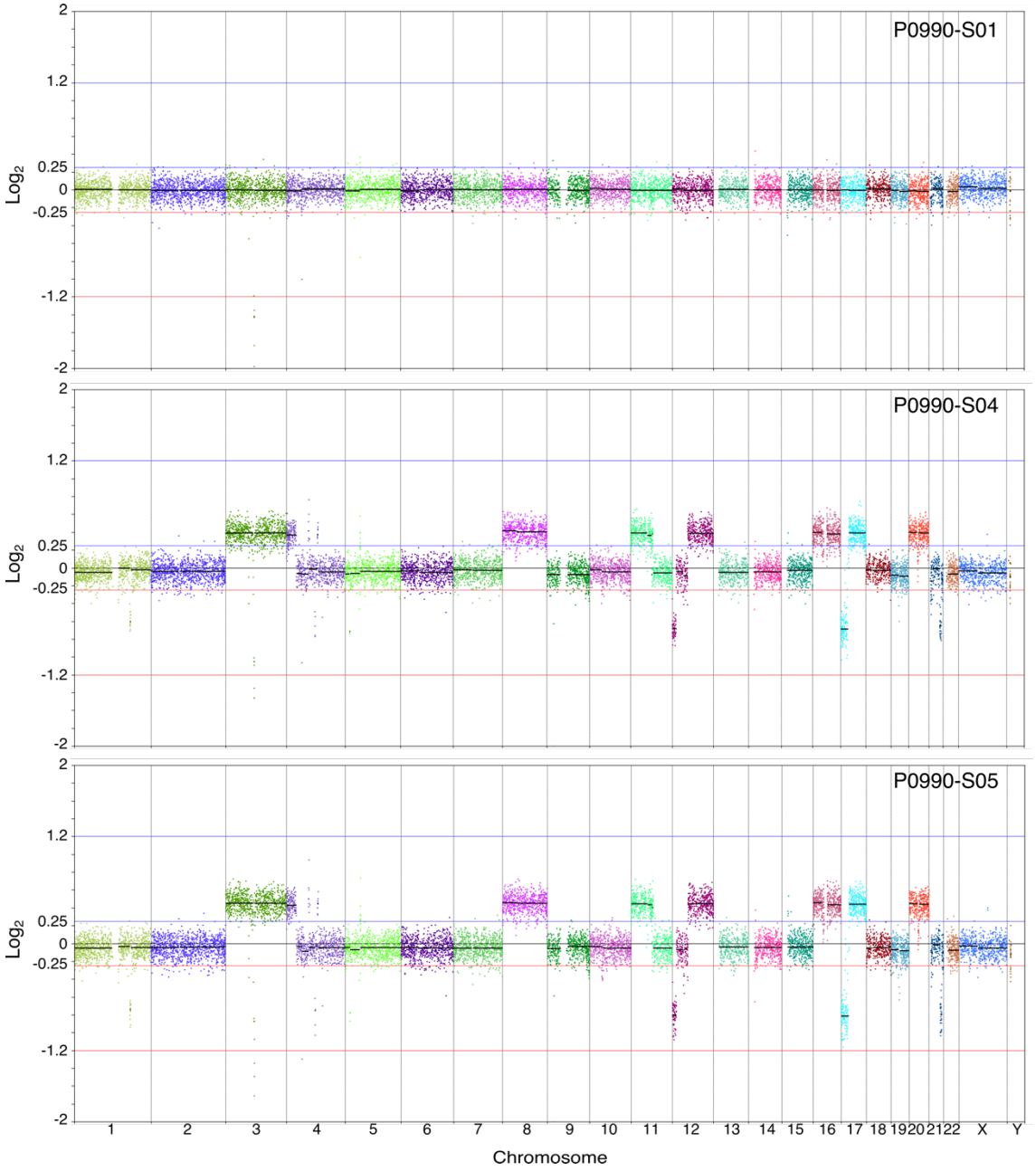
Whole genome copy number plots for tumours from patient P0926



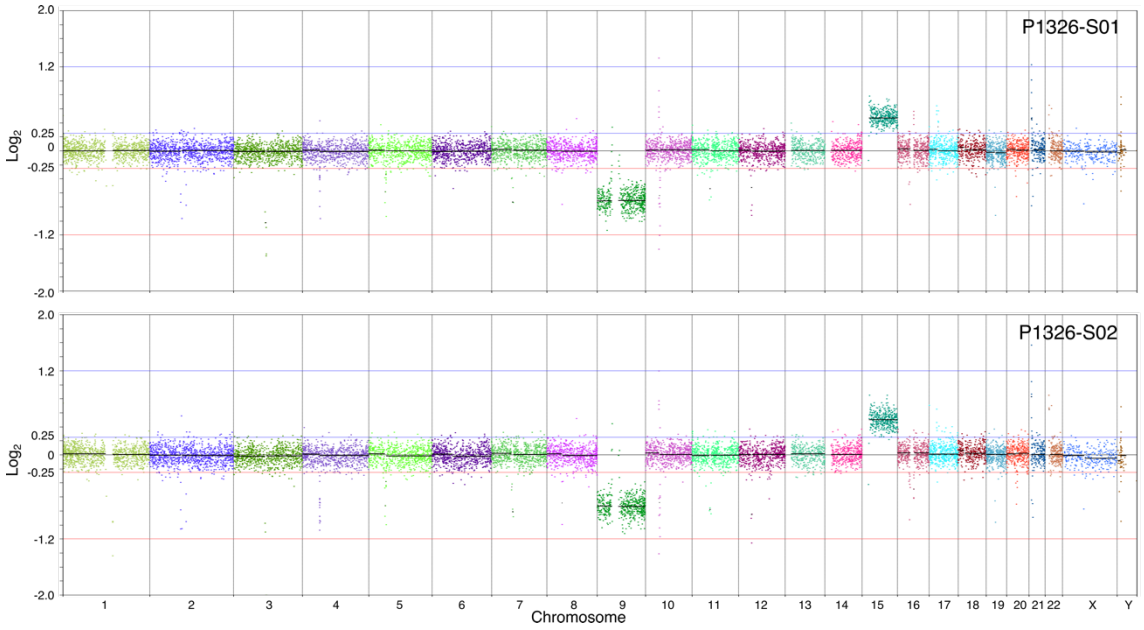
Whole genome copy number plots for tumours from patient P0933.



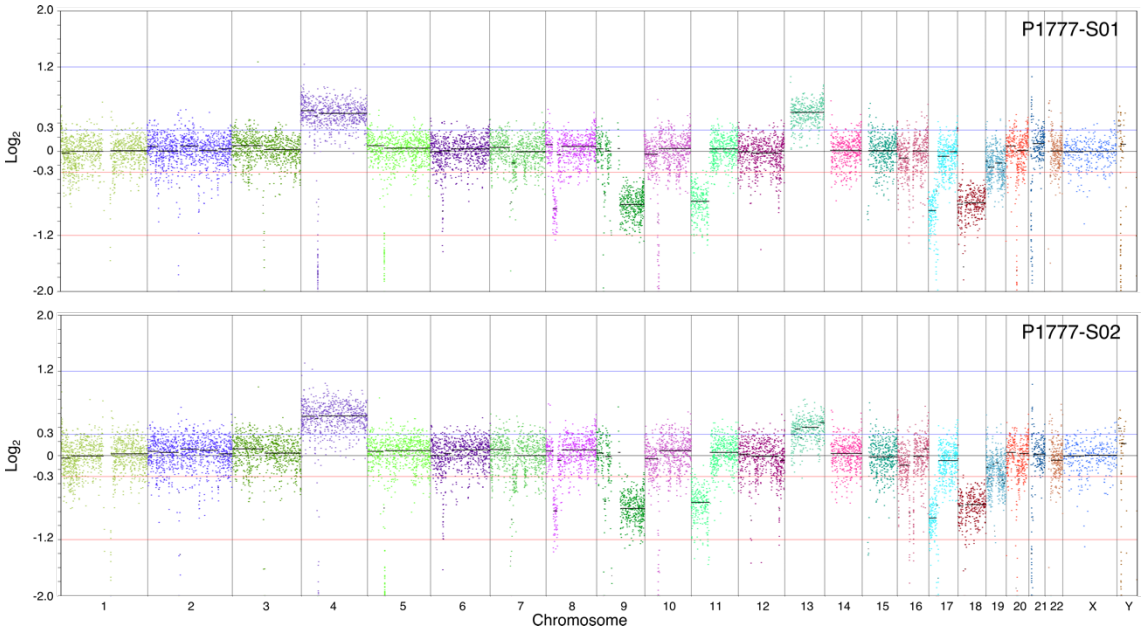
Whole genome copy number plots for tumours from patient P0960.



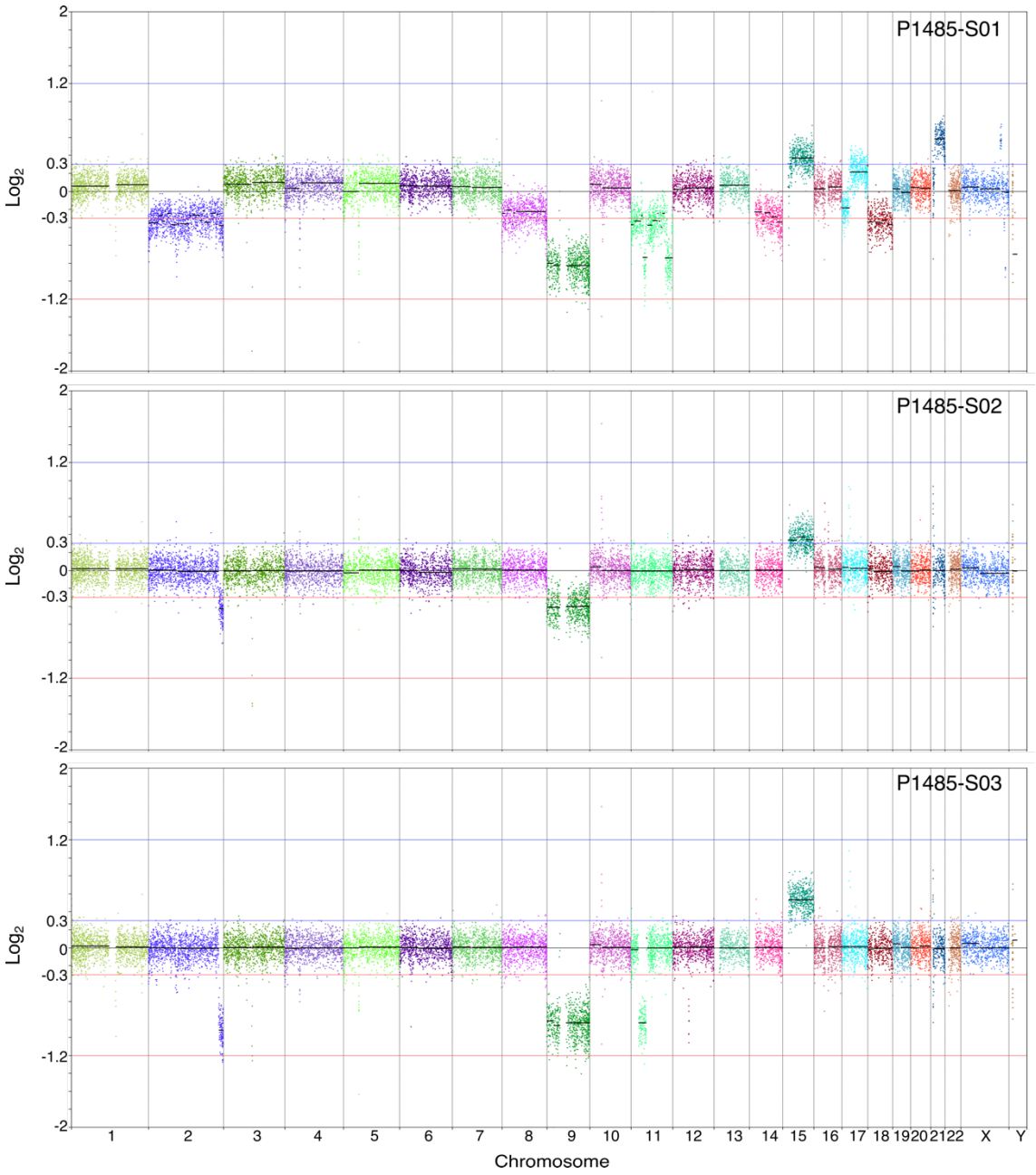
Whole genome copy number plots for tumours from patient P0990.



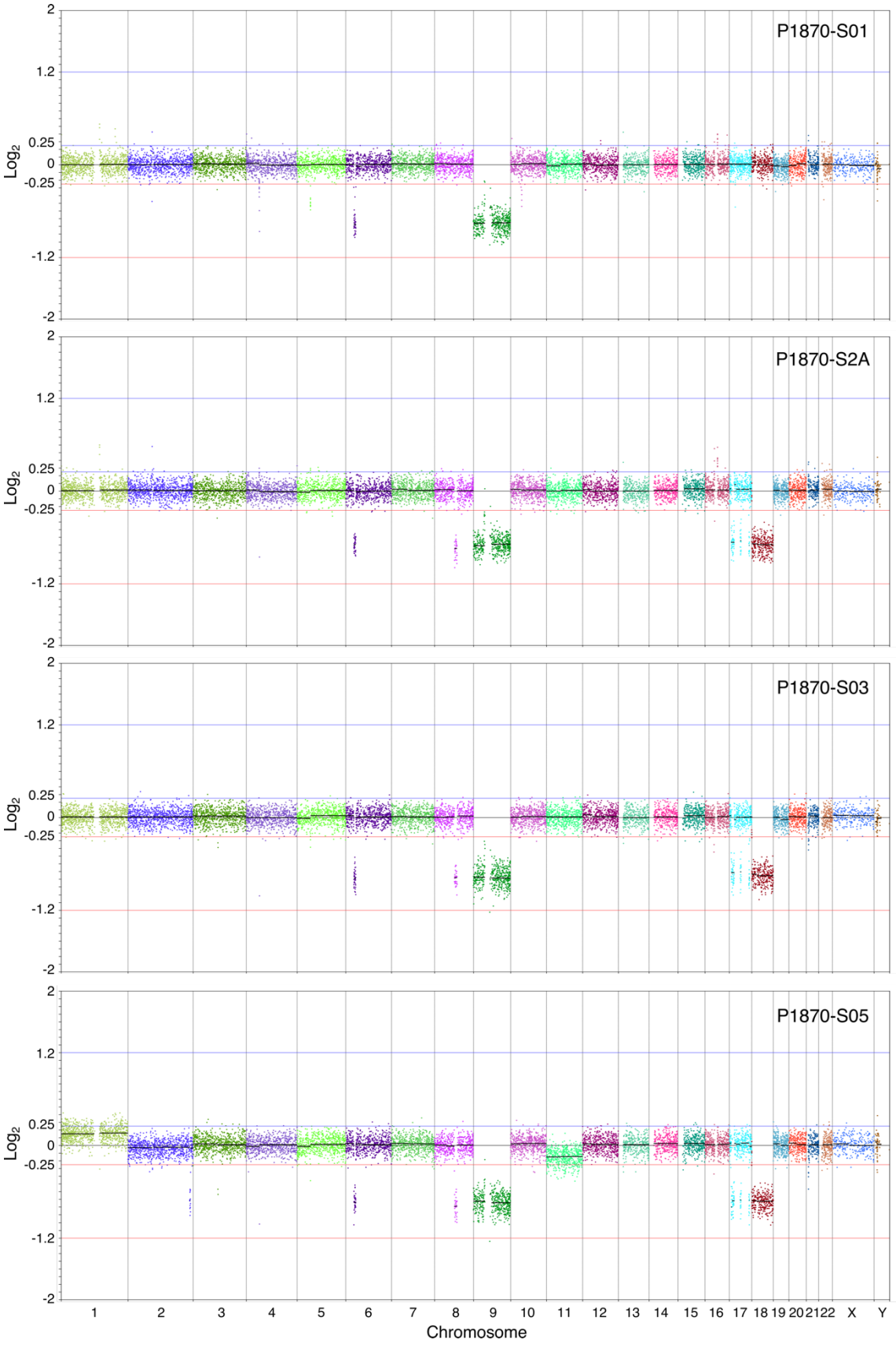
Whole genome copy number plots for tumours from patient P1326 .



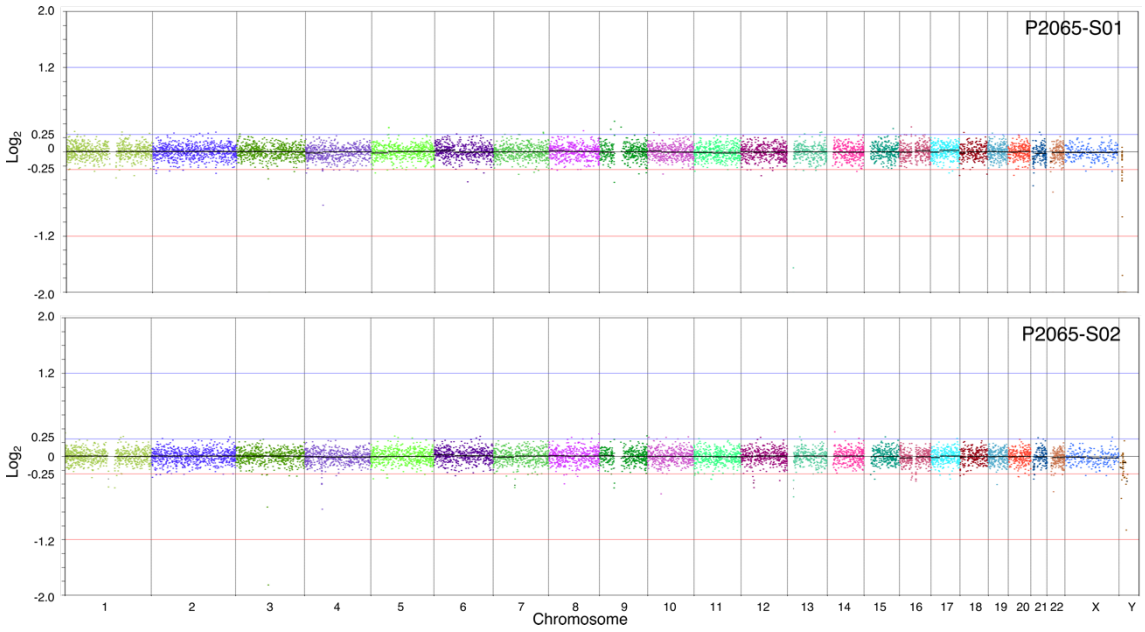
Whole genome copy number plots for tumours from patient P1777.



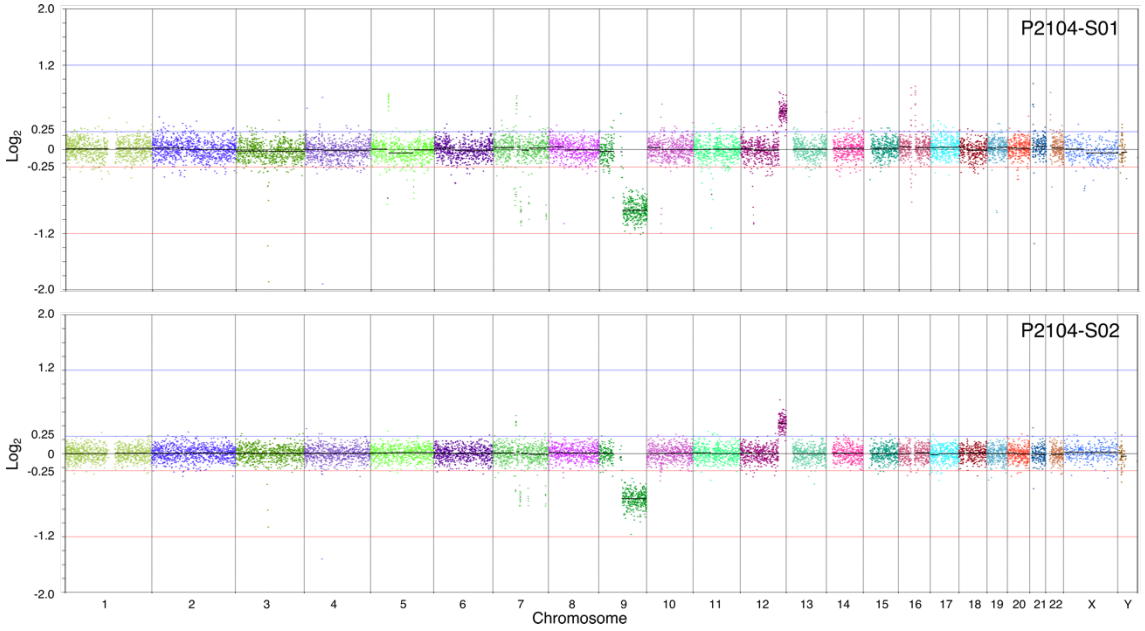
Whole genome copy number plots for tumours from patient 1485.



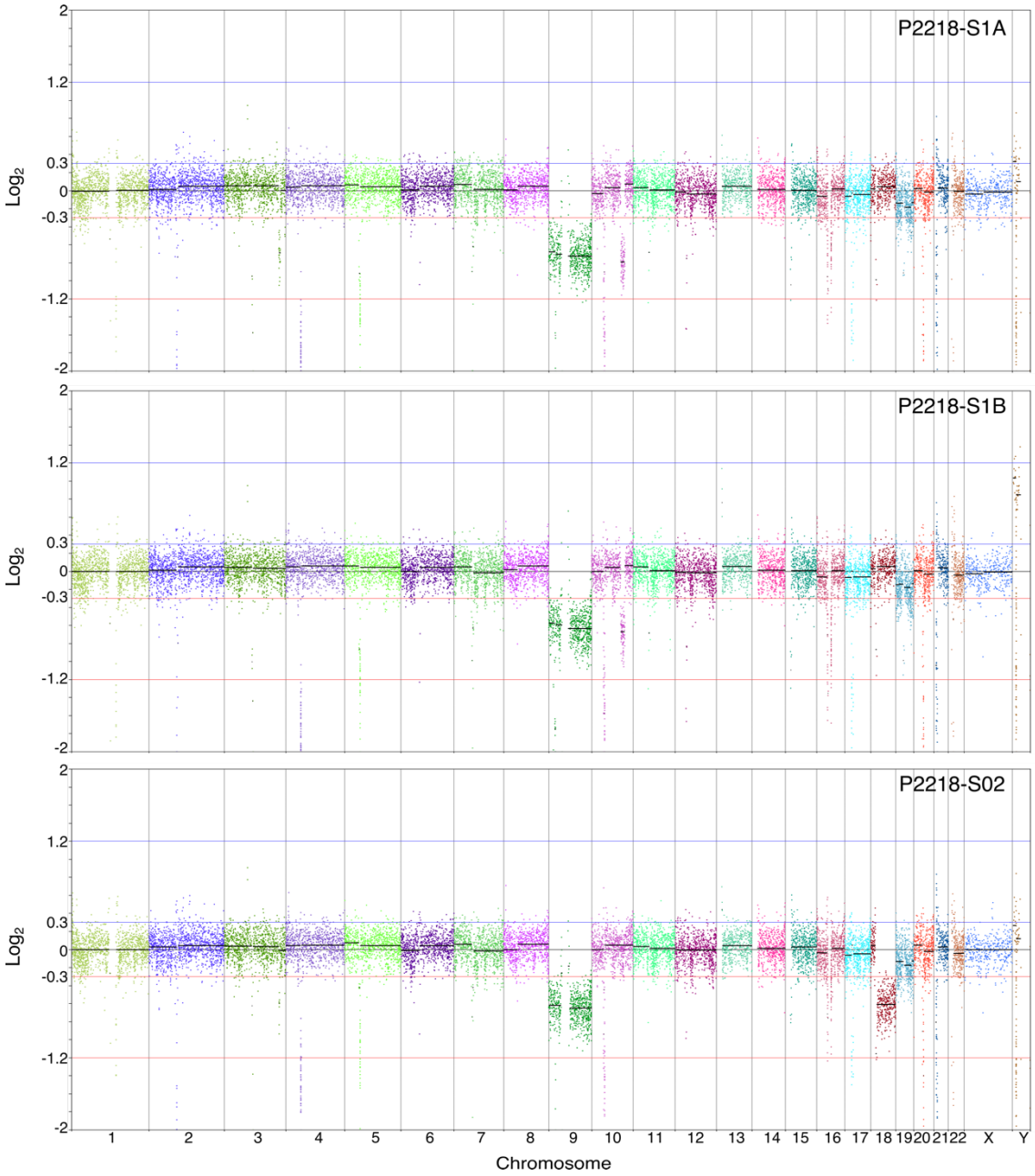
Whole genome copy number plots for tumours from patient P1870.



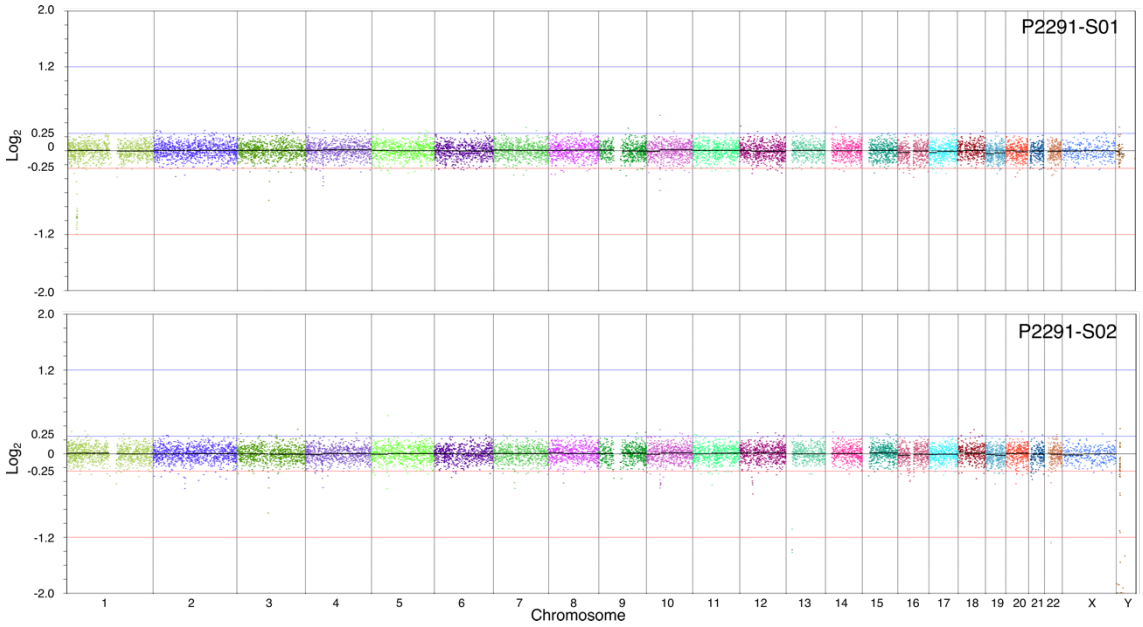
Whole genome copy number plots for tumours from patient P2065



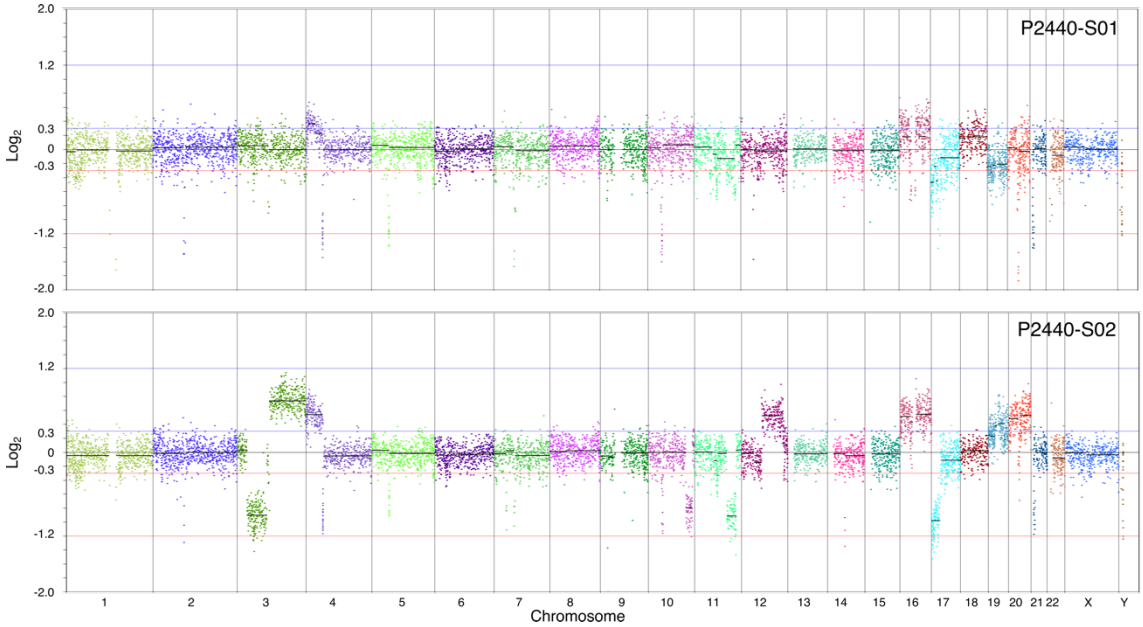
Whole genome copy number plots for tumours from patient P2104.



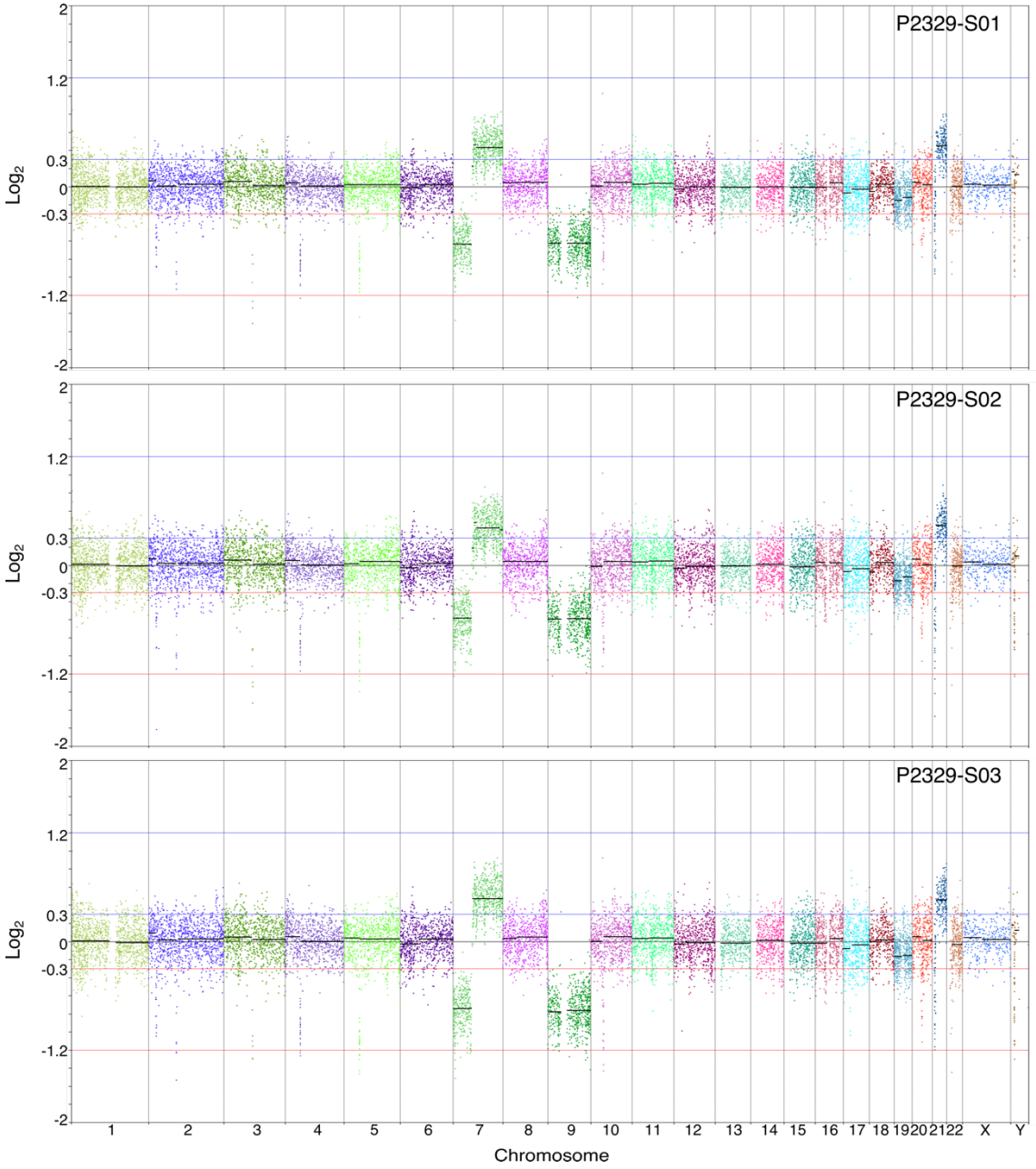
Whole genome copy number plots for tumours from patient P2218.



Whole genome copy number plots for tumours from patient P2291



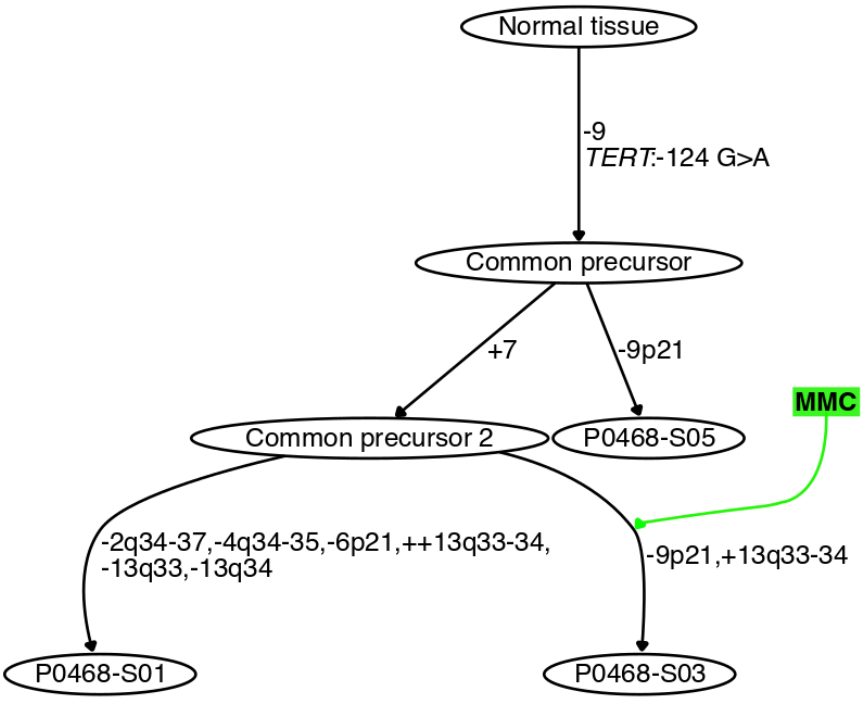
Whole genome copy number plots for tumours from patient P2440.



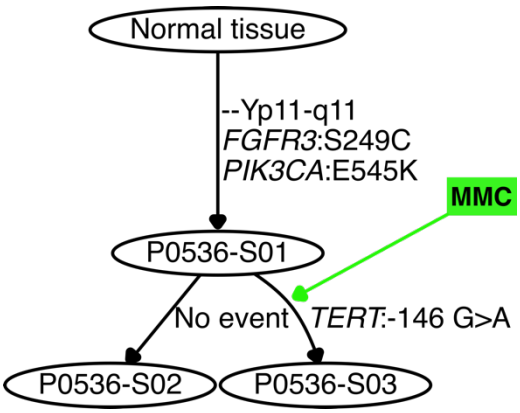
Whole genome copy number plots for tumours from patient P2329.

Appendix G

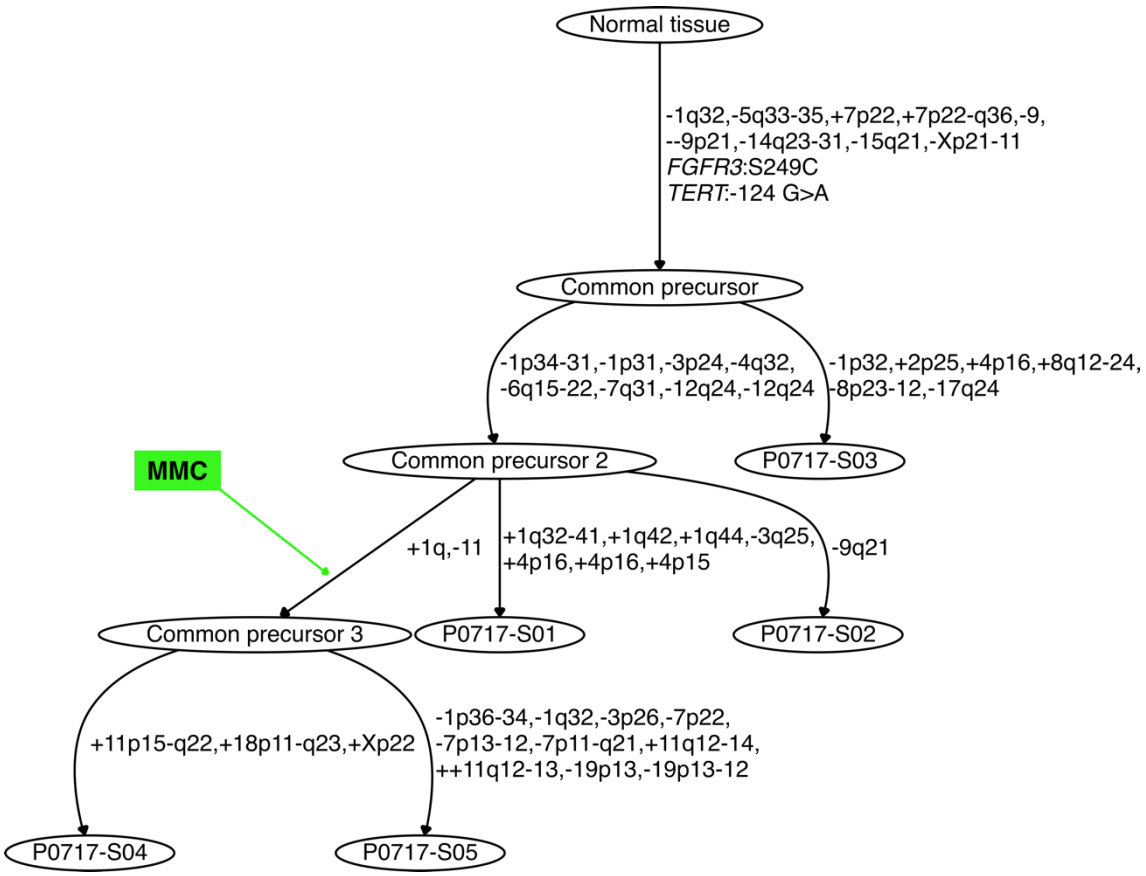
Phylogenetic trees inferred from copy number and mutation status data using TuMult



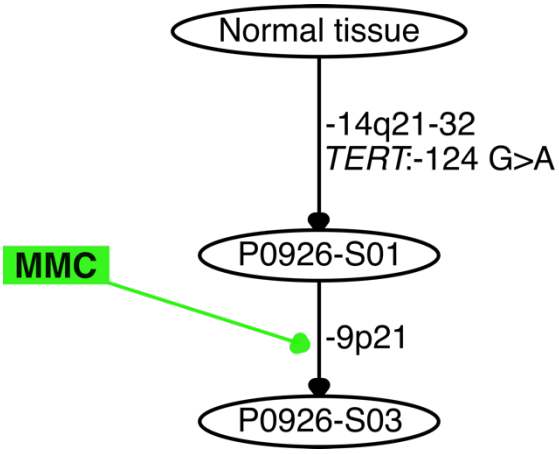
Phylogenetic tree showing the inferred relationships between 3 tumours from patient P0468 based on CNAs and hotspot mutation status.



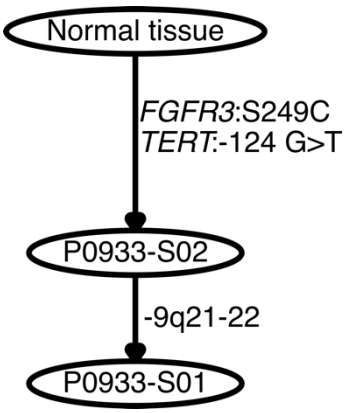
Phylogenetic tree showing the inferred relationships between 3 tumours from patient P0536 based on CNAs and hotspot mutation status.



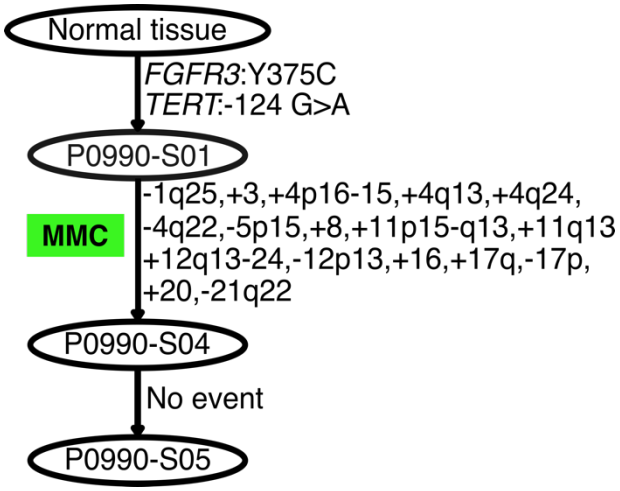
Phylogenetic tree showing the inferred relationships between 5 tumours from patient P0717 based on CNAs and hotspot mutation status.



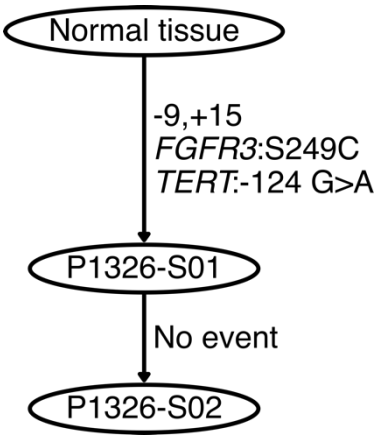
Phylogenetic tree showing the inferred relationship between 2 tumours from patient P0926 based on CNAs and hotspot mutation status.



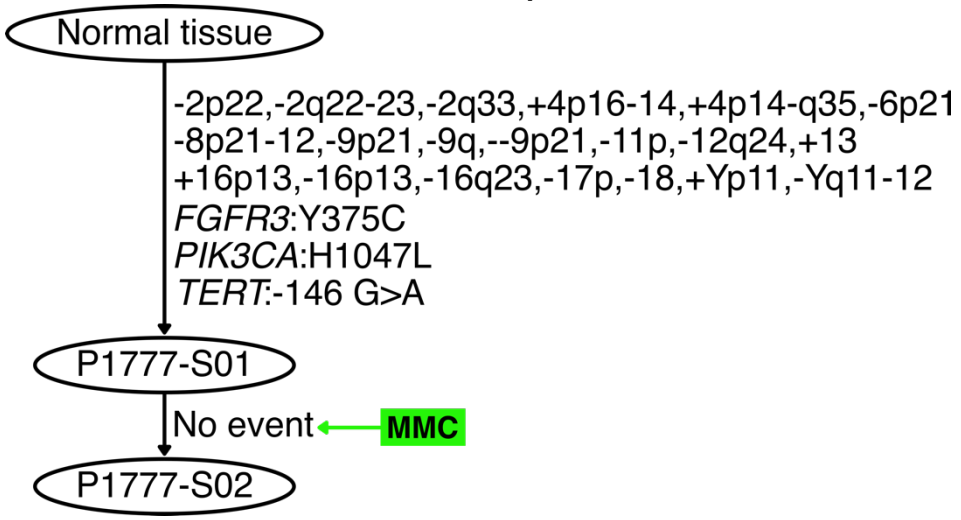
Phylogenetic tree showing the inferred relationship between 2 tumours from patient P0933 based on CNAs and hotspot mutation status.



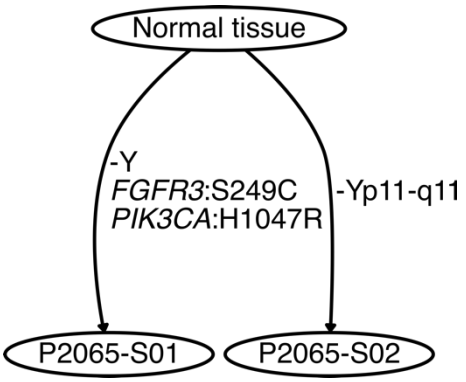
Phylogenetic tree showing the inferred relationships between 3 tumours from patient P0990 based on CNAs and hotspot mutation status.



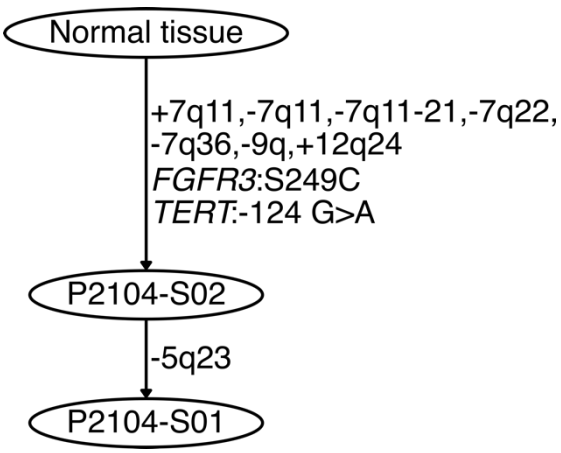
Phylogenetic tree showing the inferred relationship between 2 tumours from patient P1326 based on CNAs and hotspot mutation status.

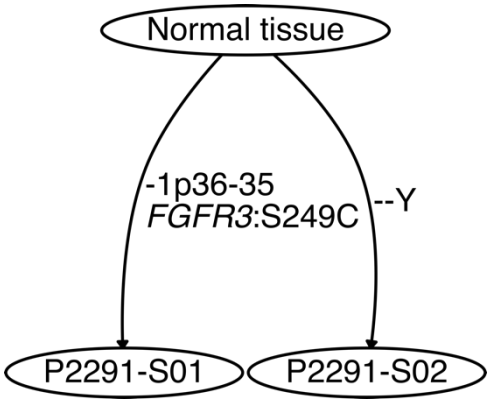


Phylogenetic tree showing the inferred relationship between 2 tumours from patient P1777 based on CNAs and hotspot mutation status.



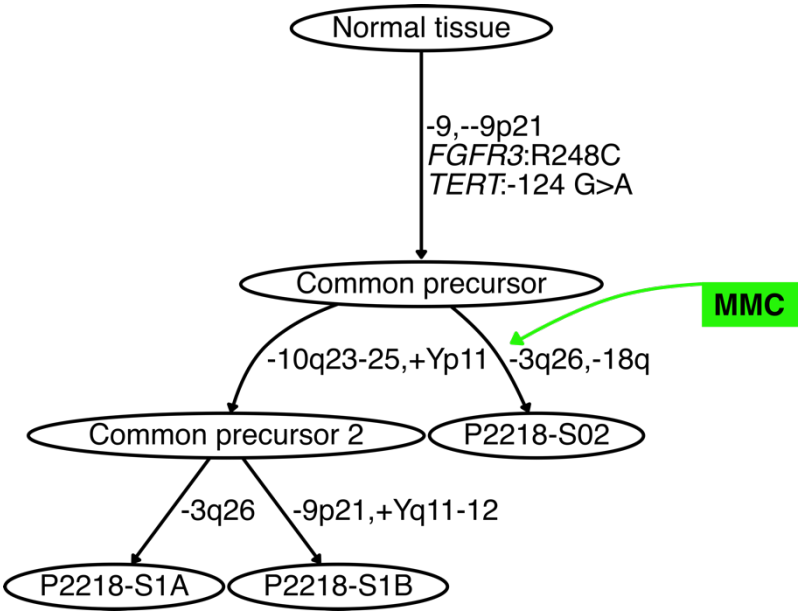
Phylogenetic tree showing the inferred relationship between 2 tumours from patient P2065 based on CNAs and hotspot mutation status.



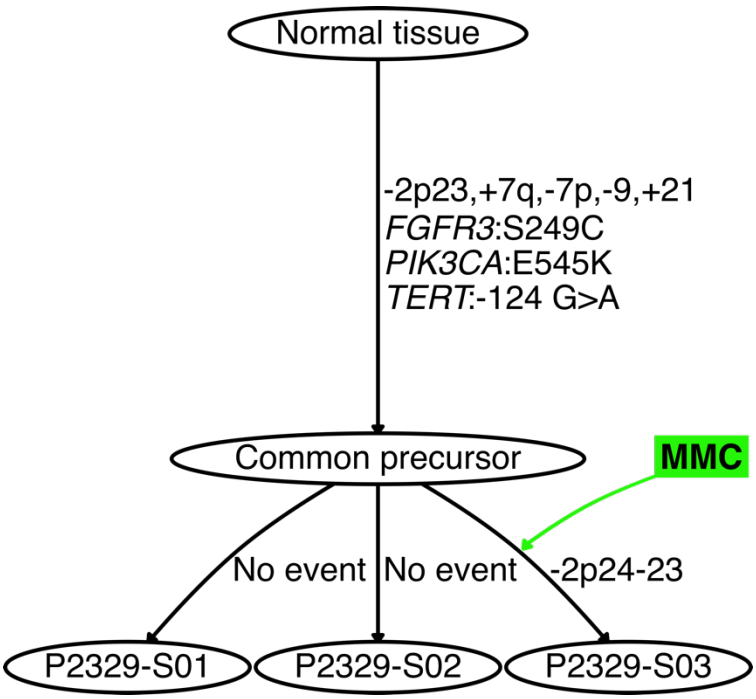


Phylogenetic tree showing the inferred relationship between 2 tumours from patient P2104 based on CNAs and hotspot mutation status.

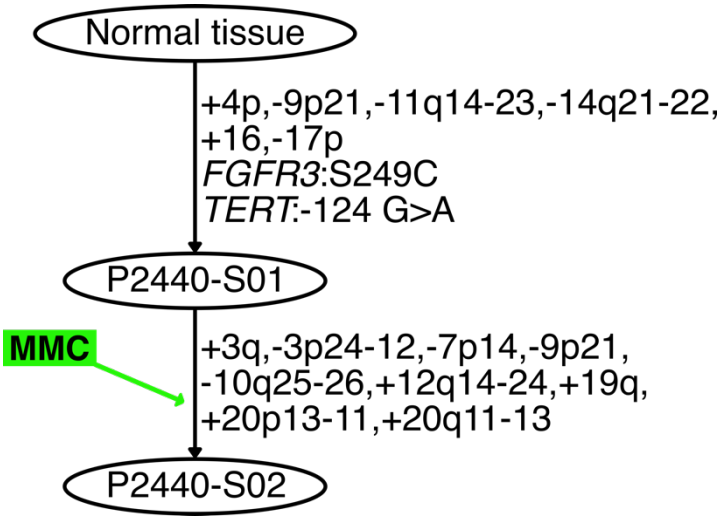
Phylogenetic tree showing the inferred relationship between 2 tumours from patient P2291 based on CNAs and hotspot mutation status.



Phylogenetic tree showing the inferred relationships between 3 tumours from patient P2218 based on CNAs and hotspot mutation status.



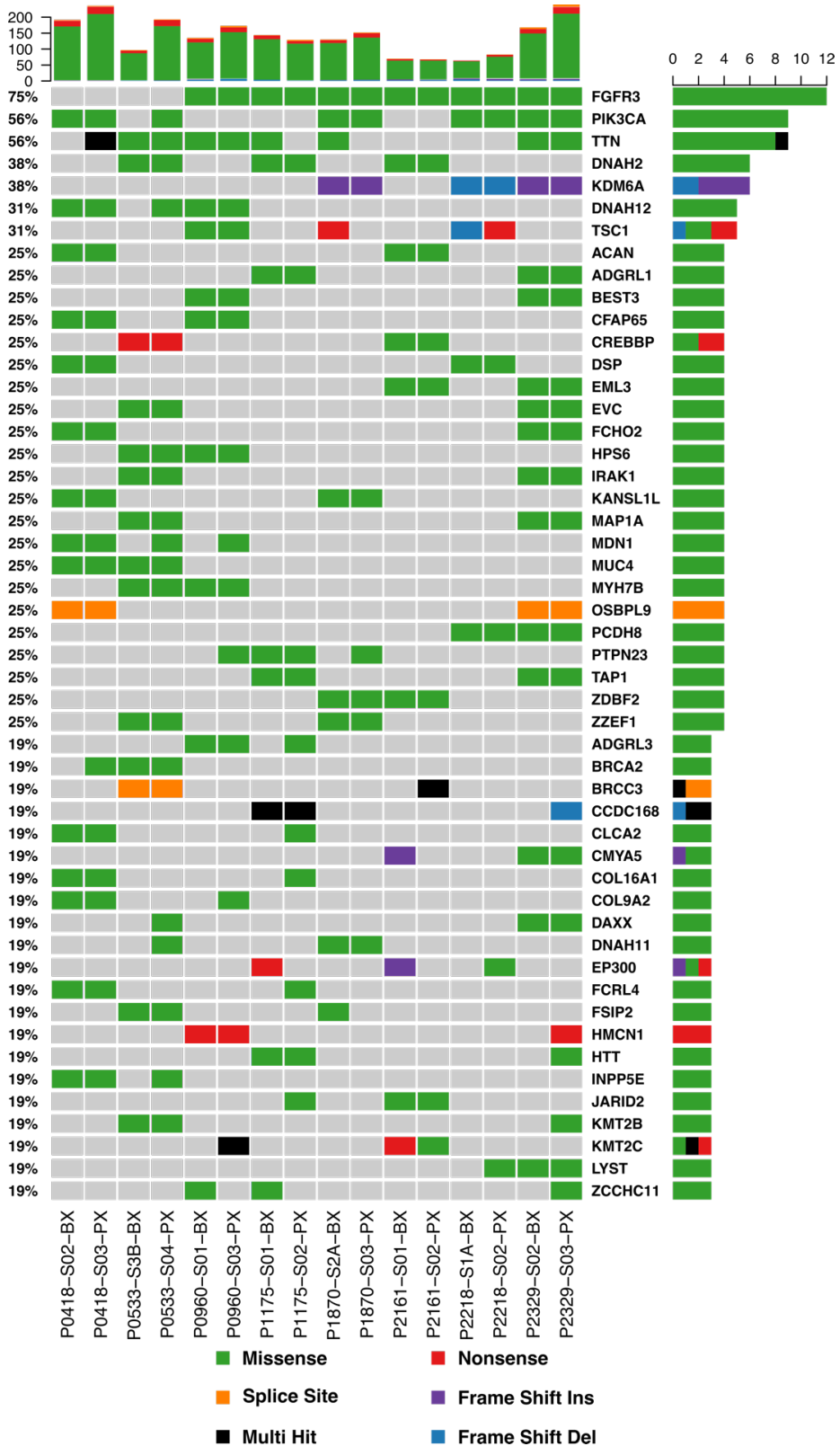
Phylogenetic tree showing the inferred relationships between 3 tumours from patient P2329 based on CNAs and hotspot mutation status.



Phylogenetic tree showing the inferred relationship between 2 tumours from patient P2440 based on CNAs and hotspot mutation status

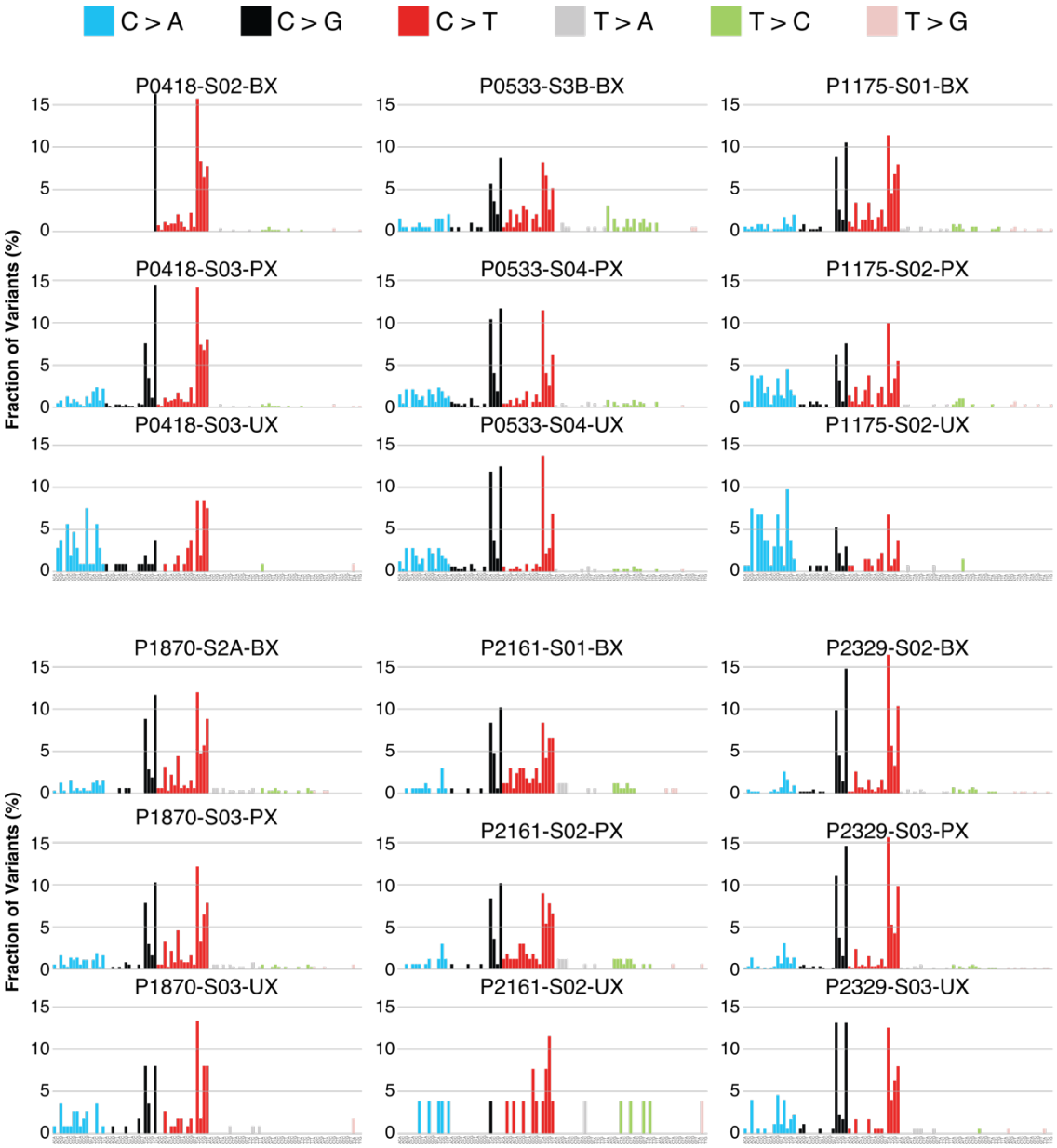
Appendix H

Oncoplot showing the distribution of mutations in the top 50 mutated genes identified by exome sequencing

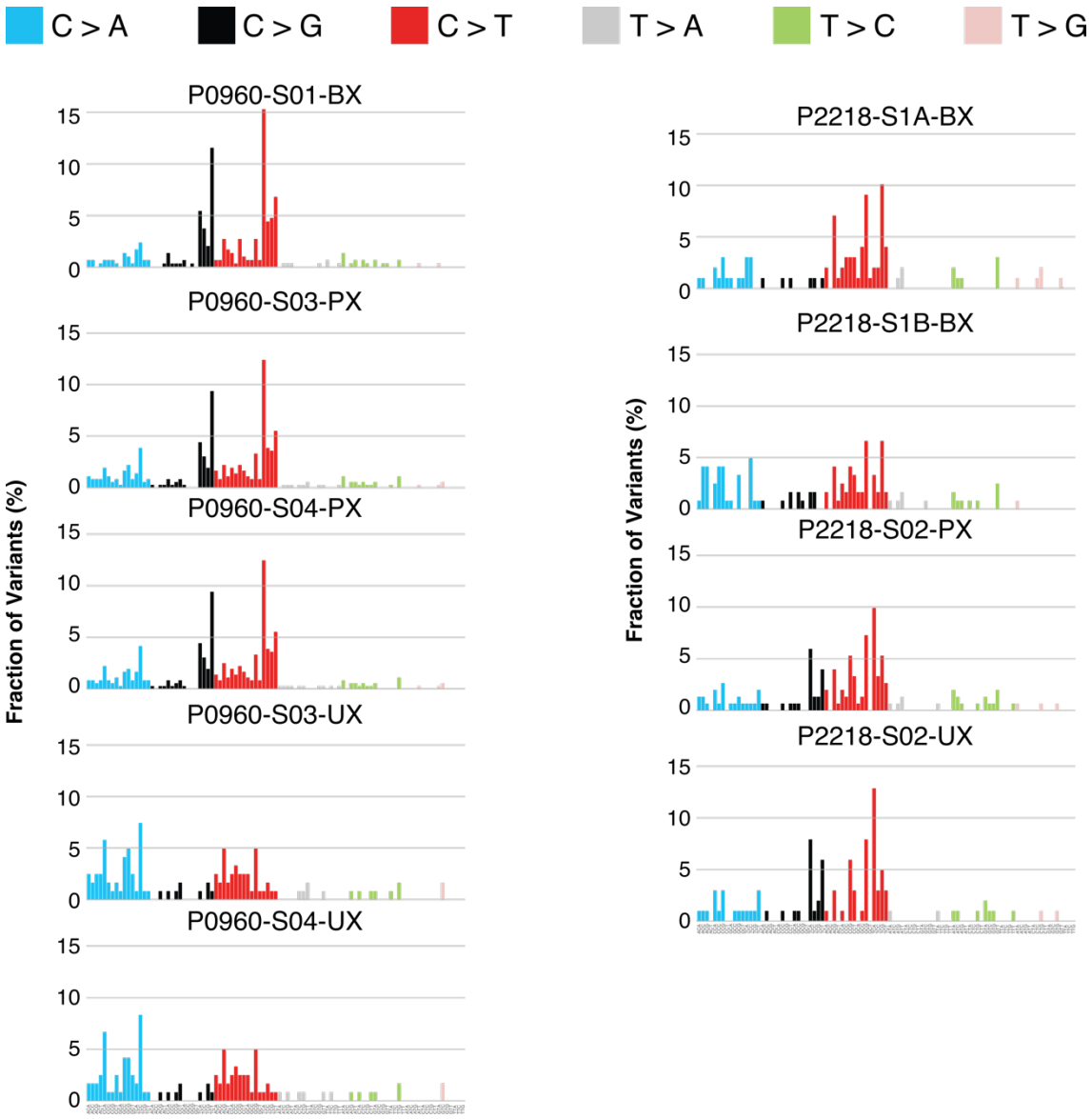


Appendix I

Mutational signature context of pre-MMC, post-MMC and post-MMC unique variants for each patient



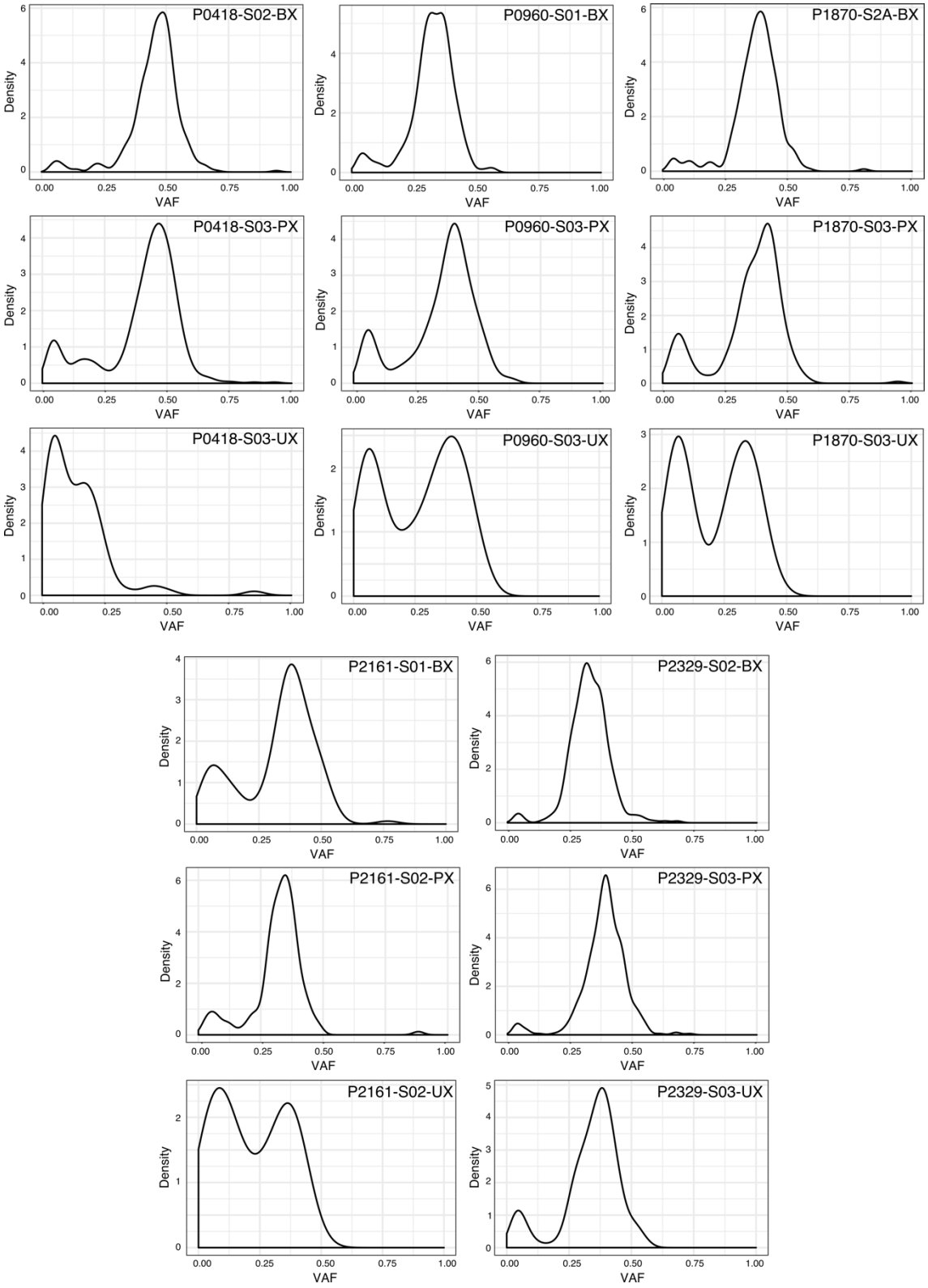
Signature context of SNVs from patients P0418, P0533, P1175, P1870, P2161 and P2329.



Signature context of SNVs from patients P0960 and P2218.

Appendix J

Kernel density plots



Kernel density plots for tumours from patients P0418, P0960, P1870, P2161 and P2329. The VAF (x-axis) is plotted against the density of the VAF (y-axis).

Appendix K
List of suppliers

Agilent Technologies LDA UK Limited, Life Sciences & Chemical Analysis Group, Lakeside, Cheadle Royal Business Park, Stockport, Cheshire SK8 3GR https://www.agilent.com/cs/agilent/en/contact-us/united-kingdom

Beckman Coulter, Oakley Court, Kingsmead Business Park, London Road, High Wycombe, HP11 1JU https://www.beckmancoulter.com/en/about-beckman-coulter/life-sciences

Bio-Rad Laboratories Ltd., The Junction, Station road, Watford, Hertfordshire WD17 1ET http://www.bio-rad.com/

BMG LABTECH Ltd., 8 Bell Business Park, Smeaton Close, Aylesbury, Bucks HP19 8JR https://www.bmglabtech.com/

Covaris Ltd., Unit 3, Brighton Office Campus, Hunns Mere Way, Woodingdean, Brighton, BN2 6AH https://covaris3.corecommerce.com/

GE Healthcare Life Sciences, Amersham Place, Little Chalfont, Buckinghamshire, HP7 9NA https://www.gelifesciences.com/en/gb

Leica Microsystems Ltd., Larch House, Woodlands Business Park, Breckland, Linford Wood, Milton Keynes MK14 6FG https://www.leicabiosystems.com/

New England Biolabs (NEB), 75-77 Knowl Piece, Wilbury Way, Hitchin, Hertfordshire, SG4 0TY https://www.neb.uk.com/

Promega, Delta House, Enterprise Road, Southampton Science Park, Southampton, SO16 7NS https://www.promega.co.uk/

QIAGEN, Skelton House, Lloyd St N, Manchester M15 6SH https://www.qiagen.com/gb/

Severn Biotech Limited, Unit 2, Park Lane, Kidderminster, Worcestershire, DY11 6TJ http://www.severnbiotech.com/

Sigma-Aldrich, The Old Brickyard, New Road, Gillingham, Dorset SP8 4XT https://www.sigmaaldrich.com/united-kingdom.html

Thermo Fisher Scientific, Fisher Scientific UK Ltd, Bishop Meadow Road, Loughborough LE11 5RG https://www.thermofisher.com/uk/en/home.html

VWR International, Hunter Boulevard, Magna Park, Lutterworth, Leicestershire LE17 4XN https://uk.vwr.com/store/

References

1. Antoni, S. *et al.* Bladder Cancer Incidence and Mortality: A Global Overview and Recent Trends. *Eur. Urol.* **71,** 96–108 (2017).
2. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **136,** E359–86 (2015).
3. Burger, M. *et al.* Epidemiology and Risk Factors of Urothelial Bladder Cancer. *Eur. Urol.* **63,** 234–241 (2013).
4. Knowles, M. A. & Hurst, C. D. Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity. *Nat. Rev. Cancer* **15,** 25–41 (2015).
5. Harris, A. L. & Neal, D. E. Bladder cancer--field versus clonal origin. *N. Engl. J. Med.* **326,** 759–761 (1992).
6. Montie, J. E. *et al.* Bladder Cancer. *J Natl Compr Canc Netw* **7,** 8–39 (2009).
7. Sobin, L. H., Gospodarowicz, M. K. & Wittekind, C. TNM Classification of Malignant Tumours. (2009).
8. WHO. *Histological Typing of Urinary Bladder Tumours. International Histological Classification of Tumours 10*. (1973).
9. Eble, J. N., Sauter, G., Epstein, J. I. & Sesterhennm, I. A. *World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of the Urinary System and Male Genital Organs*. 1–354 (IARC Press, 2004).
10. Woldu, S. L., Bagrodia, A. & Lotan, Y. Guideline of guidelines: non-muscle-invasive bladder cancer. *BJU Int* **119,** 371–380 (2017).
11. Dalbagni, G. *et al.* Clinical Outcome in a Contemporary Series of Restaged Patients with Clinical T1 Bladder Cancer. *Eur. Urol.* **56,** 903–910 (2009).
12. Stein, J. P. *et al.* Radical Cystectomy in the Treatment of Invasive Bladder Cancer: Long-Term Results in 1,054 Patients. *J. Clin. Oncol.* **19,** 666–675 (2001).
13. Aggen, D. H. & Drake, C. G. Biomarkers for immunotherapy in bladder cancer: a moving target. 1–13 (2017). doi:10.1186/s40425-017-0299-1
14. Fujii, Y. Prediction models for progression of non-muscle-invasive bladder cancer: A review. *Int J Urol* **24,** 730–7 (2017).
15. Svatek, R. S. *et al.* The Economics of Bladder Cancer: Costs and Considerations of Caring for This Disease. *Eur. Urol.* **66,** 253–262 (2014).
16. Gerlinger, M. & Swanton, C. How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. *British Journal of Cancer* **103,** 1139–1143 (2010).
17. Nowell, P. C. Clonal Evolution of Tumor-Cell Populations. *Science* **194,** 23–28 (1976).
18. Pearce, T. Convergence and Parallelism in Evolution: A Neo-Gouldian Account. *Br J Philos Sci* **63,** axr046–448 (2011).
19. Powell, R. Convergent evolution and the limits of natural selection. *Euro Jnl Phil Sci* **2,** 355–373 (2012).
20. McGranahan, N. & Swanton, C. Biological and Therapeutic Impact of Intratumor Heterogeneity in Cancer Evolution. *Cancer Cell* **27,** 15–26 (2015).
21. Gould, S. J. *Wonderful Life: The Burgess Shale and the Nature of History*. (W. W. Norton, 1990).
22. Denzinger, S. *et al.* Improved clonality analysis of multifocal bladder tumors by combination of histopathologic organ mapping, loss of heterozygosity, fluorescence in situ hybridization, and p53 analyses. *Hum. Pathol.* **37,** 143–151 (2006).

23. Vollebergh, M. A. *et al.* Lack of genomic heterogeneity at high-resolution aCGH between primary breast cancers and their paired lymph node metastases. *PLoS ONE* **9,** e103177 (2014).
24. Coons, S. W., Johnson, P. C. & Shapiro, J. R. Cytogenetic and Flow-Cytometry Dna Analysis of Regional Heterogeneity in a Low-Grade Human Glioma. *Cancer Res.* **55,** 1569–1577 (1995).
25. Teixeira, M. R. *et al.* Clonal Heterogeneity in Breast-Cancer - Karyotypic Comparisons of Multiple Intra-Tumorous and Extra-Tumorous Samples From 3 Patients. *Int. J. Cancer* **63,** 63–68 (1995).
26. Louhelainen, J., Wijkstrom, H. & Hemminki, K. Allelic losses demonstrate monoclonality of multifocal bladder tumors. *Int. J. Cancer* **87,** 522–527 (2000).
27. Tomlinson, I. P. M., Lambros, M. B. K. & Roylance, R. R. Loss of heterozygosity analysis: practically and conceptually flawed? *Genes Chromosomes Cancer* **34,** 349–353 (2002).
28. Marusyk, A. & Polyak, K. Tumor heterogeneity: Causes and consequences. *Biochim. Biophys. Acta* **1805,** 105–117 (2010).
29. Simon, R. *et al.* Cytogenetic analysis of multifocal bladder cancer supports a monoclonal origin and intraepithelial spread of tumor cells. *Cancer Res.* **61,** 355–362 (2001).
30. O'Connor, C. Karyotyping for chromosomal abnormalities. *Nature Education* **1,** 27 (2008).
31. Farabegoli, F. *et al.* Clone heterogeneity in diploid and aneuploid breast carcinomas as detected by FISH. *Cytometry* **46,** 50–56 (2001).
32. Abat, D. *et al.* Genetic alterations of chromosomes, p53 and p16 genes in low- and high-grade bladder cancer. *Oncol Lett* **8,** 25–32 (2014).
33. Fauth, E., Scherthan, H. & Zankl, H. Chromosome painting reveals specific patterns of chromosome occurrence in mitomycin C- and diethylstilboestrol-induced micronuclei. *Mutagenesis* **15,** 459–467 (2000).
34. Pinkel, D., Straume, T. & Gray, J. W. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc. Natl. Acad. Sci. U.S.A.* **83,** 2934–2938 (1986).
35. Letouzé, E., Allory, Y., Bollet, M. A., Radvanyi, F. & Guyon, F. Analysis of the copy number profiles of several tumor samples from the same patient reveals the successive steps in tumorigenesis. *Genome Biol.* **11,** (2010).
36. Weiss, M. M. *et al.* Comparative genomic hybridisation. *MP, Mol. Pathol.* **52,** 243–251 (1999).
37. Knowles, M. A. Tumor suppressor loci in bladder cancer. *Front Biosci* **12,** 2233–2251 (2007).
38. Pinkel, D. *et al.* High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.* **20,** 207–5 (1998).
39. Hartmann, L. *et al.* Detection of clonal evolution in hematopoietic malignancies by combining comparative genomic hybridization and single nucleotide polymorphism arrays. *Clin. Chem.* **60,** 1558–1568 (2014).
40. Hiley, C. T. & Swanton, C. Spatial and temporal cancer evolution: causes and consequences of tumour diversity. *Clin Med* **14 Suppl 6,** s33–7 (2014).
41. National Human Genome Research Institute. The Cost of Sequencing a Human Genome. *https://www.genome.gov* (2016). Available at: https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/. (Accessed: 23rd December 2018)
42. Fisher, R., Pusztai, L. & Swanton, C. Cancer heterogeneity: implications for targeted therapeutics. *British Journal of Cancer* **108,** 479–485 (2013).
43. Yau, C. *et al.* Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173,** 291–304.e6 (2018).

44. Bailey, M. H. *et al.* Perspective on Oncogenic Processes at the End of the Beginning of Cancer Genomics. *Cell* **173,** 305–320.e10 (2018).
45. Mina, M. *et al.* Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **173,** 321–337.e10 (2018).
46. Huang, K.-L. *et al.* Pathogenic Germline Variants in 10,389 Adult Cancers. *Cell* **173,** 355–370.e14 (2018).
47. Bailey, M. H. *et al.* Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **173,** 371–376.e18 (2018).
48. Hudson Chairperson, T. J. *et al.* International network of cancer genome projects. *Nature* **464,** 993–998 (2010).
49. NIH, TCGA, NCINHGRI. The Cancer Genome Atlas (TCGA): The next stage. *NIH The Cancer Genome Atlas* (2015).
50. Turner, N. C. & Reis-Filho, J. S. Genetic heterogeneity and cancer drug resistance. *Lancet Oncol.* **13,** e178–85 (2012).
51. Walter, M. J. *et al.* Clonal architecture of secondary acute myeloid leukemia. *N. Engl. J. Med.* **366,** 1090–1098 (2012).
52. Navin, N. E. Cancer genomics: one cell at a time. *Genome Biol.* **15,** 452 (2014).
53. Gerlinger, M. *et al.* Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.* **366,** 883–892 (2012).
54. Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472,** 90–U119 (2011).
55. Thomsen, M. B. H. *et al.* Comprehensive multiregional analysis of molecular heterogeneity in bladder cancer. *Scientific Reports* 1–9 (2017). doi:10.1038/s41598-017-11291-0
56. de Bruin, E. C. *et al.* Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* **346,** 251–256 (2014).
57. Gerlinger, M. *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* **46,** 225–+ (2014).
58. Gundem, G. *et al.* The evolutionary history of lethal metastatic prostate cancer. *Nature* **520,** 353–357 (2015).
59. Dentro, S. C. *et al.* Portraits of genetic intra-tumour heterogeneity and subclonal selection across cancer types. *bioRxiv* 1–53 (2018). doi:10.1101/312041
60. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481,** 506–510 (2012).
61. Lamy, P. *et al.* Paired Exome Analysis Reveals Clonal Evolution and Potential Therapeutic Targets in Urothelial Carcinoma. *Cancer Res.* **76,** 5894–5906 (2016).
62. Faltas, B. M. *et al.* Clonal evolution of chemotherapy-resistant urothelial carcinoma. *Nat. Genet.* **48,** 1490–1499 (2016).
63. Thomsen, M. B. H. *et al.* Spatial and temporal clonal evolution during development of metastatic urothelial carcinoma. *Molecular Oncology* **10,** 1450–1460 (2016).
64. Brastianos, P. K. *et al.* Genomic Characterization of Brain Metastases Reveals Branched Evolution and Potential Therapeutic Targets. *Cancer Discov* **5,** 1164–1177 (2015).
65. Rosenthal, R., McGranahan, N., Herrero, J. & Swanton, C. Deciphering Genetic Intratumor Heterogeneity and Its Impact on Cancer Evolution. *Annu. Rev. Cancer Biol.* **1,** 223–240 (2017).
66. Loeb, L. A. Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res.* **51,** 3075–3079 (1991).
67. Tomlinson, I. P., Novelli, M. R. & Bodmer, W. F. The mutation rate and cancer. *Proc. Natl. Acad. Sci. U.S.A.* **93,** 14800–14803 (1996).

68. Negrini, S., Gorgoulis, V. G. & Halazonetis, T. D. Genomic instability — an evolving hallmark of cancer. *Nat Rev Mol Cell Biol* **11,** 220–228 (2010).
69. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502,** 333–339 (2013).
70. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500,** 415–421 (2013).
71. Shen, Z. Genomic instability and cancer: an introduction. *Journal of Molecular Cell Biology* **3,** 1–3 (2011).
72. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458,** 719–724 (2009).
73. Helleday, T., Eshtad, S. & Nik-Zainal, S. A. Mechanisms underlying mutational signatures in human cancers. *Nature Publishing Group* **15,** 585–598 (2014).
74. Alexandrov, L. B. & Stratton, M. R. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Curr. Opin. Genet. Dev.* **24,** 52–60 (2014).
75. Cahill, D. P., Kinzler, K. W., Vogelstein, B. & Lengauer, C. Genetic instability and darwinian selection in tumours. *Trends Cell Biol.* **9,** M57–60 (1999).
76. Gerlinger, M. *et al.* Cancer: evolution within a lifetime. *Annu. Rev. Genet.* **48,** 215–236 (2014).
77. Gillies, R. J., Verduzco, D. & Gatenby, R. A. Evolutionary dynamics of carcinogenesis and why targeted therapy does not work. *Nat. Rev. Cancer* **12,** 487–493 (2012).
78. Smith, H. C., Bennett, R. P., Kizilyer, A., McDougall, W. M. & Prohaska, K. M. Functions and regulation of the APOBEC family of proteins. *Seminars in Cell & Developmental Biology* **23,** 258–268 (2012).
79. Burns, M. B. *et al.* APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **494,** 366–370 (2013).
80. Roberts, S. A. *et al.* An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.* **45,** 970–976 (2013).
81. Robertson, A. G. *et al.* Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. *Cell* **171,** 540–546.e25 (2017).
82. Hurst, C. D. *et al.* Genomic Subtypes of Non-invasive Bladder Cancer with Distinct Metabolic Profile and Female Gender Bias in KDM6A Mutation Frequency. *Cancer Cell* **32,** 701–715.e7 (2017).
83. Henderson, S., Chakravarthy, A., Su, X., Boshoff, C. & Fenton, T. R. APOBEC-Mediated Cytosine Deamination Links PIK3CA Helical Domain Mutations to Human Papillomavirus-Driven Tumor Development. *Cell Rep* **7,** 1833–1841 (2014).
84. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396,** 643–649 (1998).
85. Holland, A. J. & Cleveland, D. W. Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. *Nat Rev Mol Cell Biol* **10,** 478–487 (2009).
86. Taylor, A. M. *et al.* Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell* **33,** 676–689.e3 (2018).
87. Ciriello, G. *et al.* Emerging landscape of oncogenic signatures across human cancers. *Nature Publishing Group* 1–9 (2013). doi:10.1038/ng.2762
88. Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144,** 27–40 (2011).
89. Rajapakse, I., Scalzo, D. & Groudine, M. Losing control: cancer's catastrophic transition. *Nucleus* **2,** 249–252 (2011).
90. Rudolph, K. L., Millard, M., Bosenberg, M. W. & DePinho, R. A. Telomere dysfunction and evolution of intestinal carcinoma in mice and humans. *Nat. Genet.* **28,** 155–159 (2001).

91. Nik-Zainal, S. A. *et al.* The Life History of 21 Breast Cancers. *Cell* **149,** – 1007 (2012).
92. Bolli, N. *et al.* Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nature Communications* **5,** 2997 (2014).
93. Yates, L. R. & Campbell, P. J. Evolution of the cancer genome. *Nat. Rev. Genet.* **13,** 795–806 (2012).
94. Chinaranagari, S., Sharma, P., Bowen, N. J. & Chaudhary, J. Prostate cancer epigenome. *Methods Mol. Biol.* **1238,** 125–140 (2015).
95. Belinsky, S. A. Unmasking the lung cancer epigenome. *Annu. Rev. Physiol.* **77,** 453–474 (2015).
96. De Sousa E Melo, F., Vermeulen, L., Fessler, E. & Medema, J. P. Cancer heterogeneity-a multifaceted view. *EMBO reports* **14,** 686–695 (2013).
97. Brocks, D. *et al.* Intratumor DNA methylation heterogeneity reflects clonal evolution in aggressive prostate cancer. *Cell Rep* **8,** 798–806 (2014).
98. Roy, D. M., Walsh, L. A. & Chan, T. A. Driver mutations of cancer epigenomes. *Protein Cell* **5,** 265–296 (2014).
99. Dalgliesh, G. L. *et al.* Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* **463,** 360–363 (2010).
100. Gui, Y. *et al.* Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat. Genet.* **43,** 875–878 (2011).
101. Albini, A. & Sporn, M. B. The tumour microenvironment as a target for chemoprevention. *Nat. Rev. Cancer* **7,** 139–147 (2007).
102. Sounni, N. E. & Noel, A. Targeting the Tumor Microenvironment for Cancer Therapy. *Clin. Chem.* **59,** 85–93 (2013).
103. Cheng, X. & Chen, H. Tumor heterogeneity and resistance to EGFR-targeted therapy in advanced nonsmall cell lung cancer: challenges and perspectives. *Onco Targets Ther* **7,** 1689–1704 (2014).
104. Koti, M. *et al.* A distinct pre-existing inflammatory tumour microenvironment is associated with chemotherapy resistance in high-grade serous epithelial ovarian cancer. *British Journal of Cancer* **112,** 1215–1222 (2015).
105. Landau, D. A. *et al.* Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **152,** 714–726 (2013).
106. Lee, A. J. X. *et al.* Chromosomal Instability Confers Intrinsic Multidrug Resistance. *Cancer Res.* **71,** 1858–1870 (2011).
107. Barbieri, C. E. & Rubin, M. A. Genomic rearrangements in prostate cancer. *Curr Opin Urol* **25,** 71–76 (2015).
108. McGranahan, N., Burrell, R. A., Endesfelder, D., Novelli, M. R. & Swanton, C. Cancer chromosomal instability: therapeutic and diagnostic challenges. *EMBO reports* **13,** 528–538 (2012).
109. Bakhoum, S. F. & Compton, D. A. Chromosomal instability and cancer: a complex relationship with therapeutic potential. *The Journal of clinical investigation* (2012).
110. Ding, L., Raphael, B. J., Chen, F. & Wendl, M. C. Advances for studying clonal evolution in cancer. *Cancer Lett.* **340,** 212–219 (2013).
111. An, X. *et al.* BCR-ABL tyrosine kinase inhibitors in the treatment of Philadelphia chromosome positive chronic myeloid leukemia: a review. *Leuk. Res.* **34,** 1255–1268 (2010).
112. Mullighan, C. G. *et al.* Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. *Science* **322,** 1377–1380 (2008).
113. Keats, J. J. *et al.* Clonal competition with alternating dominance in multiple myeloma. *Blood* **120,** 1067–1076 (2012).
114. Willis, S. G. *et al.* High-sensitivity detection of BCR-ABL kinase domain mutations in imatinib-naive patients: correlation with clonal cytogenetic evolution but not response to therapy. *Blood* **106,** 2128–2137 (2005).

115. Shi, H. *et al.* Acquired resistance and clonal evolution in melanoma during BRAF inhibitor therapy. *Cancer Discov* **4,** 80–93 (2014).
116. Burrell, R. A. & Swanton, C. Tumour heterogeneity and the evolution of polyclonal drug resistance. *Molecular Oncology* **8,** 1095–1111 (2014).
117. Anderson, K. *et al.* Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature* **469,** 356–361 (2011).
118. Schuh, A. *et al.* Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. *Blood* **120,** 4191–4196 (2012).
119. Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461,** 809–813 (2009).
120. Jamal-Hanjani, M. *et al.* Tracking Genomic Cancer Evolution for Precision Medicine: The Lung TRACERx Study. *PLoS Biol.* **12,** (2014).
121. Mitchell, T. J. *et al.* Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer: TRACERx Renal. *Cell* **173,** 611–614.e17 (2018).
122. Turajlic, S. *et al.* Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal. *Cell* **173,** 595–607.e11 (2018).
123. Turajlic, S. *et al.* Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal. *Cell* **173,** 581–589.e12 (2018).
124. Cunningham, J. J., Gatenby, R. A. & Brown, J. S. Evolutionary dynamics in cancer therapy. *Mol. Pharm.* **8,** 2094–2100 (2011).
125. Hurst, C. D., Platt, F. M., Taylor, C. F. & Knowles, M. A. Novel Tumor Subgroups of Urothelial Carcinoma of the Bladder Defined by Integrated Genomic Analysis. *Clin. Cancer Res.* **18,** 5865–5877 (2012).
126. Richter, J. *et al.* Marked genetic differences between stage pTa and stage pT1 papillary bladder cancer detected by comparative genomic hybridization. *Cancer Res.* **57,** 2860–2864 (1997).
127. Gibas, Z., Prout, G. R., Connolly, J. G., Pontes, J. E. & Sandberg, A. A. Nonrandom chromosomal changes in transitional cell carcinoma of the bladder. *Cancer Res.* **44,** 1257–1264 (1984).
128. Fadl-Elmula, I. *et al.* Cytogenetic monoclonality in multifocal uroepithelial carcinomas: evidence of intraluminal tumour seeding. *British Journal of Cancer* **81,** 6–7 (1999).
129. Fadl-Elmula, I. *et al.* Karyotypic characterization of urinary bladder transitional cell carcinomas. *Genes Chromosomes Cancer* **29,** 256–265 (2000).
130. Prat, E. *et al.* Comparative Genomic Hybridization Analysis Reveals New Different Subgroups in Early-stage Bladder Tumors. *URL* **75,** 347–355 (2010).
131. Tsai, Y. C. *et al.* Allelic losses of chromosomes 9, 11, and 17 in human bladder cancer. *Cancer Res.* **50,** 44–47 (1990).
132. Bartoletti, R. *et al.* Loss of P16 expression and chromosome 9p21 LOH in predicting outcome of patients affected by superficial bladder cancer. *J. Surg. Res.* **143,** 422–427 (2007).
133. Knowles, M. A., Habuchi, T., Kennedy, W. & Cuthbert-Heavens, D. Mutation Spectrum of the 9q34 Tuberous Sclerosis Gene TSC1 in Transitional Cell Carcinoma of the Bladder. *Cancer Res.* **63,** 7652–7656 (2003).
134. Platt, F. M. *et al.* Spectrum of Phosphatidylinositol 3-Kinase Pathway Gene Alterations in Bladder Cancer. *Clin. Cancer Res.* **15,** 6008–6017 (2009).
135. Blaveri, E. Bladder Cancer Stage and Outcome by Array-Based Comparative Genomic Hybridization. *Clin. Cancer Res.* **11,** 7012–7022 (2005).
136. Zhao, J. *et al.* Chromosomal Imbalances in Noninvasive Papillary Bladder Neoplasms (pTa). *Cancer Res.* **59,** 4658–4661 (1999).

137. Knowles, M. A. Bladder cancer subtypes defined by genomic alterations. *Scandinavian Journal of Urology and Nephrology* **42,** 116–130 (2010).
138. Ornitz, D. M. & Itoh, N. Fibroblast growth factors. *Genome Biol.* **2,** 3005.1–3005.2 (2001).
139. Hart, K. C. *et al.* Transformation and Stat activation by derivatives of FGFR1, FGFR3, and FGFR4. *Oncogene* **19,** 3309–12 (2000).
140. Cappellen, D. *et al.* Frequent activating mutations of FGFR3 in human bladder and cervix carcinomas. *Nat. Genet.* **23,** 18–3 (1999).
141. Billerey, C. *et al.* Frequent FGFR3 mutations in papillary non-invasive bladder (pTa) tumors. *Am. J. Pathol.* **158,** 1955–1959 (2001).
142. Jebar, A. H. *et al.* FGFR3 and Ras gene mutations are mutually exclusive genetic events in urothelial cell carcinoma. *Oncogene* **24,** 5218–5225 (2005).
143. Kompier, L. C. *et al.* FGFR3, HRAS, KRAS, NRAS and PIK3CA Mutations in Bladder Cancer and Their Potential as Biomarkers for Surveillance and Therapy. *PLoS ONE* **5,** e13821 (2010).
144. di Martino, E., Tomlinson, D. C. & Knowles, M. A. A Decade of FGF Receptor Research in Bladder Cancer: Past, Present, and Future Challenges. *Adv Urol* **2012,** 429213–10 (2012).
145. Tomlinson, D. C., Baldo, O., Harnden, P. & Knowles, M. A. FGFR3 protein expression and its relationship to mutation status and prognostic variables in bladder cancer. *J. Pathol.* **213,** 91–98 (2007).
146. di Martino, E., L'Hôte, C. G., Kennedy, W., Tomlinson, D. C. & Knowles, M. A. Mutant fibroblast growth factor receptor 3 induces intracellular signaling and cellular transformation in a cell type- and mutation-specific manner. *Oncogene* **28,** 4306–4316 (2009).
147. Vara, J. Á. F. *et al.* PI3K/Akt signalling pathway and cancer. *Cancer Treatment Reviews* **30,** 193–204 (2004).
148. López-Knowles, E. *et al.* PIK3CAMutations Are an Early Genetic Alteration Associated with FGFR3Mutations in Superficial Papillary Bladder Tumors. *Cancer Res.* **66,** 7401–7404 (2006).
149. Macaluso, M. *et al.* Ras family genes: An interesting link between cell cycle and cancer. *J. Cell. Physiol.* **192,** 125–130 (2002).
150. Fernandez-Medarde, A. & Santos, E. Ras in Cancer and Developmental Diseases. *Genes & Cancer* **2,** 344–358 (2011).
151. Prior, I. A., Lewis, P. D. & Mattos, C. A Comprehensive Survey of Ras Mutations in Cancer. *Cancer Res.* **72,** 2457–2467 (2012).
152. Knowles, M. A. & Williamson, M. Mutation of H-ras is infrequent in bladder cancer: confirmation by single-strand conformation polymorphism analysis, designed restriction fragment length polymorphisms, and direct sequencing. *Cancer Res.* **53,** 133–139 (1993).
153. Berrada, N. & Amzazi, S. Mutational Analysis of FGFR3 and HRAS Genes in Bladder Cancer and Washing Cell Sediments of Moroccan Patients. *Epidemiology* **05,** 1–6 (2015).
154. Günes, C., Wezel, F., Southgate, J. & Bolenz, C. Implications of TERT promoter mutations and telomerase activity in urothelial carcinogenesis. *Nat Rev Urol* 1–8 (2018). doi:10.1038/s41585-018-0001-5
155. Huang, F. W. *et al.* Highly Recurrent TERT Promoter Mutations in Human Melanoma. *Science* **339,** 957–959 (2013).
156. Hurst, C. D., Platt, F. M. & Knowles, M. A. Comprehensive mutation analysis of the TERT promoter in bladder cancer and detection of mutations in voided urine. *Eur. Urol.* **65,** 367–369 (2014).
157. Pietzak, E. J. *et al.* Next-generation Sequencing of Nonmuscle Invasive Bladder Cancer Reveals Potential Biomarkers and Rational Therapeutic Targets. *Eur. Urol.* **72,** 952–959 (2017).

158. Kiemeney, L. A. L. M., Witjes, J. A., Verbeek, A. L. M., Heijbroek, R. P. & Debruyne, F. M. J. The clinical epidemiology of superficial bladder cancer . *British Journal of Cancer* **67,** 806–812 (1993).
159. Hafner, C., Knuechel, R., Stoehr, R. & Hartmann, A. Clonality of multifocal urothelial carcinomas: 10 years of molecular genetic studies. *Int. J. Cancer* **101,** 1–6 (2002).
160. Sidransky, D. *et al.* Clonal Origin of Bladder-Cancer. *N. Engl. J. Med.* **326,** 737–740 (1992).
161. Hafner, C. *et al.* Evidence for oligoclonality and tumor spread by intraluminal seeding in multifocal urothelial carcinomas of the upper and lower urinary tract. *Oncogene* **20,** 4910–4915 (2001).
162. Xu, X. H., Stower, M. J., Reid, I. N., Garner, R. C. & Burns, P. A. Molecular screening of multifocal transitional cell carcinoma of the bladder using p53 mutations as biomarkers. *Clin. Cancer Res.* **2,** 1795–1800 (1996).
163. van Tilborg, A. *et al.* Molecular evolution of multiple recurrent cancers of the bladder. *Hum. Mol. Genet.* (2000).
164. Nordentoft, I. *et al.* Mutational Context and Diverse Clonal Development in Early and Late Bladder Cancer. *Cell Rep* **7,** 1649–1663 (2014).
165. Trkova, M. *et al.* Analysis of genetic events in 17p13 and 9p21 regions supports predominant monoclonal origin of multifocal and recurrent bladder cancer. *Cancer Lett.* **242,** 68–76 (2006).
166. Malats, N. *et al.* P53 as a prognostic marker for bladder cancer: a meta-analysis and review. *Lancet Oncol.* **6,** 678–686 (2005).
167. Lindgren, D. *et al.* Recurrent and multiple bladder tumors show conserved expression profiles. *BMC Cancer* **8,** 183 (2008).
168. Majewski, T. *et al.* Understanding the development of human bladder cancer by using a whole-organ genomic mapping strategy. *Lab Invest* **88,** 694–721 (2008).
169. Li, Y. *et al.* Single-cell sequencing analysis characterizes common and cell-lineage-specific mutations in a muscle-invasive bladder cancer. *Gigascience* **1,** 12 (2012).
170. Balbás-Martínez, C. *et al.* Recurrent inactivation of STAG2 in bladder cancer is not associated with aneuploidy. *Nat. Genet.* **45,** 1464–1469 (2013).
171. Cazier, J. B. *et al.* Whole-genome sequencing of bladder cancers reveals somatic CDKN1A mutations and clinicopathological associations with mutation burden. *Nature Communications* **5,** 1–11 (2014).
172. Guo, G. *et al.* Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation. *Nat. Genet.* **45,** 1459–U213 (2013).
173. Meeks, J. J. *et al.* Genomic characterization of high-risk non-muscle invasive bladder cancer. *Oncotarget* **7,** 75176–75184 (2016).
174. Warrick, J. I. *et al.* Tumor evolution and progression in multifocal and paired non-invasive/invasive urothelial carcinoma. *Virchows Arch.* 1–15 (2014). doi:10.1007/s00428-014-1699-y
175. Mitra, A. P., Hansel, D. E. & Cote, R. J. Prognostic Value of Cell-Cycle Regulation Biomarkers in Bladder Cancer. *Seminars in Oncology* **39,** 524–533 (2012).
176. Liu, D. *et al.* Mutational patterns in chemotherapy resistant muscle-invasive bladder cancer. *Nature Communications* **8,** 2193 (2017).
177. Ler, L. D. *et al.* Loss of tumor suppressor KDM6A amplifies PRC2-regulated transcriptional repression in bladder cancer and can be targeted through inhibition of EZH2. *Science Translational Medicine* **9,** eaai8312–14 (2017).
178. Lee, M. G. *et al.* Demethylation of H3K27 Regulates Polycomb Recruitment and H2A Ubiquitination. *Science* **318,** 447–447 (2007).

179. Knowles, M. A. & Hurst, C. D. in *DeVita, Hellman, and Rosenberg's Cancer: Principles & Practice of Oncology* (2018).
180. Acar, Ö. *et al.* Determining the origin of synchronous multifocal bladder cancer by exome sequencing. *BMC Cancer* **15,** 1–7 (2015).
181. Vlachostergios, P. J. & Faltas, B. M. Treatment resistance in urothelial carcinoma: an evolutionary perspective. *Nat Rev Clin Oncol* 1–15 (2018). doi:10.1038/s41571-018-0026-y
182. Babjuk, M. *et al.* EAU Guidelines on Non–Muscle-invasive Urothelial Carcinoma of the Bladder: Update 2016. *Eur. Urol.* **71,** 447–461 (2017).
183. Babjuk, M. *et al.* EAU Guidelines on Non-Muscle-invasive Urothelial Carcinoma of the Bladder: Update 2013. *Eur. Urol.* **64,** 639–653 (2013).
184. Iyer, V. N. & Szybalski, W. A Molecular Mechanism of Mitomycin Action: Linking of Complementary DNA Strands. *Proc. Natl. Acad. Sci. U.S.A.* **50,** 355–362 (1963).
185. Basu, A. K. *et al.* Effect of site specifically located mitomycin C-DNA monoadducts on in vitro DNA synthesis by DNA polymerases. *Biochemistry* **32,** 4708–4718 (1993).
186. Tomasz, M., Mercado, C. M., Olson, J. & Chatterjie, N. Mode of interaction of mitomycin C with deoxyribonucleic acid and other polynucleotides in vitro. *Biochemistry* **13,** 4878–4887 (1974).
187. Bizanek, R., McGuinness, B. F., Nakanishi, K. & Tomasz, M. Isolation and structure of an intrastrand cross-link adduct of mitomycin C and DNA. *Biochemistry* **31,** 3084–3091 (1992).
188. Tomasz, M. *et al.* Isolation and Structure of a Covalent Cross-Link Adduct between Mitomycin C and DNA. *Science* **235,** 1204–1208 (1987).
189. Tomasz, M. Mitomycin C: small, fast and deadly (but very selective). *Chemistry Biology* **2,** 575–579 (1995).
190. Avendaño, C. & Menéndez, J. C. in *Medicinal Chemistry of Anticancer Drugs (Second Edition)* (eds. Avendaño, C. & Menéndez, J. C.) 243–271 (Elsevier, 2015).
191. Kumar, S., Lipman, R. & Tomasz, M. Recognition of specific DNA sequences by mitomycin C for alkylation. *Biochemistry* **31,** 1399–1407 (1992).
192. Millard, J. T., Weidner, M. F., Raucher, S. & Hopkins, P. B. Determination of the DNA crosslinking sequence specificity of reductively activated mitomycin C at single-nucleotide resolution: deoxyguanosine residues at CpG are crosslinked preferentially. *J. Am. Chem. Soc.* **112,** 3637–3641 (1990).
193. Palom, Y. *et al.* Relative Toxicities of DNA Cross-Links and Monoadducts: New Insights from Studies of Decarbamoyl Mitomycin C and Mitomycin C. *Chem. Res. Toxicol.* **15,** 1398–1406 (2002).
194. Shatkin, A. J., Reich, E., Franklin, R. M. & Tatum, E. L. Effect of Mitomycin C on Mammalian cells in culture. *Biochim. Biophys. Acta* **55,** 277–289 (1962).
195. Kuroda, Y. & Furuyama, J. Physiological and Biochemical Studies of Effects of Mitomycin C on Strain HeLa Cells in Cell Culture. *Cancer Res.* **23,** 682–687 (1963).
196. Cohen, M. M. & Shaw, M. W. EFFECTS OF MITOMYCIN C ON HUMAN CHROMOSOMES. *The Journal of Cell Biology* **23,** 386–395 (1964).
197. Morad, M., Jonasson, J. & Lindsten, J. Distribution of mitomycin C induced breaks on human chromosomes. *Hereditas* **74,** 273–282 (1973).
198. Bourgeois, C. A. Distribution of Mitomycin C-induced damage in human chromosomes with special reference to regions of repetitive DNA. *Chromosoma* **48,** 203–211 (1974).
199. Sonatakke, Y. A. & Fulzele, R. R. Cytogenetic Study on Genotoxicity of Antitumor-Antibiotic Mitomycin C. *Biomedical Research* **20,** 40–44 (2009).

200. Hovhannisyan, G., Aroutiounian, R. & Liehr, T. Chromosomal Composition of Micronuclei in Human Leukocytes Exposed to Mitomycin C. *Journal of Histochemistry & Cytochemistry* **60,** 316–322 (2012).
201. Srikanth, N. S., Mudipalli, A., Maccubbin, A. E. & Gurtoo, H. L. Mutations in a Shuttle Vector Exposed to Activated Mitomycin. *Molecular Carcinogenesis* 23–29 (1994).
202. Maccubbin, A. E., Mudipalli, A., Nadadur, S. S., Ersing, N. & Gurtoo, H. L. Mutations induced in a shuttle vector plasmid exposed to monofunctionally activated mitomycin C. *Environ. Mol. Mutagen.* **29,** 143–151 (1997).
203. Takeiri, A. *et al.* Molecular Characterization of Mitomycin C-Induced Large Deletions and Tandem-Base Substitutions in the Bone Marrow of gptdelta Transgenic Mice. *Chem. Res. Toxicol.* **16,** 171–179 (2003).
204. Takeiri, A. *et al.* A newly established GDL1 cell line from gpt delta mice well reflects the in vivo mutation spectra induced by mitomycin C. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* **609,** 102–115 (2006).
205. Tam, A. S., Chu, J. S. C. & Rose, A. M. Genome-Wide Mutational Signature of the Chemotherapeutic Agent Mitomycin C in Caenorhabditis elegans. *Genes, Genomes, Genetics* **6,** 133–140 (2016).
206. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. **103,** 1412–1417 (2006).
207. Mishina, T., Oda, K., Murata, S., Ooe, H. & Mori, Y. Mitomycin C bladder instillation therapy for bladder tumors. *J. Urol.* **114,** 217–219 (1975).
208. Bouffioux, C. *et al.* Intravesical adjuvant chemotherapy for superficial transitional cell bladder carcinoma: Results of 2 European Organisation for Research and Treatment of Cancer randomized trials with mitomycin C and doxyrubicin comparing early verses delayed instillations and short term verses long term treatment. *J. Urol.* **153,** 934–941 (1995).
209. Tolley, D. A. *et al.* The effect of intravesical mitomycin C on recurrence of newly diagnosed superficial bladder cancer: a further report with 7 years of follow up. *J. Urol.* **155,** 1233–1238 (1996).
210. Tolley, D. A. *et al.* Effect of intravesical mitomycin C on recurrence of newly diagnosed superficial bladder cancer: interim report from the Medical Research Council Subgroup on Superficial Bladder Cancer (Urological Cancer Working Party). *British MedicalJournal* **296,** 1756–1761 (1988).
211. Sylvester, R. J., Oosterlinck, W. & Witjes, J. A. The Schedule and Duration of Intravesical Chemotherapy in Patients with Non–Muscle-Invasive Bladder Cancer: A Systematic Review of the Published Results of Randomized Clinical Trials. *Eur. Urol.* **53,** 709–719 (2008).
212. Bosschieter, J. *et al.* Value of an Immediate Intravesical Instillation of Mitomycin C in Patients with Non–muscle-invasive Bladder Cancer: A Prospective Multicentre Randomised Study in 2243 patients. *Eur. Urol.* **73,** 226–232 (2018).
213. Sylvester, R. J. *et al.* Systematic Review and Individual Patient Data Meta-analysis of Randomized Trials Comparing a Single Immediate Instillation of Chemotherapy After Transurethral Resection with Transurethral Resection Alone in Patients with Stage pTa–pT1 Urothelial Carcinoma of the Bladder: Which Patients Benefit from the Instillation? *Eur. Urol.* **69,** 231–244 (2016).
214. Sylvester, R. J., Oosterlinck, W. & van der Meijden, A. A single immediate postoperative instillation of chemotherapy decreases the risk of recurrence in patients with stage Ta T1 bladder cancer: A meta-analysis of published results of randomized clinical trials. *J. Urol.* **171,** 2186–2190 (2004).

215. Perlis, N. *et al.* Immediate Post–Transurethral Resection of Bladder Tumor Intravesical Chemotherapy Prevents Non–Muscle-invasive Bladder Cancer Recurrences: An Updated Meta-analysis on 2548 Patients and Quality-of-Evidence Review. *Eur. Urol.* **64,** 421–430 (2013).
216. Colombo, R. *et al.* Neoadjuvant Short-term Intensive Intravesical Mitomycin C Regimen Compared with Weekly Schedule for Low-grade Recurrent Non–muscle-invasive Bladder Cancer: Preliminary Results of a Randomised Phase 2 Study. *Eur. Urol.* **62,** 797–802 (2012).
217. Shelley, M. D. *et al.* A systematic review of intravesical bacillus Calmette-Guérin plus transurethral resection vs transurethral resection alone in Ta and T1 bladder cancer. *BJU Int* **88,** 209–216 (2001).
218. Kapoor, R., Vijjan, V. & Singh, P. Bacillus Calmette-Guerin in the management of superficial bladder cancer. *Indian J Urol* **24,** 72–76 (2008).
219. Lundholm, C. *et al.* A Randomized Prospective Study Comparing Long-Term Intravesical Instillations of Mitomycin C and Bacillus Calmette-Guerin in Patients with Superficial Bladder Carcinoma. *J. Urol.* **156,** 372–376 (1996).
220. Friedrich, M. G., Pichlmeier, U., Schwaibold, H., Conrad, S. & Huland, H. Long-Term Intravesical Adjuvant Chemotherapy Further Reduces Recurrence Rate Compared with Short-Term Intravesical Chemotherapy and Short-Term Therapy with Bacillus Calmette-Guérin (BCG) in Patients with Non–Muscle-Invasive Bladder Carcinoma. *Eur. Urol.* **52,** 1123–1130 (2007).
221. Järvinen, R., Kaasinen, E., Sankila, A. & Rintala, E. Long-term Efficacy of Maintenance Bacillus Calmette-Guérin versus Maintenance Mitomycin C Instillation Therapy in Frequently Recurrent TaT1 Tumours without Carcinoma In Situ: A Subgroup Analysis of the Prospective, Randomised FinnBladder I Study with a 20-Year Follow-up. *Eur. Urol.* **56,** 260–265 (2009).
222. Shelley, M. D. *et al.* Intravesical bacillus Calmette-Guérin is superior to mitomycin C in reducing tumour recurrence in high-risk superficial bladder cancer: a meta-analysis of randomized trials. *BJU Int* **93,** 485–490 (2004).
223. Böhle, A. & Bock, P. R. Intravesical bacille calmette-guérin versus mitomycin c in superficial bladder cancer: formal meta-analysis of comparative studies on tumor progression. *Urology* **63,** 682–686 (2004).
224. Malmström, P.-U. *et al.* An Individual Patient Data Meta-Analysis of the Long-Term Outcome of Randomised Studies Comparing Intravesical Mitomycin C versus Bacillus Calmette-Guérin for Non–Muscle-Invasive Bladder Cancer. *Eur. Urol.* **56,** 247–256 (2009).
225. Schmidt, S. *et al.* Intravesical bacillus Calmette-Guérin versus mitomycin C for Ta and T1 bladder cancer. *Cochrane Database of Systematic Reviews* **11,** 477–15 (2015).
226. Sylvester, R. J. Bacillus Calmette-Guérin versus Mitomycin C for the Treatment of Intermediate-Risk Non–Muscle-Invasive Bladder Cancer: The Debate Continues. *Eur. Urol.* **56,** 266–268 (2009).
227. Huncharek, M., Geschwind, J.-F., Witherspoon, B., McGarry, R. & Adcock, D. Intravesical chemotherapy prophylaxis in primary superficial bladder cancer: a meta-analysis of 3703 patients from 11 randomized trials. *Journal of Clinical Epidemiology* **53,** 676–680 (2000).
228. Messing, E. M. The BCG Shortage. *BLC* **3,** 227–228 (2017).
229. Bandari, J., Maganty, A., MacLeod, L. C. & Davies, B. J. Manufacturing and the Market: Rationalizing the Shortage of Bacillus Calmette-Guérin. *European Urology Focus* **4,** 481–484 (2018).
230. Dean, F. B. *et al.* Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 5261–5266 (2002).

231. Lasken, R. S. Genomic DNA amplification by the multiple displacement amplification (MDA) method. *Biochem. Soc. Trans* **37,** 450 (2009).
232. van Oers, J. *et al.* A simple and fast method for the simultaneous detection of nine fibroblast growth factor receptor 3 mutations in bladder cancer and voided urine. *Clin. Cancer Res.* **11,** 7743–7748 (2005).
233. Hurst, C. D., Zuiverloon, T. C. M., Hafner, C., Zwarthoff, E. C. & Knowles, M. A. A SNaPshot assay for the rapid and simple detection of four common hotspot codon mutations in the PIK3CA gene. *BMC Res Notes* **2,** 66 (2009).
234. Andrews, S. FastQC: A quality control tool for high throughput sequence data. (2010). Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. (Accessed: 10 July 2018)
235. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26,** 589–595 (2010).
236. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
237. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* **46,** W537–W544 (2018).
238. Kim, S. *et al.* Strelka2: fast and accurate calling of germline and somatic variants. *Nat. Methods* 1–10 (2018). doi:10.1038/s41592-018-0051-x
239. Shiraishi, Y. *et al.* An empirical Bayesian framework for somatic mutation detection from cancer genome sequencing data. *Nucleic Acids Res.* **41,** e89–e89 (2013).
240. Koboldt, D. C., Larson, D. E. & Wilson, R. K. Using VarScan 2 for Germline Variant Calling and Somatic Mutation Detection. *Current Protocols in Bioinformatics* **467,** 15.4.1–15.4.17 (2013).
241. Fan, Y. *et al.* MuSE: accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling from sequencing data. *Genome Biol.* 1–11 (2016). doi:10.1186/s13059-016-1029-6
242. Droop, A. *et al.* How to analyse the spatiotemporal tumour samples needed to investigate cancer evolution: A case study using paired primary and recurrent glioblastoma. *Int. J. Cancer* **142,** 1620–1626 (2017).
243. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17,** (2016).
244. Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **28,** 1747–1756 (2018).
245. Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res.* **44,** e131–e131 (2016).
246. Roth, A. *et al.* PyClone: statistical inference of clonal population structure in cancer. *Nat. Methods* **11,** 396–398 (2014).
247. Dang, H. X. *et al.* ClonEvol: clonal ordering and visualization in cancer sequencing. *Ann. Oncol.* **28,** 3076–3082 (2017).
248. Miller, C. A. *et al.* Visualizing tumor evolution with the fishplot package for R. *BMC Genomics* **17,** 880 (2016).
249. Chapman, E. J. Comprehensive Analysis of CDKN2A Status in Microdissected Urothelial Cell Carcinoma Reveals Potential Haploinsufficiency, a High Frequency of Homozygous Co-deletion and Associations with Clinical Phenotype. *Clin. Cancer Res.* **11,** 5740–5747 (2005).

250. Beothe, T., Zubakov, D. & Kovacs, G. Homozygous losses detected by array comparative genomic hybridization in multiplex urothelial carcinomas of the bladder. *Cancer Genetics* **208,** 434–440 (2015).
251. Zhao, M., Wang, Q., Wang, Q., Jia, P. & Zhao, Z. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. *BMC Bioinformatics* **14,** S1 (2013).
252. Zhou, B. *et al.* Whole-genome sequencing analysis of CNV using low-coverage and paired-end strategies is efficient and outperforms array-based CNV analysis. *J Med Genet* **55,** 735 (2018).
253. Magi, A., Tattini, L., Pippucci, T., Torricelli, F. & Benelli, M. Read count approach for DNA copy number variants detection. *Bioinformatics* **28,** 470–478 (2011).
254. Scheinin, I. *et al.* DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res.* **24,** 2022–2032 (2014).
255. Liang, D. *et al.* Copy Number Variation Sequencing for Comprehensive Diagnosis of Chromosome Disease Syndromes. *The Journal of Molecular Diagnostics* **16,** 519–526 (2014).
256. Tattini, L., D'Aurizio, R. & Magi, A. Detection of Genomic Structural Variants from Next-Generation Sequencing Data. *Front. Bioeng. Biotechnol.* **3,** 1–8 (2015).
257. Medvedev, P., Stanciu, M. & Brudno, M. Computational methods for discovering structural variation with next-generation sequencing. *Nat. Methods* **6,** S13–S20 (2009).
258. Krijgsman, O., Carvalho, B., Meijer, G. A., Steenbergen, R. D. M. & Ylstra, B. Focal chromosomal copy number aberrations in cancer—Needles in a genome haystack. *BBA - Molecular Cell Research* **1843,** 2698–2704 (2014).
259. Junker, K. *et al.* Fibroblast Growth Factor Receptor 3 Mutations in Bladder Tumors Correlate with Low Frequency of Chromosome Alterations. *Neoplasia* **10,** 1–7 (2008).
260. Sylvester, R. J. *et al.* Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: A combined analysis of 2596 patients from seven EORTC trials. *Eur. Urol.* **49,** 466–477 (2006).
261. Waldman, F. M. *et al.* Centromeric copy number of chromosome 7 is strongly correlated with tumor grade and labeling index in human bladder cancer. *Cancer Res.* **51,** 3807–3813 (1991).
262. Santos, L. *et al.* Chromosome Instability and Progression in Urothelial Cell Carcinoma of the Bladder. *Acta Oncologica* **42,** 169–173 (2011).
263. Rebouissou, S. *et al.* CDKN2A homozygous deletion is associated with muscle invasion in FGFR3-mutated urothelial bladder carcinoma. *J. Pathol.* **227,** 315–324 (2012).
264. Zhang, L., Yuan, Y., Lu, K. H. & Zhang, L. Identification of recurrent focal copy number variations and their putative targeted driver genes in ovarian cancer. *BMC Bioinformatics* 1–12 (2016). doi:10.1186/s12859-016-1085-7
265. Nord, H. *et al.* Focal amplifications are associated with high grade and recurrences in stage Ta bladder carcinoma. *Int. J. Cancer* **16,** NA–NA (2009).
266. Veltman, J. A. *et al.* Array-based Comparative Genomic Hybridization for Genome-Wide Screening of DNA Copy Number in Bladder Tumors. *Cancer Res.* 1–10 (2003).
267. Qie, S. & Diehl, J. A. Cyclin D1, cancer progression, and opportunities in cancer treatment. *J Mol Med* **94,** 1313–1326 (2016).

268. Yun, Y.-R. *et al.* Fibroblast Growth Factors: Biology, Function, and Application for Tissue Regeneration. *J Tissue Eng* **1,** 218142–18 (2010).
269. Dearth, R. K., Cui, X., Kim, H.-J., Hadsell, D. L. & Lee, A. V. Oncogenic Transformation by the Signaling Adaptor Proteins Insulin Receptor Substrate (IRS)-1 and IRS-2. *Cell Cycle* **6,** 705–713 (2014).
270. Day, E. *et al.* IRS2is a candidate driver oncogene on 13q34 in colorectal cancer. *Int. J. Exp. Path.* **94,** 203–211 (2013).
271. Veeriah, S. *et al.* The tyrosine phosphatase PTPRD is a tumor suppressor that is frequently inactivated and mutated in glioblastoma and other human cancers. *PNAS* **106,** 9435–9440 (2009).
272. Solomon, D. A. *et al.* Mutational Inactivation of PTPRD in Glioblastoma Multiforme and Malignant Melanoma. *Cancer Res.* **68,** 10300–10306 (2008).
273. Solomon, D. A. *et al.* Frequent truncating mutations of STAG2 in bladder cancer. *Nat. Genet.* **45,** 1428–1430 (2013).
274. Taylor, C. F., Platt, F. M., Hurst, C. D., Thygesen, H. H. & Knowles, M. A. Frequent inactivating mutations of STAG2 in bladder cancer are associated with low tumour grade and stage and inversely related to chromosomal copy number changes. *Hum. Mol. Genet.* **23,** 1964–1974 (2014).
275. Remeseiro, S. & Losada, A. Cohesin, a chromatin engagement ring. *Current Opinion in Cell Biology* **25,** 63–71 (2013).
276. Walter, M. J. *et al.* Acquired copy number alterations in adult acute myeloid leukemia genomes. *Proc Natl Acad Sci USA* **106,** 12950–12955 (2009).
277. Solomon, D. A. *et al.* Mutational Inactivation of STAG2 Causes Aneuploidy in Human Cancer. *Science* **333,** 1039–1043 (2011).
278. Li, X. *et al.* Loss of STAG2 causes aneuploidy in normal human bladder cells. *Genet. Mol. Res.* **14,** 2638–2646 (2015).
279. Qiao, Y., Zhu, X., Li, A., Yang, S. & Zhang, J. Complete loss of STAG2 expression is an indicator of good prognosis in patients with bladder cancer. *Tumor Biol.* **37,** 10279–10286 (2016).
280. Lelo, A. *et al.* STAG2 Is a Biomarker for Prediction of Recurrence and Progression in Papillary Non-Muscle-Invasive Bladder Cancer. *Clin. Cancer Res.* **24,** 4145–4153 (2018).
281. Natarajan, A. T., Tates, A. D., Meijers, M., Neuteboom, I. & de Vogel, N. Induction of sister-chromatid exchanges (SCEs) and chromosomal aberrations by mitomycin C and methyl methanesulfonate in Chinese hamster ovary cells An evaluation of methodology for detection of SCEs and of persistent DNA lesions towards the frequencies of observed SCEs. *Mutation Research* **121,** 211–223 (1983).
282. van Rhijn, B. W. G. *et al.* Molecular Grade (FGFR3/MIB-1) and EORTC Risk Scores Are Predictive in Primary Non-Muscle-Invasive Bladder Cancer. *Eur. Urol.* **58,** 433–441 (2010).
283. Juanpere, N. *et al.* Mutations in FGFR3 and PIK3CA, singly or combined with RAS and AKT1, are associated with AKT but not with MAPK pathway activation in urothelial bladder cancer. *Hum. Pathol.* **43,** 1573–1582 (2012).
284. Mendoza, M. C., Er, E. E. & Blenis, J. The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation. *Trends in Biochemical Sciences* **36,** 320–328 (2011).
285. Wu, S. *et al.* Telomerase Reverse Transcriptase Gene Promoter Mutations Help Discern the Origin of Urogenital Tumors: A Genomic and Molecular Study. *Eur. Urol.* **65,** 274–277 (2014).

286. Isharwal, S. *et al.* Prognostic Value of TERT Alterations, Mutational and Copy Number Alterations Burden in Urothelial Carcinoma. *European Urology Focus* 1–4 (2017). doi:10.1016/j.euf.2017.07.004
287. McGranahan, N. *et al.* Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Science Translational Medicine* **7,** 283ra54–283ra54 (2015).
288. Temko, D., Tomlinson, I. P. M., Severini, S., Schuster-Böckler, B. & Graham, T. A. The effects of mutational processes and selection on driver mutations across cancer types. *Nature Communications* 1–10 (2018). doi:10.1038/s41467-018-04208-6
289. Allory, Y. *et al.* Telomerase Reverse Transcriptase Promoter Mutations in Bladder Cancer: High Frequency Across Stages, Detection in Urine, and Lack of Association with Outcome. *Eur. Urol.* **65,** 360–366 (2014).
290. Descotes, F. C. O. *et al.* Non-invasive prediction of recurrence in bladder cancer by detecting somatic TERT promoter mutations in urine. *British Journal of Cancer* **117,** 583–587 (2017).
291. Dahmcke, C. M. *et al.* A Prospective Blinded Evaluation of Urine-DNA Testing for Detection of Urothelial Bladder Carcinoma in Patients with Gross Hematuria. *Eur. Urol.* **70,** 916–919 (2016).
292. van Kessel, K. E. M., Van Neste, L., Lurkin, I., Zwarthoff, E. C. & Van Criekinge, W. Evaluation of an Epigenetic Profile for the Detection of Bladder Cancer in Patients with Hematuria. *J. Urol.* **195,** 601–607 (2016).
293. Tan, W. S. *et al.* Novel urinary biomarkers for the detection of bladder cancer: A systematic review. *Cancer Treatment Reviews* **69,** 39–52 (2018).
294. Gakis, G. & Fahmy, O. Systematic Review and Meta-Analysis on the Impact of Hexaminolevulinate- Versus White-Light Guided Transurethral Bladder Tumor Resection on Progression in Non-Muscle Invasive Bladder Cancer. *BLC* **2,** 293–300 (2016).
295. Lipponen, P. K., Eskelinen, M. J., Jauhiainen, K., Harju, E. & Terho, R. Tumour infiltrating lymphocytes as an independent prognostic factor in transitional cell bladder cancer. *European Journal of Cancer* **29A,** 69–75 (1992).
296. Sharma, P. *et al.* CD8 tumor-infiltrating lymphocytes are predictive of survival in muscle-invasive urothelial carcinoma. *Proc Natl Acad Sci USA* **104,** 3967–3972 (2007).
297. Huang, H.-S. *et al.* Prognostic impact of tumor infiltrating lymphocytes on patients with metastatic urothelial carcinoma receiving platinum based chemotherapy. *Scientific Reports* 1–7 (2018). doi:10.1038/s41598-018-25944-1
298. Rouanne, M. *et al.* Stromal lymphocyte infiltration is associated with tumour invasion depth but is not prognostic in high-grade T1 bladder cancer. *European Journal of Cancer* **108,** 111–119 (2019).
299. Ray, F. A. *et al.* Directional genomic hybridization for chromosomal inversion discovery and detection. *Chromosome Res* **21,** 165–174 (2013).
300. Johnson, B. E. *et al.* Mutational Analysis Reveals the Origin and Therapy-Driven Evolution of Recurrent Glioma. *Science* **343,** 189–193 (2014).
301. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31,** 213–219 (2013).
302. Xu, C. A review of somatic single nucleotide variant calling algorithms for next-generation sequencing data. *Computational and Structural Biotechnology Journal* **16,** 15–24 (2018).
303. Kim, S. Y. & Speed, T. P. Comparing somatic mutation-callers: beyond Venn diagrams. *BMC Bioinformatics* **14,** 1–16 (2013).

304. Ewing, A. D. *et al.* Combining tumor genome simulation with crowdsourcing to benchmark somatic single-nucleotide-variant detection. *Nat. Methods* **12,** 623–630 (2015).
305. Spencer, D. H. *et al.* Performance of Common Analysis Methods for Detecting Low-Frequency Single Nucleotide Variants in Targeted Next-Generation Sequence Data. *The Journal of Molecular Diagnostics* **16,** 75–88 (2014).
306. The DREAM5 Consortium *et al.* Wisdom of crowds for robust gene network inference. *Nat. Methods* **9,** 796–804 (2012).
307. Margolin, A. A. *et al.* Systematic Analysis of Challenge-Driven Improvements in Molecular Prognostic Models for Breast Cancer. *Science Translational Medicine* **5,** 181re1–181re1 (2013).
308. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499,** 214–218 (2013).
309. Tan, H., Bao, J. & Zhou, X. Genome-wide mutational spectra analysis reveals significant cancer-specific heterogeneity. *Scientific Reports* **5,** (2015).
310. Hedegaard, J. *et al.* Comprehensive Transcriptional Analysis of Early-Stage Urothelial Carcinoma. *Cancer Cell* **30,** 27–42 (2016).
311. Alexandrov, L. B. *et al.* Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47,** 1402–1407 (2015).
312. Alexandrov, L. B. *et al.* The Repertoire of Mutational Signatures in Human Cancer. *bioRxiv* 1–44 (2018). doi:10.1101/322859
313. Spencer, D. H. *et al.* Comparison of Clinical Targeted Next-Generation Sequence Data from Formalin-Fixed and Fresh-Frozen Tissue Specimens. *The Journal of Molecular Diagnostics* **15,** 623–633 (2013).
314. Prentice, L. M. *et al.* Formalin fixation increases deamination mutation signature but should not lead to false positive mutations in clinical practice. *PLoS ONE* **13,** e0196434–14 (2018).
315. Ma, X. *et al.* Analysis of error profiles in deep next- generation sequencing data. 1–15 (2019). doi:10.1186/s13059-019-1659-6
316. Salk, J. J., Schmitt, M. W. & Loeb, L. A. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nature Publishing Group* **19,** 269–285 (2018).
317. Schulze, K. *et al.* Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat. Genet.* **47,** 505–511 (2015).
318. Hsu, I. C. *et al.* Mutational hot spot in the p53 gene in human hepatocellular carcinomas. *Nature* **350,** 427–428 (1991).
319. Letouzé, E. *et al.* Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nature Communications* 1–13 (2017). doi:10.1038/s41467-017-01358-x
320. Zhang, W. *et al.* Genetic Features of Aflatoxin-Associated Hepatocellular Carcinoma. *Gastroenterology* **153,** 249–262.e2 (2017).
321. Huang, M. N. *et al.* Genome-scale mutational signatures of aflatoxin in cells, mice, and human tumors. *Genome Res.* **27,** 1475–1486 (2017).
322. Smela, M. E., Currier, S. S., Bailey, E. A. & Essigmann, J. M. The chemistry and biology of aflatoxin B_1: from mutational spectrometry to carcinogenesis. *Carcinogenesis* **22,** 535–545 (2001).
323. Essigmann, J. M. *et al.* Structural identification of the major DNA adduct formed by aflatoxin B_1 *in vitro*. *Proc Natl Acad Sci USA* **74,** 1870–1874 (1977).
324. Foster, P. L., Eisenstadt, E. & Miller, J. H. Base substitution mutations induced by metabolically activated aflatoxin B_1. *Proc Natl Acad Sci USA* **80,** 2695–2698 (1983).

325. Network, T. C. G. A. R. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **507,** 315–322 (2014).
326. Hanane, O., Gianluca, S. & Vittorio, P. Computational tools to detect signatures of mutational processes in DNA from tumours: a review and empirical comparison of performance. 1–53 (2019). doi:10.1101/483982
327. Ramazzotti, D., Lal, A., Liu, K., Tibshirani, R. & Sidow, A. De Novo Mutational Signature Discovery in Tumor Genomes using SparseSignatures. 1–25 (2018). doi:10.1101/384834
328. Baez-Ortega, A. & Gori, K. Computational approaches for discovery of mutational signatures in cancer. *Briefings in Bioinformatics* **20,** 77–88 (2017).
329. Huang, P.-J. *et al.* mSignatureDB: a database for deciphering mutational signatures in human cancers. *Nucleic Acids Res.* **46,** D964–D970 (2017).
330. Alexandrov, L. B., Nik-Zainal, S. A., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering Signatures of Mutational Processes Operative in Human Cancer. *Cell Rep* **3,** 246–259 (2013).
331. Blokzijl, F., Janssen, R., van Boxtel, R. & Cuppen, E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. 1–11 (2018). doi:10.1186/s13073-018-0539-0
332. Grzywa, T. M., Paskal, W. & odarski, P. K. W. Intratumor and Intertumor Heterogeneity in Melanoma. *Translational Oncology* **10,** 956–975 (2017).
333. Wang, Y. *et al.* Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* **512,** 155–160 (2014).
334. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481,** 306–313 (2012).
335. Hajirasouliha, I., Mahmoody, A. & Raphael, B. J. A combinatorial approach for analyzing intra-tumor heterogeneity from high-throughput sequencing data. *Bioinformatics* **30,** i78–86 (2014).
336. Jiang, Y., Qiu, Y., Minn, A. J. & Zhang, N. R. Assessing intratumor heterogeneity and tracking longitudinal and spatial clonal evolutionary history by next-generation sequencing. *Proc. Natl. Acad. Sci. U.S.A.* 201522203 (2016). doi:10.1073/pnas.1522203113
337. Miller, C. A. *et al.* SciClone: Inferring Clonal Architecture and Tracking the Spatial and Temporal Patterns of Tumor Evolution. *Plos Computational Biology* **10,** (2014).
338. Malikic, S., McPherson, A. W., Donmez, N. & Sahinalp, C. S. Clonality Inference in Multiple Tumor Samples using Phylogeny. *Bioinformatics* btv003 (2015). doi:10.1093/bioinformatics/btv003
339. Amarasinghe, K. C. *et al.* Inferring copy number and genotype in tumour exome data. *BMC Genomics* **15,** 732–12 (2014).
340. Sathirapongsasuti, J. F. *et al.* Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. *Bioinformatics* **27,** 2648–2654 (2011).
341. Morrison, C. D. *et al.* Whole-genome sequencing identifies genomic heterogeneity at a nucleotide and chromosomal level in bladder cancer. *Proc Natl Acad Sci USA* **111,** E672–E681 (2014).
342. Macaulay, I. C. & Voet, T. Single Cell Genomics: Advances and Future Perspectives. *PLoS Genet.* **10,** (2014).
343. Ma, Q. C., Ennis, C. A. & Aparicio, S. Opening Pandora's Box - the new biology of driver mutations and clonal evolution in cancer as revealed by next generation sequencing. *Curr. Opin. Genet. Dev.* **22,** 3–9 (2012).
344. Hughes, A. E. O. *et al.* Clonal architecture of secondary acute myeloid leukemia defined by single-cell sequencing. *PLoS Genet.* **10,** e1004462 (2014).
345. Xu, X. *et al.* Single-Cell Exome Sequencing Reveals Single-Nucleotide Mutation Characteristics of a Kidney Tumor. *Cell* **148,** 886–895 (2012).

346. Gust, K. M. *et al.* Fibroblast Growth Factor Receptor 3 Is a Rational Therapeutic Target in Bladder Cancer. *Molecular Cancer Therapeutics* **12,** 1245–1254 (2013).
347. Chae, Y. K. *et al.* Inhibition of the fibroblast growth factor receptor (FGFR) pathway: the current landscape and barriers to clinical application. *Oncotarget* **8,** 1–23 (2016).
348. Kosaka, T. *et al.* Analysis of Epidermal Growth Factor Receptor Gene Mutation in Patients with Non-Small Cell Lung Cancer and Acquired Resistance to Gefitinib. *Clin. Cancer Res.* **12,** 5764–5769 (2006).
349. Engelman, J. A. *et al.* MET Amplification Leads to Gefitinib Resistance in Lung Cancer by Activating ERBB3 Signaling. *Science* **316,** 1039–1043 (2007).
350. Li, H.-T., Duymcih, C. E., Welsenberger, D. J. & Liang, G. Genetic and Epigenetic Alterations in Bladder Cancer. *Int Neurourol J* **20,** S84–94 (2016).
351. Wu, C. *et al.* Targeting AURKA-CDC25C axis to induce synthetic lethality in ARID1A-deficient colorectal cancer cells. *Nature Communications* 1–14 (2018). doi:10.1038/s41467-018-05694-4
352. Bitler, B. G. *et al.* Synthetic lethality by targeting EZH2 methyltransferase activity in ARID1A-mutated cancers. *Nat. Med.* **21,** 231–238 (2015).
353. Knudsen, E. S. & Wang, J. Y. J. Targeting the RB-pathway in Cancer Therapy. *Clin. Cancer Res.* **16,** 1094–1099 (2010).
354. Moon, N. S., Di Stefano, L. & Dyson, N. A Gradient of Epidermal Growth Factor Receptor Signaling Determines the Sensitivity of rbf1 Mutant Cells to E2F-Dependent Apoptosis. *Molecular and Cellular Biology* **26,** 7601–7615 (2006).
355. Morris, E. J. *et al.* Functional Identification of Api5 as a Suppressor of E2F-Dependent Apoptosis In Vivo. *PLoS Genet.* **2,** e196–15 (2006).
356. Nip, J. *et al.* E2F-1 Cooperates with Topoisomerase II Inhibition and DNA Damage To Selectively Augment p53-Independent Apoptosis. *Molecular and Cellular Biology* **17,** 1049–1056 (1997).
357. Agerbaek, M., Alsner, J., Marcussen, N., Lundbeck, F. & Maase, von der, H. Retinoblastoma protein expression is an independent predictor of both radiation response and survival in muscle-invasive bladder cancer. *British Journal of Cancer* **89,** 298–304 (2003).
358. Ojha, J. *et al.* Deep sequencing identifies genetic heterogeneity and recurrent convergent evolution in chronic lymphocytic leukemia. *Blood* **125,** 492–498 (2015).
359. Swanton, C. Intratumor Heterogeneity: Evolution through Space and Time. *Cancer Res.* **72,** 4875–4882 (2012).
360. Kim, H. *et al.* Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. *Genome Res.* **25,** 316–327 (2015).
361. Spruck, C. H., III *et al.* Two Molecular Pathways to Transitional Cell Carcinoma of the Bladder. *Cancer Res.* **54,** 784–788 (1994).
362. Cheng, L. *et al.* Precise microdissection of human bladder carcinomas reveals divergent tumor subclones in the same tumor. *Cancer* **94,** 104–110 (2001).
363. Hu, J., Lieb, J. D., Sancar, A. & Adar, S. Cisplatin DNA damage and repair maps of the human genome at single-nucleotide resolution. *Proc Natl Acad Sci USA* **113,** 11507–11512 (2016).
364. Tran, A. N. *et al.* High-resolution detection of chromosomal rearrangements in leukemias through mate pair whole genome sequencing. *PLoS ONE* **13,** e0193928–10 (2018).
365. Smadbeck, J. B. *et al.* Copy number variant analysis using genome-wide mate-pair sequencing. *Genes Chromosomes Cancer* **57,** 459–470 (2018).

366. Jain, M. *et al.* Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* **36,** 338–345 (2018).
367. Payne, A., Holmes, N., Rakyan, V. & Loose, M. Whale watching with BulkVis: A graphical viewer for Oxford Nanopore bulk fast5 files. *bioRxiv* 1–28 (2018). doi:10.1101/312256
368. Jain, M. *et al.* Improved data analysis for the MinION nanopore sequencer. *Nat. Methods* **12,** 351–356 (2015).
369. Jain, M. *et al.* Linear assembly of a human centromere on the Y chromosome. *Nat. Biotechnol.* **36,** 321–323 (2018).
370. Wang, X. *et al.* Copy number alterations detected by whole-exome and whole-genome sequencing of esophageal adenocarcinoma. *Human Genomics* 1–15 (2015). doi:10.1186/s40246-015-0044-0
371. Brown, S. D. *et al.* Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. *Genome Res.* **24,** 743–750 (2014).
372. Rizvi, N. A. *et al.* Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348,** 124–128 (2015).
373. Snyder, A. *et al.* Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma. *N. Engl. J. Med.* **371,** 2189–2199 (2014).
374. Rosenberg, J. E. *et al.* Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *The Lancet* **387,** 1909–1920 (2016).
375. Balar, A. V. *et al.* Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: a single-arm, multicentre, phase 2 trial. *The Lancet* **389,** 67–76 (2017).
376. Kaasinen, E. *et al.* Alternating Mitomycin C and BCG Instillations versus BCG Alone in Treatment of Carcinoma in Situ of the Urinary Bladder: A Nordic Study. *Eur. Urol.* **43,** 637–645 (2003).
377. Svatek, R. S. *et al.* Sequential Intravesical Mitomycin plus Bacillus Calmette-Guerin for Non-Muscle-Invasive Urothelial Bladder Carcinoma: Translational and Phase I Clinical Trial. *Clin. Cancer Res.* **21,** 303–311 (2015).
378. Solsona, E. *et al.* Sequential Combination of Mitomycin C Plus Bacillus Calmette-Gue´rin (BCG) Is More Effective but More Toxic Than BCG Alone in Patients with Non–Muscle-invasive Bladder Cancer in Intermediate- and High-risk Patients: Final Outcome of CUETO 93009, a Randomized Prospective Trial. *Eur. Urol.* **67,** 508–516 (2015).
379. Oosterlinck, W. *et al.* Sequential Intravesical Chemoimmunotherapy with Mitomycin C and Bacillus Calmette-Gue´rin and with Bacillus Calmette-Gue´rin Alone in Patients with Carcinoma in Situ of the Urinary Bladder: Results of an EORTC Genito-Urinary Group Randomized Phase 2 Trial (30993). *Eur. Urol.* **59,** 438–446 (2011).
380. Järvinen, R., Kaasinen, E., Rintala, E. & Group, T. F. Long-term results of maintenance treatment of mitomycin C or alternating mitomycin C and bacillus Calmette–Guérin instillation therapy of patients with carcinoma in situ of the bladder: A subgroup analysis of the prospective FinnBladder 2 study with a 17-year follow-up. *Scandinavian Journal of Urology and Nephrology* **46,** 411–417 (2012).
381. Deng, T. *et al.* Systematic Review and Cumulative Analysis of the Combination of Mitomycin C plus Bacillus Calmette-Guérin (BCG) for Non–Muscle-Invasive Bladder Cancer. *Scientific Reports* **7,** 225–10 (2017).
382. Karasaki, T. *et al.* Prediction and prioritization of neoantigens: integration of RNA sequencing data with whole-exome sequencing. *Cancer Sci.* **108,** 170–177 (2017).

383. Kompier, L. C. *et al.* The development of multiple bladder tumour recurrences in relation to the FGFR3 mutation status of the primary tumour. *J. Pathol.* **218,** 104–112 (2009).
384. Chandrasekar, T. AUA 2018: Results of CALIBER: A Phase II Randomized Feasibility Trial of Chemoablation vs Surgical Management in Low Risk Non-Muscle Invasive Bladder Cancer. *Uro Today* (2018). Available at: https://www.urotoday.com/conference-highlights/aua-2018/aua-2018-bladder-cancer/104386-aua-2018-results-of-caliber-a-phase-ii-randomized-feasibility-trial-of-chemoablation-vs-surgical-management-in-low-risk-non-muscle-invasive-bladder-cancer.html. (Accessed: 9 January 2019)
385. McConkey, D. J. & Choi, W. Molecular Subtypes of Bladder Cancer. *Curr Oncol Rep* **20,** 1–7 (2018).
386. Khojasteh, M., Lam, W. L., Ward, R. K. & MacAulay, C. A stepwise framework for the normalization of array CGH data. *BMC Bioinformatics* **6,** (2005).
387. Chen, Y.-C., Liu, T., Yu, C.-H., Chiang, T.-Y. & Hwang, C.-C. Effects of GC Bias in Next-Generation-Sequencing Data on De Novo Genome Assembly. *PLoS ONE* **8,** e62856 (2013).