

**Exploring tumour evolution through mutational
patterns in bone tumours**

Matthew Fittall

University College London
and
The Francis Crick Institute

PhD Supervisors: Peter Van Loo and Adrienne M. Flanagan

A thesis submitted for the degree of
Doctor of Philosophy
University College London

October 2019

Declaration

I, Matthew Fittall confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Cancer is a continuation of the evolutionary process on a cellular scale. The mutations that define this evolutionary process show a marked variety of complexity, which I have explored in this work.

First, I have explored the genomics of osteoblastoma, a rare benign bone tumour. This work, for the first time, demonstrates that osteoblastoma and the related entity, osteoid osteoma, are defined by structural rearrangements in the AP-1 family of genes, *FOS* and *FOSB*. This original work is the first report of a *FOS* mutation in a human bone-forming tumour since its discovery as one of the archetypal proto-oncogenes, forming the basis of a much-needed diagnostic test. Giant cell tumours (GCTs) of bone are characterised by an H3.3 gene mutation. I have explored a group of benign (GCTs), benign metastasising and malignant bone tumours which possess this mutation. Methylation profiling and evolutionary analysis suggest that malignant tumours have transformed from GCTs, acquiring replicative immortality or an additional epigenetic regulatory mutation. In contrast, my analyses show that benign metastatic disease can occur without any additional mutational changes.

Finally, I have studied complex mutational events more broadly in cancer and benign neoplastic disease. I reported chromothripsis and chromoplexy in malignant GCTs and osteoblastoma respectively for the first time. I explored the detection, frequency and evolutionary onset of chromoplexy in a collection of 2,626 human tumours. Found across almost all cancer types, the particularly striking and novel finding was the high frequency of chromoplexy in thyroid cancers, creating many of the known driver fusions.

Altogether, focusing on primary bone tumours, I have demonstrated how both simple and complex mutational events can define the earliest steps in tumour evolution. The analysis of complex patterns of mutation can also give new insights into the patterns of progression of both malignant and metastatic disease.

Impact Statement

Every cell in the body is defined by a blueprint within it, its genetic code. Cancer is a disease caused by errors in that genetic code. Some errors already reside in the code inherited from any individual's parents, but errors continue to develop throughout an individual's lifetime. The acquisition of these errors marks every cell as unique but setup a process of competition between cells for survival. This same competitive selection has also guided the evolution of species. Cancer develops when a collection of these errors allows a cell to replicate unrestrained, forming a tumour that can damage the tissues around it and spread to other parts of the body. The work presented in this thesis explores some of the patterns in those errors and how they have arisen in cancer, with a specific focus on primary bone tumours.

For the first time, in this work, I have analysed the genetic code of a rare bone tumour called osteoblastoma. This affects younger people and can cause significant pain and disability. Making an accurate diagnosis is critical to treat the tumour properly. Prior to this work, there could be difficulty in distinguishing osteoblastoma from osteosarcoma, another primary bone tumour but which, by contrast, is highly aggressive, requiring toxic chemotherapy to control systemic disease. I have defined a genetic error (mutation) in one of two related genes, *FOS* or *FOSB* that is found in virtually all cases of osteoblastoma. This will lead to further work to understand how these errors cause osteoblastoma to grow. This finding has also led to the first diagnostic test, helping to distinguish it from osteosarcoma.

In another disease called giant cell tumour of bone, I have explored some of the genetic differences that underlie more aggressive and potentially life-threatening behaviour of these tumours. Analysing the genetic changes in more aggressive tumours, which have the same mutation as giant cell tumours of bone, I have provided further evidence that they reflect a spectrum of the same disease. I have also uncovered the family of changes that allow these life-threatening tumours to change in behaviour, features which could, if confirmed in further studies, provide tests to guide the right treatments for patients.

Finally, I was the first to explore and time chromoplexy, a specific type of highly complex error that was originally identified in prostate cancer. I explored this in a large collection of 2,626 cancers, spanning 27 cancer types. I found it to be widespread across cancer types, including prostate cancer where these events were found to have developed early in tumour development. They also generated lots of the crucial mutations in thyroid cancers. These chromoplexy mutations produce a lot of the errors that cause a cancer to grow and are likely produced in one evolutionary leap. Recognising these events, as presented here, should instigate further work to find out how they occur and understand if there is a way to predict or potentially prevent them happening, thereby preventing some cancers from developing.

Acknowledgement

First and foremost, this work would not have been possible without the patients and their families' commitment. They contributed, knowing this work will never directly benefit them, but I hope that it has not been in vain.

I am also indebted to Cancer Research UK and all those who raise funds for them. None of this work, or the training I have received in doing it, would be possible without them.

On a personal level I am enormously grateful to my supervisory team of Peter Van Loo, Adrienne Flanagan, and my unofficial third supervisor, Sam Behjati. Each has set up and nurtured various parts of this work and guided my initial misguided steps into the world of genomics. I have to thank Adrienne for letting me be part of her inexhaustible search for the genetic causes of bone and soft tissue tumours, Sam for his laser sharp vision of what results mean, and Peter for his limitless patience. My thesis committee of Francesca Ciccarelli, Charles Swanton and Sandra Strauss have also provided valuable insights and direction. I could never have got started in computational analysis without the support of Kerstin Haase, Jonas Demeulemeester and Maxime Tarabichi. Jonas and Maxime, particularly, have guided me through our joint foray into the punctuated evolution of cancer, never making me feel the inferior partner I undoubtedly was. Annelien Verfaillie also provided significant support in the validation of *FOS* fusions in chapter 4 as well as invaluable early lessons in figure design.

My parents, Barbara and William, have supported me throughout a never-ending education, and have still ended up holding their granddaughter, Iona, while her parents have worked. I have never stinted in taking the credit for successes and blamed them for my failures, whilst most certainly the reverse was always true. Finally, Candida, my wife, has lived through every frustration and challenge and forgiven countless missed meals, weekends and holidays together. I dedicate this thesis to you all.

Table of Contents

Abstract	5
Impact Statement	7
Acknowledgement	9
Table of Contents	11
List of figures	15
List of tables	17
Abbreviations	18
Chapter 1. Introduction	19
1.1. Cancer as a genetic disease	19
1.1.1. The detection of mutations	20
1.1.2. The types of mutation	22
Structural variants	22
Copy number alterations	23
Substitutions	25
Insertions/Deletions (Indels)	26
Epigenetic variation.....	27
1.1.3. The impact of mutations	28
1.1.4. The patterns of mutation.....	29
1.2. Cancer as an evolving disease	31
1.2.1. The dynamics of cancer evolution	32
The clonal origin of tumours.....	32
Ongoing genetic divergence - intratumour heterogeneity	33
The pressure of selection.....	35
Co-evolution with the tumour micro-environment	35
1.2.2. The clinical relevance of cancer evolution	37
Evolution-guided therapy	37
Evolutionary monitoring	40
Cancer prevention, screening and stratification.....	41
1.3. Bone tumours as a case study of the patterns of mutation in tumour evolution	43
1.4. Thesis aims and objectives	46
Chapter 2. Materials & Methods	47
2.1. Chapter specific analysis flows	47
2.1.1. Chapter 3 methods	47
2.1.2. Chapter 4 methods	48
2.1.3. Chapter 5 methods	49
2.2. Patient samples	50
2.3. DNA and RNA extraction	50
2.4. SNP array analysis	50
2.5. Sequencing	51

2.6. Data analysis	51
2.7. Whole genome sequencing alignment and variant detection	51
2.8. Variant validation	52
2.9. Analysis of mutations in cancer genes.....	53
2.10.Fusion detection (Chapter 3)	53
2.11.FOS fusion validation	54
2.12.Allele-specific expression analysis.....	54
2.13.Fluorescence In-Situ Hybridisation (FISH) for FOS and FOSB.....	55
2.14.Immunohistochemistry.....	55
2.15.Copy Number Scoring (Chapter 4)	56
2.16.Mutation clustering, purity estimation and phylogenetic reconstruction	56
2.17.Simple timing analysis (Chapter 4)	57
2.18.Methylation array analysis and data pre-processing	58
2.19.Methylation data clustering.....	59
2.20.Methylation genomic analysis	59
2.21.Methylation based copy number analysis.....	60
2.22.Pan Cancer Analysis of Whole Genomes (PCAWG) variant calling.....	61
2.23.PCAWG structural variant detection and classification.....	61
2.24.Application of ChainFinder	62
2.25.Detection of deletion bridges.....	62
2.26.PCAWG mutation clustering and timing.....	63
2.27.Defining kataegis and chromothripsis.....	63
2.28.Punctuated event timing	64
Chapter 3. Osteoblastoma is defined by recurrent rearrangements in <i>FOS/FOSB</i>.....	67
3.1. Introduction	67
3.1.1. Clinical and epidemiological features of osteoblastoma	67
3.1.2. Pathology	68
3.1.3. Genomics.....	68
3.2. Aim	69
3.2.1. Objectives	69
3.3. Results	69
3.3.1. Osteoblastoma possess few somatic alterations.....	70
3.3.2. Recurrent FOS structural rearrangements	73
3.3.3. FOSB rearrangement by chromoplexy	76
3.3.4. The impact of <i>FOS</i> and <i>FOSB</i> rearrangement	79
3.3.5. Validation of <i>FOS</i> or <i>FOSB</i> alterations.....	80
3.3.6. <i>FOS</i> or <i>FOSB</i> rearrangement is unique to osteoblastoma amongst bone-forming tumours	81
3.4. Discussion	82
3.4.1. <i>FOS</i> and <i>FOSB</i> alteration as diagnostic biomarkers of osteoblastoma	82
3.4.2. <i>FOS</i> rearrangement resembles the murine retroviral <i>v-fos</i>	83
3.4.3. Mutational recurrence suggests common disease entities and evolutionary origins	84
Chapter 4. H3.3-mutant bone tumours – the pattern of progression.....	87

4.1. Introduction	87
4.1.1. Conventional giant cell tumours of bone.....	87
4.1.2. The genomics of GCT.....	89
4.1.3. Histone Coding Variants	90
K27M Mechanism	92
G34R/V Mechanism.....	92
K36M Mechanism	92
G34W Mechanism	93
4.1.4. Metastatic giant cell tumours of bone	93
4.1.5. Malignant H3.3 bone tumours.....	94
4.2. Aim	96
4.2.1. Objectives	96
4.3. Results	96
4.3.1. The genomic landscape of H3.3 mutant bone tumours	96
4.3.2. Late aneuploidy and genome duplication mark malignant H3.3 tumour progression	104
4.3.3. Case Study – H3.3 mutation potentially lost in malignant progression	110
4.3.4. Malignant H3.3 bone tumours have a distinct methylation profile	113
4.3.5. Benign H3.3 tumours have a hypermethylated <i>CCND1</i> promoter	118
4.3.6. GCT metastases are polyclonally seeded	122
4.4. Discussion	124
4.4.1. Malignant H3.3 tumours acquire additional driver mutations.....	124
4.4.2. Malignant H3.3 tumours have a distinct methylation profile	125
4.4.3. Do malignant H3.3-mutated tumours transform from GCTs?	126
4.4.4. Benign metastatic GCTs resemble conventional GCT and may be polyclonally seeded.....	127
Chapter 5. Identification and timing of chromoplexy across cancer types	129
5.1. Introduction	129
5.1.1. A mechanism and definition of chromoplexy	129
5.1.2. Chromoplexy in prostate cancer	131
5.1.3. Chromoplexy in other cancers	131
5.1.4. Methods to identify chromoplexy	132
5.2. Aim	133
5.2.1. Objectives	133
5.3. Results	133
5.3.1. The Pan Cancer Analysis of Whole Genomes (PCAWG) dataset.....	133
5.3.2. The ChainFinder algorithm finds clusters of structural variants across all cancer types	135
5.3.3. ChainFinder chains are predominantly short and lack the expected deletion bridges.....	136
5.3.4. Many ChainFinder chains represent chromothripsis or retrotransposon insertions	138
5.3.5. Breakpoint and copy number configurations define rearrangement events	141

5.3.6. The appearances of chromoplexy and templated insertion footprints can overlap	142
5.3.7. Chromoplexy involves recurrent disease specific driver genes	146
5.3.8. Differentiated thyroid cancer driver rearrangements are predominantly produced by complex rearrangements and chromoplexy 148	
5.3.9. Mutational timing of events suggests chromoplexy is an early and clonal event.....	150
5.4. Discussion.....	153
Chapter 6. Discussion.....	155
6.1. Key Findings.....	155
6.2. Cross-cutting themes	157
6.3. Future work.....	159
Chapter 7. Appendix.....	161
7.1. Chapter 3 Appendix	161
7.1.1. DNA discordant reads supporting <i>FOS/FOSB</i> rearrangement.....	161
7.1.2. DNA discordant reads supporting <i>FOS/FOSB</i> rearrangement.....	162
7.1.3. Validation of <i>FOS</i> rearrangements	163
7.1.4. Discovery cohort sequencing results and <i>FOS/FOSB</i> rearrangements 165	
7.1.5. Discovery cohort copy number profiles.....	166
7.1.6. Immunohistochemistry and histology images	175
7.1.7. Validation cohort immunohistochemistry and FISH results	179
7.1.8. Allele specific expression results	181
7.2. Chapter 4 Appendix	189
7.2.1. SNP Samples	189
7.2.2. Methylation array samples.....	190
7.2.3. Additional Timing Figures	192
7.3. Chapter 5 Appendix	194
7.3.1. PCAWG cohort	194
7.4. Related authored papers	197
Reference List	230

List of figures

Figure 1 Schematic of Structural Variant Types.	23
Figure 2 Copy number example.	24
Figure 3 Substitution example.	26
Figure 4 Indel example.	27
Figure 5 Schematic of adaptive therapy.	39
Figure 6 Chapter 3 discovery cohort.	47
Figure 7 Validation cohort analysis.	48
Figure 8 H3.3 mutated tumour sequencing analysis.	48
Figure 9 Methylation analysis.	49
Figure 10 SNP array analysis.	49
Figure 11 Chromoplexy workflow.	49
Figure 12 Osteoblastoma mutation burden compared with osteosarcoma.	70
Figure 13 Osteoblastoma mutation burden compared with all cancer types. ...	71
Figure 14 Chromosome 22 copy number data for PD13480.	72
Figure 15 <i>FOS</i> breakpoints in osteoblastoma.	75
Figure 16 Validation of <i>FOS/FOSB</i> rearrangements.	76
Figure 17 PD7525 Chromoplexy around <i>FOSB</i>	77
Figure 18 Alternative chromoplexy derivative configurations.	78
Figure 19 <i>FOS</i> rearranged transcripts.	80
Figure 20 <i>FOSB</i> rearrangements.	80
Figure 21 Validation for <i>FOS</i> alteration.	81
Figure 22 Landscape of H3.3 mutant tumours.	100
Figure 23 PD38328a <i>KDM4B</i> homozygous deletion.	101
Figure 24 PD4922e Circos plot of genome wide copy number and structural variants.	102
Figure 25 Circos plot of chromothripsis around <i>TERT</i> in PD3788d.	103
Figure 26 Example copy number profiles and scores.	105
Figure 27 Validation cohort of CN scores.	106
Figure 28 SNP array copy number profiles from a GCT and its recurrence. ...	107
Figure 29 Methylation array based copy number scores.	107

Figure 30 Timing of PD4922e.....	109
Figure 31 PD38328a biopsy GCT histological images.	110
Figure 32 PD38328a resection malignant histology.	111
Figure 33 PD38328a circos plot of copy number and structural rearrangements.	112
Figure 34 H3.3 gene LOH in osteosarcoma and undifferentiated sarcoma....	113
Figure 35 Methylation based unsupervised clustering.....	114
Figure 36 Genomic features of differentially methylated sites.	115
Figure 37 Gene set enrichment analysis for methylation data.....	116
Figure 38 Differential methylation across histone clusters.	117
Figure 39 Differentially methylation around olfactory gene clusters.	118
Figure 40 Focal differentially methylated regions (DMRs).....	119
Figure 41 CCND1 differential methylation.	120
Figure 42 CCND1 promoter methylation across sarcoma types.	121
Figure 43 CCND1 promoter methylation in clustered tumours.	121
Figure 44 Mutation clustering reveals the pattern of metastatic seeding in benign- metastasising GCT.	123
Figure 45 PCAWG ChainFinder chains.....	136
Figure 46 Breakpoint adjacency distance density plot.....	137
Figure 47 Example deletion bridge.....	138
Figure 48 Chromothripsis-like ChainFinder chain.....	139
Figure 49 Retrotransposon-like ChainFinder Chain.	140
Figure 50 Footprint schematic.	143
Figure 51 Schematic of templated insertion phasing support.....	144
Figure 52 Schematic of inverted orientation balanced breakpoint footprints. .	145
Figure 53 PCAWG Chromoplexy proportions by disease type.....	146
Figure 54 Chromoplexy footprint drivers.....	147
Figure 55 Thyroid cancer chromoplexy chain.....	148
Figure 56 Circos plot of thyroid cancer breakpoints.	149
Figure 57 Timing of PCAWG punctuated events.....	152

List of tables

Table 1 Classification of bone tumours.....	45
Table 2 Sequencing algorithms.	52
Table 3 <i>FOS</i> fusion validation primers.....	54
Table 4 Clinical characteristics of discovery cohort.	70
Table 5 Structural rearrangements in osteoblastoma genomes identified by the genomic structural variation caller BRASS.	74
Table 6 Cosmic histone mutations.....	91
Table 7 Clinical characteristics of whole genome sequenced H3.3 mutant bone tumours.....	97
Table 8 Abbreviated PCAWG cohort.	134
Table 9 PCAWG SV classifications of ChainFinder chains.	142

Abbreviations

bp	Base pair
CCF	Cancer Cell Fraction
CNA	Copy Number Aberration
CT	Computed Tomography
CTC	Circulating Tumour Cell
ctDNA	Circulating Tumour DNA
DMP	Differentially Methylated Probe
DMR	Differentially Methylated Region
DNA	Deoxyribonucleic Acid
GCT	Giant Cell Tumour
GSEA	Gene Set Enrichment Analysis
LOH	Loss of Heterozygosity
MCN	Mutation Copy Number
MDS	Multi-Dimensional Scaling
MRI	Magnetic Resonance Imaging
PCA	Principal Component Analysis
PCAWG	Pan-Cancer Analysis of Whole Genomes
PCR	Polymerase Chain Reaction
RNA	Ribonucleic Acid
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
SV	Structural Variant
TCN	Total Copy Number
tSNE	T distributed stochastic neighbour embedding
VAF	Variant Allele Fraction
WGD	Whole Genome Duplication

Chapter 1. Introduction

Life expectancy in the developed world now exceeds the span optimised by natural selection (ONS, 2015, DeGregori, 2011). Concomitant with this has been the increasing incidence of cancer and neurodegenerative diseases, becoming leading causes of death (ONS, 2017). Billions of dollars have been invested in cancer research, however the greatest impact on outcomes across cancer types has been yielded from the introduction of safe surgery, radiotherapy and conventional chemotherapy. These improvements were largely introduced based on empirical observations alone (Mukherjee, 2011). Insights into numerous facets of cancer biology, such as the relationship of the microbiome, host immune response, the tumour microenvironment and cancer cellular metabolism have instigated new therapeutic research and delineated the divergence of the phenotype of a cancer from its host (Weinberg, 2014). Fundamentally however, cancer is a disorder of genotypic divergence and its progression is a continuation of evolutionary processes on a cellular scale.

This thesis aims to explore the range of complexity in changes to the tumour genome and how they define tumour evolution. A focus will be on bone tumours, which, though rare, exhibit much of the spectrum of mutational patterns seen across other cancer types.

1.1. Cancer as a genetic disease

The genetic basis of cancer is now widely accepted (Lawrence *et al.*, 2019, Weinberg, 2014, Kumar, 2015). Our current genetic understanding of cancer can be traced to four landmarks that span a century, beginning at the turn of the 20th century. Theodor and Marcella Boveri's observations, in their 1914 monograph 'The Origin of Malignant Tumours', localised the root of malignant tumour behaviour to the nucleus and to specific combinations of chromosomes (Boveri and Boveri, 1914, Boveri, 2008):

"We begin by assuming that the properties of malignant cells are due to an inherent defect... we have compelling reasons for assuming that the individual chromosomes of metazoan nuclei have different properties. Their differences are

not only quantitative, as appears to be the case in protozoa, but also qualitative. They are specialised in different ways and can only maintain the normality of the cell if they are present in a certain specific combination.”

The elucidation of the molecular structure of the chromosomes, their constituent DNA, by Crick, Franklin, Watson and Wilkins (Watson and Crick, 1953, Franklin and Gosling, 1953, Wilkins *et al.*, 1953) paved the way to understand the nature of the nuclear changes underlying cancer. The first specific example that changes in host DNA could underlie malignant transformation was provided by Harold Varmus and Michael Bishop in 1976. This stemmed from work 65 years earlier by Peyton Rous discovering the Rous Sarcoma Virus (RSV), a transmissible element that could induce sarcoma in chickens (Rous, 1910). Bishop and Varmus discovered that the viral RSV gene responsible for tumour growth, *v-src* was highly homologous to normal avian DNA (Stehelin *et al.*, 1976). This avian gene was named *c-src*, with its homologue in humans, *SRC*. Finally, a quarter of a century later, the completion of the human genome project in 2003 provided the full genome map to delineate the genetic changes in cancer.

The full catalogue of genetic changes or mutations that cause cancer is still far from complete. There are however several different methods by which they can be classified, either guided by their genomic appearance, which is implicitly influenced by how they are detected, or by their functional impact on the cell.

1.1.1. The detection of mutations

The study of mutation has advanced immeasurably in both scale and resolution since the observations made by Boveri and others regarding aberrant chromosome numbers in cancer cells. These early findings were made using plain light microscopy, predominantly in model organisms such as nematodes and sea urchins. For much of the 20th century, genetic discoveries were guided by three principle approaches: direct cytogenetic observations, low-resolution linkage analyses and focussed biological assay of genetic sequence of interest. Cytogenetic observations, with improvements in microscopy resolution, and DNA labelling techniques, such as fluorescent-in situ hybridisation, could define changes in individual cancer cells but only at the scale of the chromosomes. This could clearly define critical recurrent translocations. Most famous is the

Philadelphia chromosome which creates the *BCR-ABL1* fusion as one half of the reciprocal balanced $t(9;22)(q34;q11)$ translocation, seen in 90% of chronic myeloid leukaemia (Nowell and Hungerford, 1960, Rowley, 1973).

Linkage analysis localised cancer genes inherited in familial syndromes but required markers that were physically mapped in the genome. For a long while these were restriction fragment length polymorphisms (RFLPs) but they required extensive manual laboratory work (Strachan and Read, 2010). They have become replaced by single nucleotide polymorphisms (SNPs), assayed by array hybridisation with the potential to interrogate millions of loci in one experiment. Informed by such low-resolution techniques, painstaking techniques could identify sequence variants and their functional impact but only at very local scale. Next-generation sequencing techniques have transformed the ability to identify cancer variants from the human reference genome. This could now be performed at the scale of the whole genome and across large numbers of individuals (Meyerson *et al.*, 2010). Most studies have restricted the genomic resolution to panels of genes of interest or coding sequences (exome) but decreasing cost allows large datasets of whole cancer genomes to be collated (Priestley *et al.*, 2019, Turnbull, 2018, PCAWG, 2020). Accompanying approaches have extended analysis to other layers or scales of genetic changes including: non-sequence mediated heritable changes (the epigenome) (Li and Tollefsbol, 2011, Feinberg *et al.*, 2006), the expressed genome or the transcriptome, larger blocks of the genome with linked-read and long read sequencing (Viswanathan *et al.*, 2018), or at the individual cellular scale with single-cell sequencing (Wang *et al.*, 2014, Navin, 2015).

Accompanying the technical ability to sequence the genome has been the raft of computational tools required to analyse this wealth of data. These have continuously improved but still require expert use, frequent adaptation and multiple tools to generate the catalogue of mutations for any sequenced sample (**Methods 2.7**).

1.1.2. The types of mutation

The genetic changes underlying cancer are either inherited and therefore present in every cell or acquired in individual somatic cells and passed onto their progeny. Tumour evolution relies on the accumulation of new mutations, this thesis focuses on somatic variants, though the following classification applies broadly to both. The classification of mutations by their type is implicitly influenced by historical methods to detect them. Based on current sequencing methods they can be divided into larger scale copy number and structural variants, sequence-level substitutions and small insertions/deletions (indels), and epigenetic changes.

Structural variants

Gross chromosomal changes were the first variants to be recognised as they could be directly visualised with microscopy during metaphase and are the basis of cytogenetics. In next-generation sequencing analysis, structural variants are identified using sequenced fragments not aligned to the reference genome as expected (Cameron *et al.*, 2019) (**Figure 1**). Paired-end sequencing provides sequence information (reads) at the ends of DNA fragments. If paired reads align to different parts of the genome they are 'discordant' and most structural variant detection rely on this information. Individual reads that partially align to two different parts of the genome, are called 'split' or 'spanning' reads, and these therefore inform about the exact basepair position of a structural variant. Many structural variant callers will also conduct *de novo* assembly of all reads mapping nearby a structural variant to identify new sequence; inserted or deleted at the breakpoint. Structural variants are usually subclassified into four main classes: translocations, which are inter-chromosomal structural variants, tandem-duplications, deletions and inversions. All of these have relevance to cancer because they can have a marked impact on the sequence and regulation of genes. This can either be deleterious to gene function by disrupting or deleting them, amplify the gene's function by increasing the copy number or changing its promoter or enhancer function, or alternatively by creating entirely new fusion gene sequences.

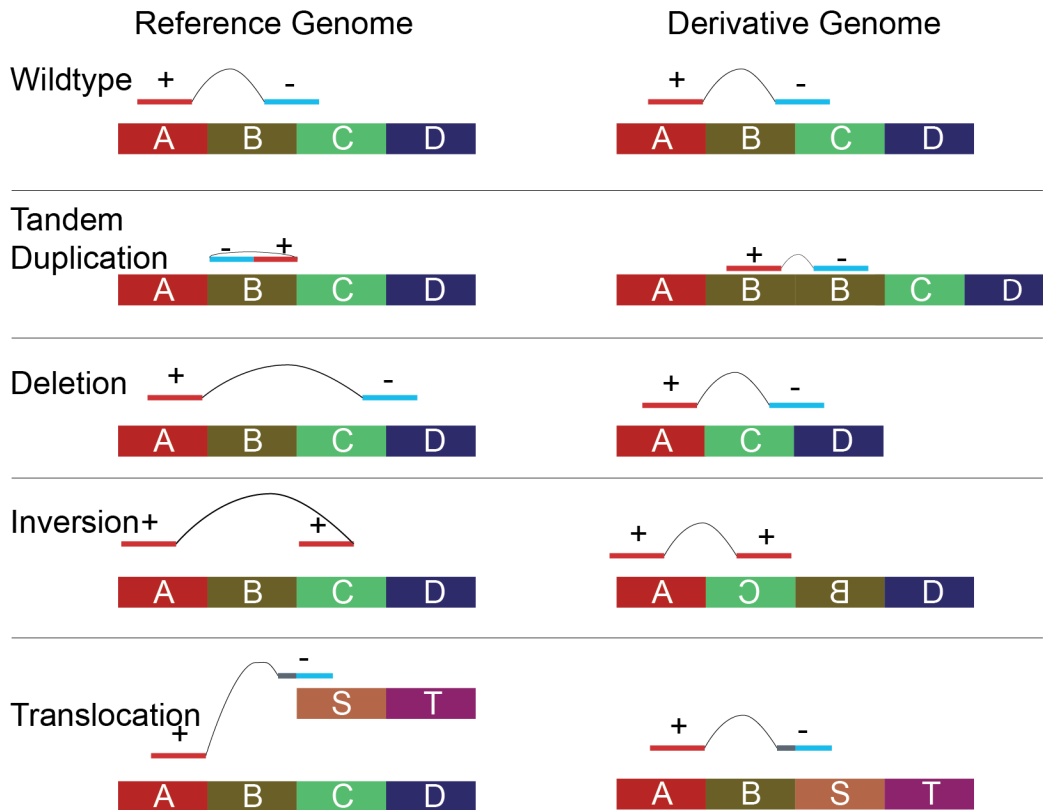


Figure 1 Schematic of Structural Variant Types. Each rearrangement type is shown both with its appearance relative to the reference genome and the true derivative genome. Note the derivative genome is also inferred from the reference. Reads have an orientation indicated whether the sequence required reverse-complementation during alignment (red (+) reads are reference sequence while blue (-) reads were reverse complemented). Reads that only partially align will be 'clipped' (shown in grey), but this sequence will be interpreted either from chimeric alignment or assembly, depending on the aligner and rearrangement caller respectively. Note that both the inversion and translocation examples, as shown, are unbalanced unless there are further rearrangements detected: BD and RC respectively.

Copy number alterations

Aneuploidy, which is any deviation from the normal chromosomal contingent, is frequent in cancer. Changes in the number of copies of chromosomes or parts of chromosomes are primarily informed by the number of sequencing reads (coverage) aligning at any one location (**Figure 2**). The copy number of each chromosomal allele can be inferred from the relative frequency of single nucleotide polymorphisms (SNPs) which were inherited in a heterozygous fashion; the B-allele frequency (BAF). Amplifications of oncogenes and deletions of tumour suppressors (**Introduction 1.1.3**), either of both copies (homozygous)

or in combination with another deleterious mutation, are frequent in cancer genomes.

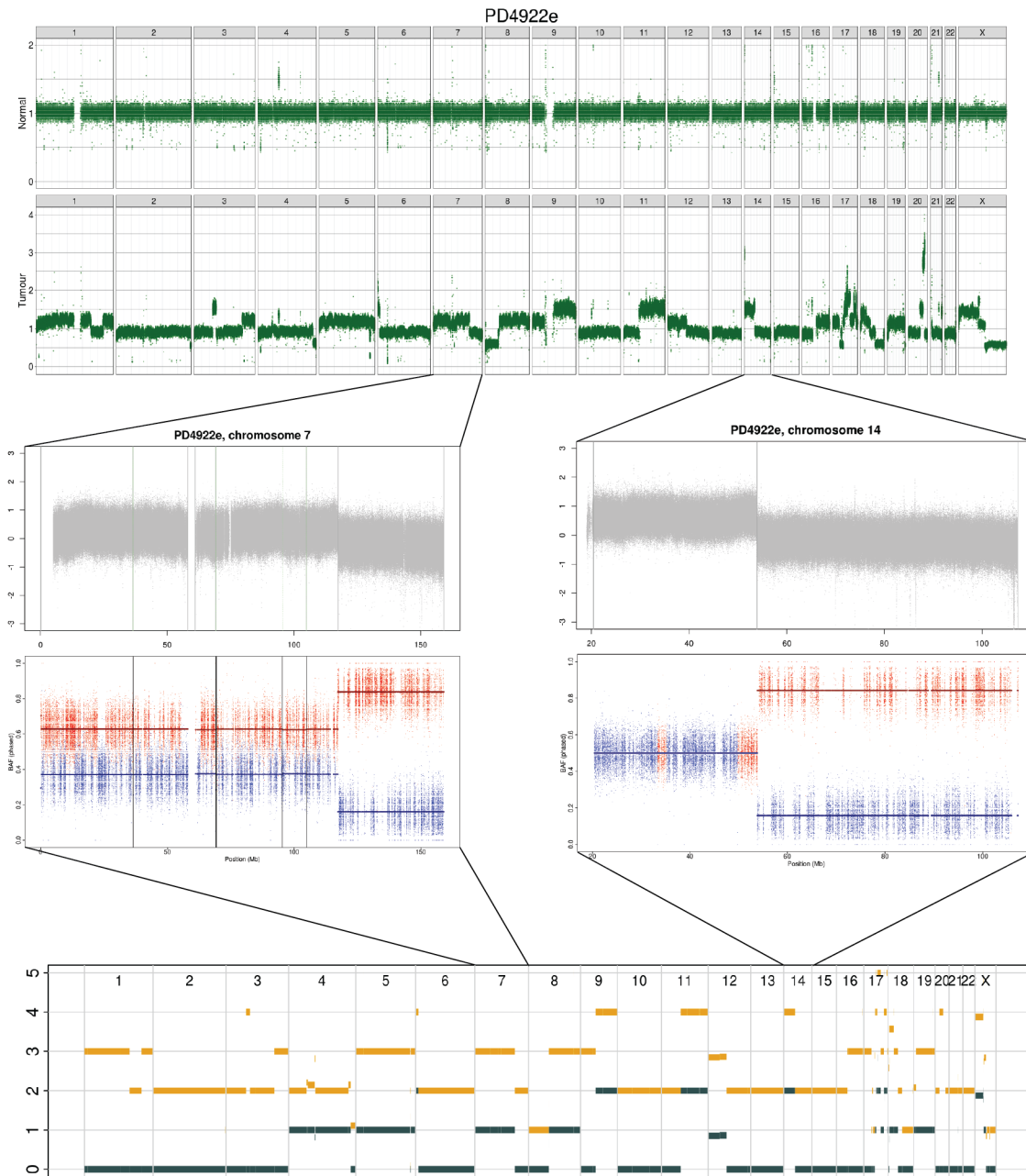


Figure 2 Copy number example. A copy number workflow using the Battenberg algorithm (**Methods 2.7**). Top to bottom: Normal and tumour coverage across the whole genome, LogR (log normalised ratio of tumour:normal coverage) plot for selected chromosomes, the B-Allele frequency for SNPs across these selected chromosomes (the SNPs are coloured according to the parental allele on which they are carried), the average whole genome copy number plot, with total copy number (orange) and minor allele (black). Note that chromosome 7 is predominantly allelically imbalanced (2+1) but 7q has LOH (2+0) with a wider BAF separation because of the greater allelic imbalance (not at 1/0 because of normal tissue contamination) and lower LogR. Chromosome 14p is allelically balanced but genome duplicated (2+2) with a BAF of 0.5 while 14q has LOH (2+0).

In most, but not all circumstances, rearrangements involve copy number changes and *vice-versa*, therefore copy number analyses can integrate pre-computed structural rearrangements (Dentro *et al.*, 2018). There are significant exceptions to this overlap and technical reasons why both methods do not always detect them.

Rearrangements can result from DNA damage that is conservatively repaired and therefore involve no genomic loss; commonly seen in reciprocal translocations (**Results 5.3.6**). Alternatively, copy number calling methods are relatively insensitive to short copy number changes (**Results 5.3.3**) that may still be detectable by rearrangement methods. Conversely whole chromosomal or whole genome level copy number changes will not have associated structural rearrangements.

Substitutions

Single nucleotide variants (SNVs) are the most common variants seen in cancer genomes, typically ranging from 0.1-100 SNVs/MB (Lawrence *et al.*, 2013). These are relatively simply detected as differences from the reference genome but require extensive filtering to exclude technical sequencing artefacts or normal genetic variation (**Figure 3**). Common normal variants are easily filtered but others are excluded by sequencing a paired 'normal' tissue specimen. The impact of SNVs is best understood in the coding genome where they may induce a change in the protein amino acid sequence (non-synonymous). This can have a functional impact on the gene, for example the classical *KRAS*:p.G12D substitution causing constitutive activation in many colorectal cancers (Valenzuela and Groffen, 1986, Bos *et al.*, 1987).

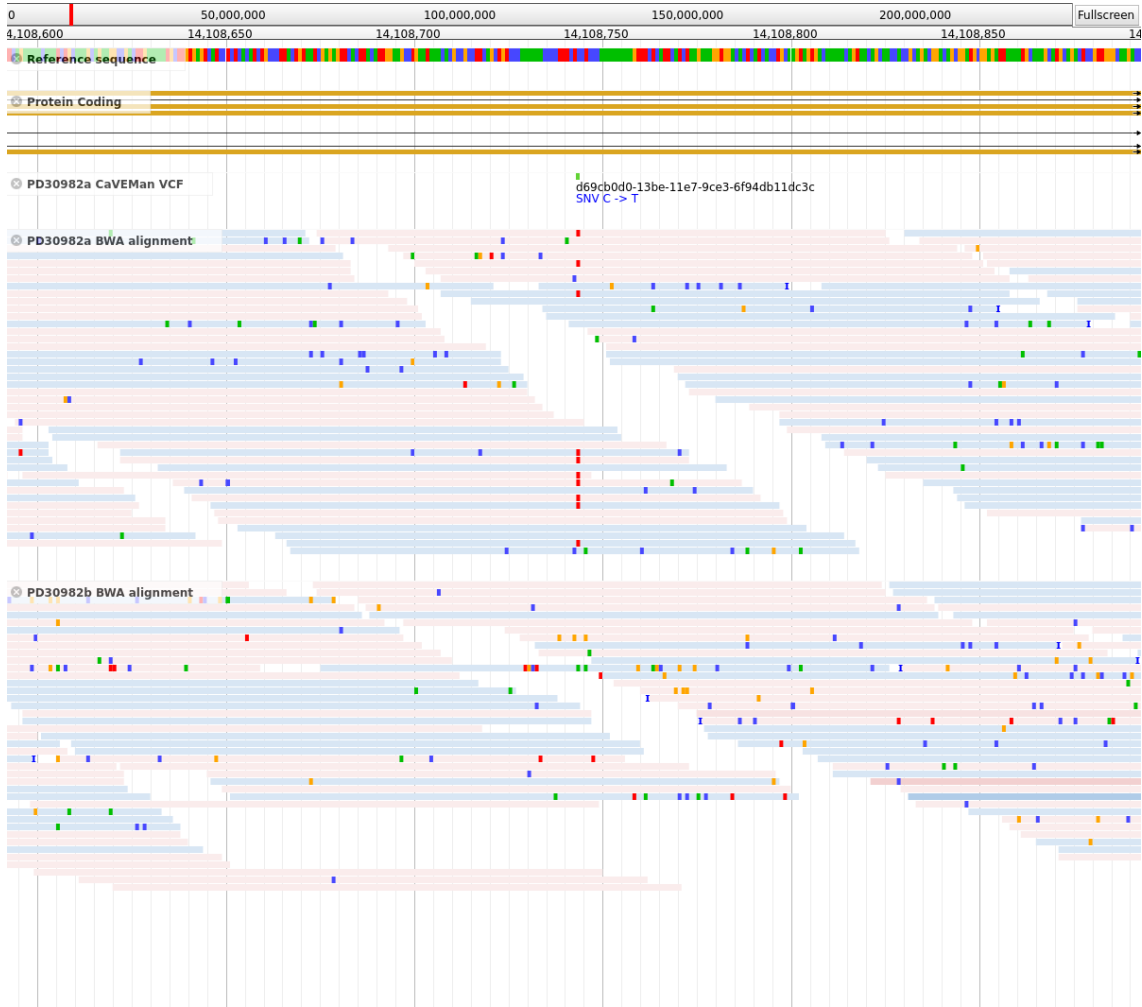


Figure 3 Substitution example. Image of sequencing reads (from the genomics browser, JBrowse) showing the sequencing reads supporting the *H3F3A* G34W mutation in PD30982a (**Results 4.3.1**). Reference sequence and location are shown at the top as well as the multiple alternative transcripts for the *H3F3A* gene. The substitution can be seen by the multiple red bars seen in multiple reads of both orientation (red and blue read colouring). These are seen in the tumour sample (PD30982a) but not the paired normal/germline sample (PD30982b).

Insertions/Deletions (Indels)

Indels, typically of <50bp are detected using read-pairs in which one read is initially unmapped. The remaining read is remapped in fragments and then the assembly of the remaining unmapped fragments can reconstruct the indel, this is done in part by both primary aligners and specialist indels callers (**Figure 4**). Indels are challenging to detect but less frequent in cancer, though they can also lead to significant changes in the coding genome, particularly where they induce a frameshift in the coding sequence. This is a frequent mechanism of inactivation

of tumour suppressor genes (**Introduction 1.1.3**) (Zehir *et al.*, 2017, Nik-Zainal *et al.*, 2016).

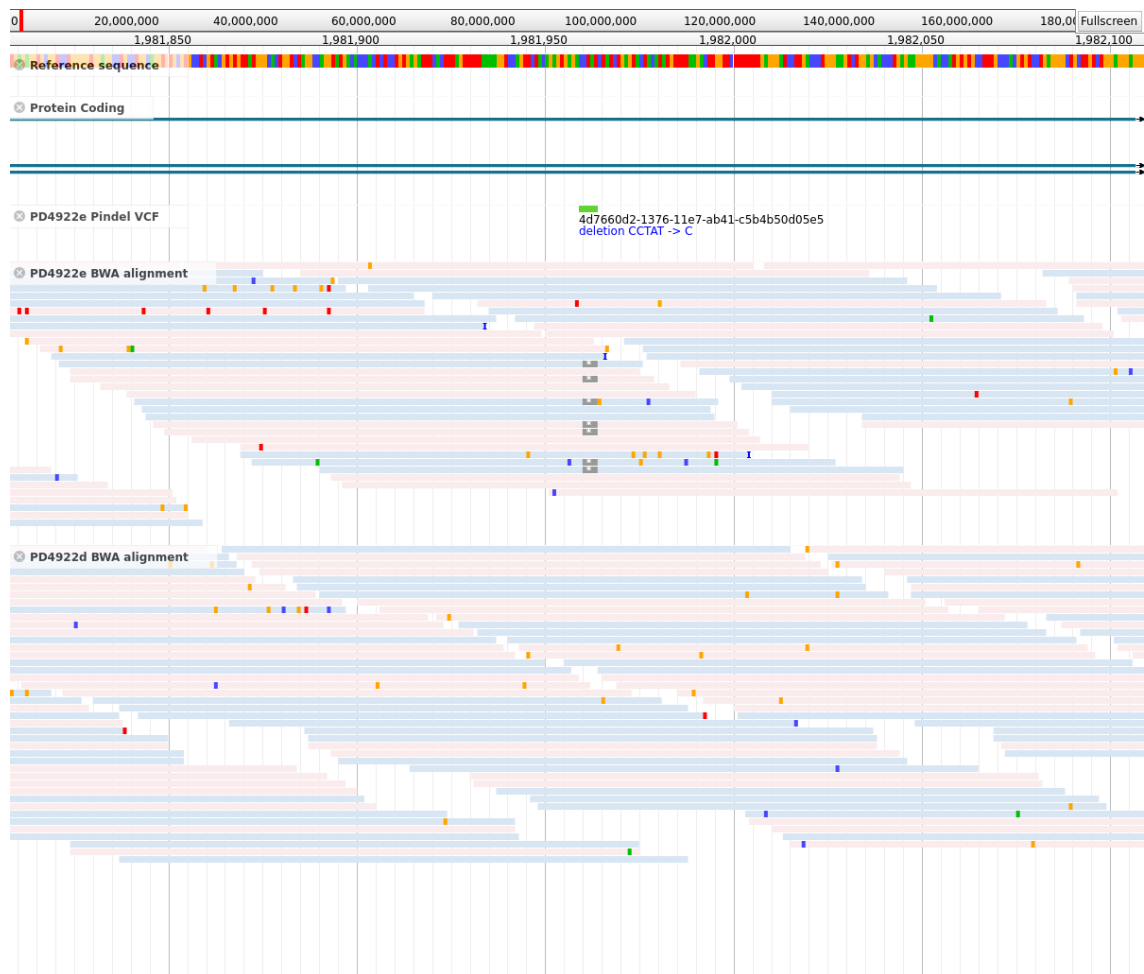


Figure 4 Indel example. Figure format is as for Figure 3. The small deletion of the sequence ‘CTAT’ is shown in multiple reads by the grey bars in the tumour sequencing reads (PD4922e). This deletion was not a normal variant as it was not seen in the germline sample (PD4922d).

Epigenetic variation

The epigenome, defined as the catalogue of non-sequence-related heritable traits, is also susceptible to important variants in cancer. Epigenetic states are preserved during cell division and therefore may be transmitted through a lineage of cells. Components of epigenetic regulation are themselves genetically encoded and therefore susceptible to genetic mutation. For example, point mutations in the genes encoding an essential chromatin-associated proteins, the histones, are mutated in all giant cell tumours of bone (**Results 4.1.3**). Histone modifying enzymes, such as *SETD2*, are also frequently mutated (Dalglish *et*

al., 2010). Epigenetic dysregulation can have a significant functional impact in a cancer cell, through the regulation of gene expression, but may not have associated sequence changes and therefore represents a class of variation in its own right. The best studied is direct DNA covalent modification by methylation or covalent modification by methylation or acetylation of the chromatin-associated proteins, the histones. DNA methylation is in general associated with the repression of gene expression, while the impact of histone modifications depend on the type of modification and the exact protein residue affected (Flavahan *et al.*, 2017). For example the tumour suppressor genes *MLH1* (Esteller *et al.*, 1998) and *CDKN2A* (Merlo *et al.*, 1995) have both been found to be silenced by the methylation of their sequence.

1.1.3. The impact of mutations

Mutations can also be considered from the perspective of their impact on cell function. Those that are irrelevant to cell function, are considered passengers. These mutations can still be highly informative to understand the clonal relationship of cells, because unless a locus is deleted, all progeny of a cell will possess its mutations (**Introduction 1.2.1**). Conversely, mutations that denote a functional advantage to a cell, contributing to its ability to outcompete its neighbours and develop into a tumour, can be considered 'drivers'. Driver genes can be further subdivided into those that once mutated promote tumour growth, oncogenes, or those that ordinarily repress tumour growth, tumour suppressors. Once again, such a division was foreseen by Boveri, albeit considered at the level of the chromosome (Boveri and Boveri, 1914, Boveri, 2008):

"...there are chromosomes that stimulate cell multiplication...The unrestrained proliferation of malignant tumour cells would then be due to a permanent excess of these stimulatory chromosomes"

"Another possibility is that there is a specific inhibitory mechanism in every normal cell that only permits cell division to take place when this mechanism is overcome by some special stimulus. It would accord with our basic concept if one assumed that there were specific chromosomes that inhibited cell division...A tumour cell that proliferated without restraint would be generated if these 'inhibitory chromosomes' were eliminated"

The functional capabilities of a tumour are clearly more diverse than merely the ability to proliferate ‘without restraint’ and include the ability to invade and metastasise. The various capabilities, endowed by genetic and epigenetic mutation are known as the ‘hallmarks of cancer’ (Hanahan and Weinberg, 2011). These ten hallmarks are each fields of study in themselves and their broad discussion is beyond the scope of this thesis. Particularly relevant hallmarks are the ability to generate genome instability and mutation, which is likely responsible for the development of tumour aneuploidy (**Results 4.3.2**), and replicative immortality (**Results 4.3.1**).

1.1.4. The patterns of mutation

A further classification of mutations, much of which this thesis is focussed upon, is their pattern in the genome. Thus far most mutations have been considered in isolation, with regards their type or functional impact on the cell. Clearly the assortment of potential mutations within and between tumours is highly varied, as once again Boveri already described:

“The fact that there may be countless different abnormal chromosome combinations, of which the vast majority are, in our view, incompatible with the survival of the cell, provides us with a simple explanation for the varied nature of the malignant tumours that arise in the one tissue of origin”

Increasingly, mutations are recognised to be acquired either as a result of a distinct process or in complex and simultaneous patterns. Mutational signatures have been defined as patterns of particular mutations with common features. Most explored are substitution signatures, which are often defined by their sequence change and trinucleotide sequence context (Alexandrov *et al.*, 2013). They have been identified from large datasets, using non-negative matrix factorisation (NMF), to identify a number of co-occurring mutational types, labelled a signature. Experimental validation has attributed processes causing some of these mutational signatures, for example spontaneous deamination occurring in all cells, producing C>T at CpG contexts, C>A mutations caused by benzopyrenes in cigarette smoke and C>T mutations creating pyrimidine dimers by UV radiation (Petljak *et al.*, 2019). Equivalent mutation signatures now exist for dinucleotide substitutions, indels, as well as structural variants and copy

number changes (Alexandrov *et al.*, 2019, Li *et al.*, 2020, Steele *et al.*, 2019, Macintyre *et al.*, 2018, Nik-Zainal *et al.*, 2016). Distinct from this are complex patterns of mutation, normally occurring in close genomic proximity. The number of mutations within these clusters is too great for them to occur independently, therefore they are thought to result from a single event. The earliest pattern described was the breakage-fusion-bridge cycle, initially described by Barbara McClintock in Maize (McClintock, 1938). These cycles, thought to occur after telomere erosion and end-to-end chromosomal joining, lead to patterns of duplication, inversion and potentially inter-chromosomal translocation and develop over a small number of cell cycles. These events have been found in human cancers, contribute to chromosomal instability and induce driver amplification (O'Hagan *et al.*, 2002, Rudolph *et al.*, 2001).

Patterns of co-localised substitution and structural rearrangements have also been observed. Substitution clusters, named kataegis, ordinarily involve the same change in DNA sequence and are often associated with nearby breakpoints. Most kataegis is thought to result from the aberrant activity of the single-stranded DNA cytosine deaminases of the APOBEC family (Taylor *et al.*, 2013, Swanton *et al.*, 2015, Nik-Zainal *et al.*, 2014, Nik-Zainal *et al.*, 2012b).

There are two patterns of clustered structural rearrangements both of which have particular relevance to cancer and this thesis, and are introduced further in Chapter 5. Chromothripsis is a pattern of clustered rearrangements on one, or a small number of chromosomes, associated with an alternating pattern of copy number, and a random pattern of rearrangements. This pattern was initially discovered in a single case of chronic lymphocytic leukaemia but has been observed across many cancer types including a significant proportion of the malignant bone tumour, osteosarcoma (Stephens *et al.*, 2011, Maher and Wilson, 2012, Govind *et al.*, 2014, Behjati *et al.*, 2017). Chromothripsis also can combine with breakage-fusion-bridge cycles in well-differentiated and dedifferentiated liposarcoma to create neochromosomes containing the highly amplified oncogenes *MDM2* and *CDK4* (Garsed *et al.*, 2014). Finally, chromoplexy, the focus of Chapter 5, is a seemingly less random pattern of rearrangements that was initially reported in prostate cancer. Multiple rearrangements often involving

genes and spanning multiple chromosomes are thought to result from a series of simultaneous double stranded DNA breaks which are erroneously repaired (Baca *et al.*, 2013). When considered with respect to the reference genome chromoplexy can appear as a chain or cycle of rearrangements, though in reality these represent distinct inter-chromosomal translocations. The resultant karyotype would possess a series of derivative chromosomes that result from these translocations.

1.2. Cancer as an evolving disease

Genetic changes are the basis of variation on which natural selection acts. Cancer results from the inevitable continuation of genetic evolution at the cellular level within an individual. Hypothetically, an evolutionary family tree could be traced from the earliest primordial self-replicating molecule, through the development of species and down to the cells within an individual. The somatic cellular evolution that occurs within an individual organism is largely disregarded from one generation to the next. Each organism contributes only a single cell, the gamete, to the generation of a new organism therefore most somatic diversification is lost. This also highlights the one genetic process at play in the evolution of species that does not impact on cellular evolution, namely genetic recombination through sexual reproduction.

The application of evolutionary theory to cancer was first proposed by Peter Nowell (Nowell, 1976). This has two essential components: all cancers have a clonal origin, that is their ancestry can be traced back to a single cell of origin and their cellular composition is guided by the evolutionary principles of variation and selection. It is important to recognise, that evolution of tumours cells does not occur in isolation but in competition and interaction with other cells in its micro-environment.

1.2.1. The dynamics of cancer evolution

The clonal origin of tumours

The evidence for the clonal origin of tumours existed prior to Nowell's theory in 1976. Three observations of tumour cells support their clonal ancestry. Tumour cells often share a number of aberrations unlikely to have been acquired independently, such as the Philadelphia chromosome in CML (A A Sandberg and Hossfeld, 1970). The inactivation of an X chromosome in females is normally random in somatic tissues however cancers that arise in females have the same copy of an X chromosome inactivated (Linder and Gartler, 1965). Finally, somatic recombination and hypermutation of the immunoglobulin locus generates huge diversity of immunoglobulin repertoire. Immunoglobulins are therefore unique to a clonal lineage of somatic cells. Malignancies that produce immunoglobulins all produce the same immunoglobulin (Preud'homme and Seligmann, 1972).

Clonal expansions occur in normal tissues with a surprising frequency, likely reflecting the high rate at which the early precursors of cancer arise. The prognostic significance of these populations is not always certain. Martincorena *et al.* (Martincorena *et al.*, 2015, Martincorena *et al.*, 2018) identified multiple clonal expansions of cells containing mutations in *TP53*, *NOTCH1* and other recognised cancer genes, in both normal sun-exposed eyelids and in aging normal oesophagus. Interestingly, in two studies of normal oesophagus (Martincorena *et al.*, 2018, Yokoyama *et al.*, 2019), mutations were much more common in *NOTCH1* than *TP53*, the inverse of the pattern seen in oesophageal cancer. This suggests a number of interesting interactions between these two gene mutations, including that early *NOTCH1* mutations may protect against subsequent *TP53* mutation and cancer development. Demeulemeester *et al.* (Demeulemeester *et al.*, 2016) analysed epithelial cells found in bone marrow aspirates of breast cancer patients, identifying cells with copy number aberrations that were completely distinct from the primary breast cancer and therefore from an unknown origin. Gao *et al.* (Gao *et al.*, 2016) also detected similar aberrant cells, in tissue adjacent to breast tumours, which were once again unrelated to tumour cells. Clonal expansions of haematopoietic cells containing leukaemia-

associated mutations are well recognised in the circulation of otherwise healthy adults (Xie *et al.*, 2014, Genovese *et al.*, 2014). These confer an increased risk of a subsequent haematological malignancy, though many do not arise (Welch *et al.*, 2012, Gibson and Steensma, 2018, Abelson *et al.*, 2018).

Ongoing genetic divergence - intratumour heterogeneity

The evolution of tumour cells requires ongoing genetic variation. Nowell surmised this from the correlation of genetic aberration with more aggressive and advanced malignancies (Levan and Mitelman, 1975). Many morphological and phenotypic observations of more advanced malignancies were essentially circumstantial. Evidence supporting genetic variation within tumours has essentially accrued in two phases. Initially this was based on gross morphological and cytogenetic observations of heterogeneity within tumours. Pathologists have long noted the histological variation in morphological appearance in tumours (Fitzgerald, 1986, Hirsch *et al.*, 1983, Kruger *et al.*, 2003, Van Der Poel *et al.*, 1997). Clinical decision making only reflects the most aggressive appearing part of the tumour (Ignatiadis and Sotiriou, 2008). FISH and karyotyping demonstrated that gross genetic differences reflect the degree of morphologic heterogeneity (Farabegoli *et al.*, 2001, Maley *et al.*, 2006, Mora *et al.*, 2001, Pantou *et al.*, 2005, Roka *et al.*, 1998, Sauter *et al.*, 1995, Coons *et al.*, 1995).

In the last decade, next generation sequencing has enriched the evidence of intra-tumoural heterogeneity and begun the much more detailed process of describing how mutations mark the evolutionary process. This has demonstrated heterogeneity at the cellular level using single cell sequencing (Eirew *et al.*, 2015, Navin *et al.*, 2011, Gao *et al.*, 2016, Kim *et al.*, 2018, Zhang *et al.*, 2018a, Voet *et al.*, 2013) and across tumour regions and metastases (Campbell *et al.*, 2010, Yachida *et al.*, 2010, Gerlinger *et al.*, 2012, Gudem *et al.*, 2015). This has benefitted from economic, technical and computational advances in the application of next-generation sequencing. The decreasing cost of sequencing has made possible sequencing of large numbers of regions of tumours both across space and time. Initial studies of individual tumours have progressed to large scale programmes to detail tumour evolution (Abbosh *et al.*, 2017, Jamal-Hanjani *et al.*, 2017, Mitchell *et al.*, 2018, Turajlic *et al.*, 2018a, Turajlic *et al.*,

2018b). These have demonstrated that genetic heterogeneity is widespread in tumours, though can be of various different patterns. Many tumours demonstrate spatially segregated regions that represent distinct branches in evolution (Campbell *et al.*, 2010, Yachida *et al.*, 2010, Gerlinger *et al.*, 2012). Others demonstrate intermixed subclonal populations, more consistent with an early evolutionary divergence, the so-called 'big bang' model (Sottoriva *et al.*, 2015, Navin *et al.*, 2010).

Most significant in recent technical advances has been the ability to explore changes within individual tumour cells at the genetic, epigenetic and transcriptomic levels, sometimes simultaneously (Wang and Navin, 2015, Navin and Hicks, 2011, Van Loo and Voet, 2014, Navin, 2014). High throughput techniques, which are most advanced for transcriptome sequencing, have developed for analysing large numbers of single cells (Zheng *et al.*, 2017). Exciting high-throughput approaches are now emerging for other 'omic layers (Zahn *et al.*, 2017) or to analyse multiple layers simultaneously (Macaulay *et al.*, 2015, Macaulay *et al.*, 2017, Dey *et al.*, 2015) but current applications are still costly. These 'multi-omic' approaches are likely to significantly improve the interpretation of non-genetic cellular heterogeneity. This is otherwise confounded by heterogeneity among non-tumour cells, which results from the variety of cell types and states (Ryser *et al.*, 2018, Bian *et al.*, 2018).

Critical computational and bioinformatic frameworks have developed to harness the richness of information yielded by sequencing experiments for evolutionary analysis. At the simplest level this has used the frequency at which variants are detected in sequencing data to understand the proportion of cells that possess them (Campbell *et al.*, 2008, Nik-Zainal *et al.*, 2012b, Dentro *et al.*, 2017). Integrating different levels of information, copy number and allele frequency, has allowed mutations to be ordered in time (Mitchell *et al.*, 2018, Gerstung *et al.*, 2018, Jolly and Van Loo, 2018). Reconstructed phylogenies can then be compared across tumours, either manually (Turajlic *et al.*, 2018b, Turajlic *et al.*, 2018a, Jamal-Hanjani *et al.*, 2017) or potentially with the application of machine learning (Caravagna *et al.*, 2018) to infer conserved paths in evolution.

The pressure of selection

Shaping the process of evolution are selective pressures; promoting the survival of cells acquiring advantageous mutations and driving to extinction those with relatively deleterious mutations. Some degree of heterogeneity clearly derives from mutational drift in populations, which can result in subpopulations defined by mutations with only neutral survival impact (Williams *et al.*, 2016). The degree to which this occurs in cancer is still a matter for debate (Tarabichi *et al.*, 2018, Heide *et al.*, 2018), however it is undeniable that selection does shape the majority of tumour evolution. A study of selection by Martincorena *et al.* (Martincorena *et al.*, 2017), using the normalised ratio of non-synonymous to synonymous mutations (dN/dS), demonstrated a marked preponderance of positive over negative selection. As would be expected, all known cancer genes, whether oncogenes or tumour suppressors are found to be under positive selection. This approach has identified a large number of genes, many not yet recognised as cancer genes, that are positively selected in cancer. These candidate cancer genes require further validation. The lack of genes under negative selection suggests that relatively few genes, if mutated, are deleterious to tumour growth. This enormous resilience of the genome in part results from the redundancy created by haplosufficiency in a diploid genome. As expected, negative selection of genes was noted when one copy has already been lost by deletion. In addition it suggests that cells either have enormous intrinsic resilience or that a pre-existing driver mutation renders most new mutations irrelevant (Bakhoun and Landau, 2017). For example, most mutations predicted to generate presentable neoantigens are not negatively selected, probably because tumours already have a mechanism of immune evasion that renders them invisible (Van den Eynden *et al.*, 2019).

Co-evolution with the tumour micro-environment

Tumour cells exist not merely with each other but in complex relationships with their host tissues which are comprised of numerous different cell types (Hanahan and Coussens, 2012). As an example, Chapter 4 explores a tumour type which is characterised by the presence of a non-tumour osteoclast-like giant cell in the

tumour environment, though its function is not entirely clear (**Results 4.1.1**). Interactions with the tumour environment are contrasting and dynamic, imposing both selective pressures on the evolution of tumour cells, as well as being manipulated by them. The numerous cell types involved, including fibroblasts, immune cells, endothelial cells and pericytes, can all control tumour growth as well as facilitate it. For example, immune cells can provide a chronic inflammatory environment conducive to tumour cell growth and invasion (Qian and Pollard, 2010, Grivennikov *et al.*, 2010, DeNardo *et al.*, 2010) as well as an effective anti-tumour response (Galon *et al.*, 2006). Adaptive immunity provides a powerful selective pressure (Zhang *et al.*, 2018b, Milo *et al.*, 2018), resulting in tumour cell adaptations to evade it (Rosenthal *et al.*, 2019, McGranahan *et al.*). These interactions may also not be merely intrinsic but influenced by exogenous exposures. For examples cigarette smoke is not only a powerful mutagen but also has a complex affect on the bronchial environment in which tumour cells grow (Wu *et al.*, 2016).

Interactions with the microenvironment vary at different stages of tumour evolution, most visibly between the primary tumour and the metastatic niche, requiring further adaptation.

The evolution of bone tumours, around which this thesis centres, is almost entirely unexplored. There are three reasons to consider bone tumours of particular interest. First, the mesenchymal tissues from which they arise have one of the lowest cell replication rates, presumed to explain much of the low incidence of these tumours. This is likely to produce the lowest rate of intrinsic mutation and tumour development but this is largely unknown beyond the most common malignant tumour, osteosarcoma (Wu *et al.*, 2016). Second the microenvironment of the bone is complex but radically different from that of either the sites of common epithelial tumours or haematological malignancies which are only partly located amidst the bone marrow. Finally, the environmental factors to which the bone is exposed are clearly distinct from those in other sites and potentially more limited. The net effect of these on bone tumour evolution is unknown, and though it is likely to still reflect common evolutionary patterns with other tumours, the dynamics may differ and warrant further investigation.

1.2.2. The clinical relevance of cancer evolution

The biological insights yielded from considering the cancer genome in an evolutionary framework have many potential clinical benefits. These range from the clinical management and monitoring of established cancers to the prediction and prevention of cancer.

Evolution-guided therapy

The heterogeneity within tumours and their ability to evolve in response to therapy has proved a major hurdle to the implementation of personalised medicine (Dagogo-Jack and Shaw, 2018, Swanton *et al.*, 2011). Appreciating this challenge might allow it to be overcome. Targeting mutations present in all tumour cells (clonal mutations) will clearly have the greatest chance of tumour eradication. Targeted therapies, proven effective to date, have implicitly relied on this principle because they have often targeted mutations that are highly prevalent between tumours. These are also likely to be early events in tumorigenesis and are therefore likely to be clonal.

Monotherapy, even targeting clonal mutations, invariably results in the acquisition of treatment resistance. Resistance arises through multiple mechanisms which have been well explored in the context of hormone therapies in breast and prostate cancer (Jeselson *et al.*, 2015, Watson *et al.*, 2015, Gudem *et al.*, 2015). Resistance mutations may pre-exist or appear subsequent to the therapy, or relate to non-genetic factors. Pre-existing mutations might be rare, potentially found in only a single cell, and therefore undetectable by present assays. Treatment creates selective pressure allowing resistant cells to increase in relative terms. This might be true even for conventional cytotoxic therapies (Kim *et al.*, 2018). Resistance mutations may also occur *de novo* after treatment exposure, perhaps induced through iatrogenic mutagenesis (Alexandrov *et al.*, 2013, Szikriszt *et al.*, 2016). It is, however, difficult to exclude that these mutations did not exist prior to treatment at a very low and undetectable frequency. Non-genetic resistance may either be related to cell state plasticity or a specific molecular resistance pathway, such as aurora kinase activation in anti-EGFR treated lung cancer (Sharma *et al.*, 2018, Shah *et al.*, 2019) but are generally not

well understood. The widespread evolution of resistance suggests that clonal monotherapies are unlikely to achieve permanent tumour control or cure. For those patients with slow-paced advanced disease, or who would not tolerate more intensive therapy, single agents will continue to play an important role. Most responses to targeted therapies, however, are both incomplete and short-lived and require improvement.

A combination of clonal-targeted therapies might reduce the emergence of resistance. In addition to oestrogen dependency, many breast cancers have cell cycle dysregulation (Dean *et al.*, 2010), and the addition of CDK4/6 inhibition to aromatase inhibition can prolong treatment response. This can delay the use of cytotoxic chemotherapy but increases toxicity (Cristofanilli *et al.*, 2016). Conversely resistance to BRAF inhibitors in metastatic melanoma was thought to result from downstream MEK activation (Nazarian *et al.*, 2010, Johannessen *et al.*, 2010), however MEK and BRAF inhibition only demonstrated modest clinical benefit (Long *et al.*, 2014, Moriceau *et al.*, 2015). Resistant tumours often have multiple different detectable MAPK mutations, suggesting convergent evolution (Wagle *et al.*, 2014). Development of effective combination therapies requires more comprehensive understanding of mutation clonality and resistance mechanisms. Ideally, larger numbers of drivers could be targeted simultaneously or sequentially, depending on the pace and nature of the evolutionary response of the tumour.

Predicting a tumour's evolutionary response to a therapy might allow pre-emptive measures to prevent resistance, most likely involving combination therapies. For example, preclinical combination of ABL1 inhibitors in models of chronic myeloid leukaemia can pre-empt the emergence of resistant subclones and results in durable responses (Wylie *et al.*, 2017).

Adaptive therapy may also be a way of increasing the duration of response to therapy (Enriquez-Navas *et al.*, 2016, Gatenby *et al.*, 2009). This relies on the principle that tumour cells compete with one another for survival. Resistance mutations, in the absence of a relevant therapy, are intrinsically disadvantageous. In an adaptive approach, sensitive subclones are treated to the point at which tumour size is reduced or growth is suppressed to achieve symptomatic benefit

(**Figure 5**). Thereafter, treatment is reduced or withdrawn. This allows the competitive suppression of resistant subclones by sensitive subclones, that in the absence of therapy have a growth advantage. This approach is currently under evaluation in metastatic prostate cancer, using individualised serum Prostate Specific Antigen (PSA) thresholds to guide the use of the CYP17A1 inhibitor, Abiraterone (Zhang *et al.*, 2017). Small numbers of patients have been treated with this approach with good clinical outcome and reduced cumulative exposure to medication.

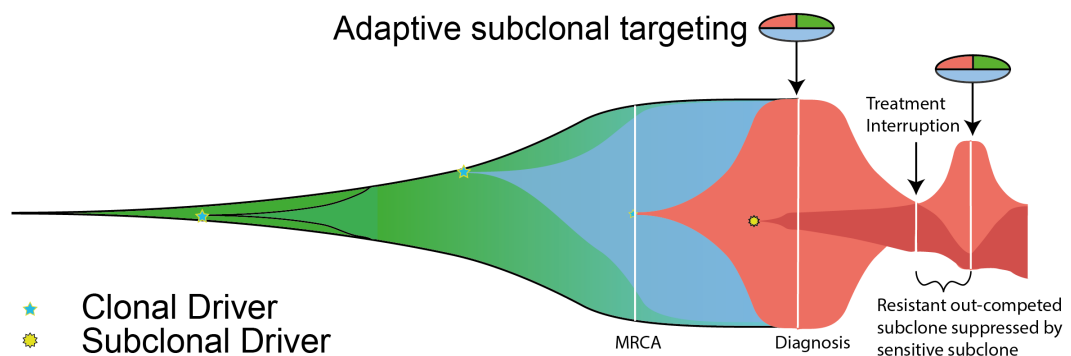


Figure 5 Schematic of adaptive therapy. A schematic of the relative proportions of different clonal and subclonal cell populations. Colours display populations of cells defined by a common set of mutations, set apart by the initial acquisition of a positively selected driver. The ellipse above represents a combination therapy against the drivers defined by colour. The maroon population, which increases in proportion to other cells during treatment is resistant to therapy but has a relative survival disadvantage in its absence. Figure adapted from (Fittall and Van Loo, 2019).

Effective pharmacological options for targeting driver mutations are relatively limited and many tumours have few detected clonal drivers, therefore require alternative approaches (Turajlic *et al.*, 2018b, Turajlic *et al.*, 2018a, Jamal-Hanjani *et al.*, 2017). Some drivers may be targetable indirectly by synthetic lethality, a treatment approach that exploits a cellular vulnerability exposed by a clonal driver mutation. As an example, *BRCA* mutations increase genomic instability, which produces variation during tumorigenesis, but increases their reliance on other DNA repair mechanisms. Therapeutic inhibition of the single-stranded DNA repair PARP enzymes (Lord and Ashworth, 2016, Ashworth, 2008) causes the accumulation of lethal DNA damage specifically in tumour cells. PARP inhibition resistance can still emerge via the somatic reversal of *BRCA* mutation (Patch *et al.*, 2015, Weigelt *et al.*, 2017, Christie *et al.*, 2017).

Alternatively, collateral lethality has been proposed to harness susceptibilities created by the loss of genes genomically adjacent to deleted tumour suppressors (Muller *et al.*, 2015, Muller *et al.*, 2012). Finally, the adaptive anti-tumour immune response released by immunotherapy may exploit the antigenicity of clonal passenger neoantigen mutations. One of the potential predictive markers of response in non-small cell lung cancer and melanoma is the clonal neoantigen load (McGranahan *et al.*, 2016, Snyder *et al.*, 2014), highlighting the need to detail fully a tumour's genomic diversity to plan optimal treatment.

Evolutionary monitoring

Monitoring the dynamics of different tumour cell populations may substantially enhance the treatment of advanced disease. Liquid biopsies are non-invasive assays, easily repeated over time, and therefore ideal for this purpose. They are already proposed to reduce the invasiveness of clinical assays used for diagnosis (Newman *et al.*, 2014, Beaver *et al.*, 2014), prognosis (Bettegowda *et al.*, 2014, Khan *et al.*, 2018), molecular profiling (Oxnard *et al.*, 2016) and response assessment (Dawson *et al.*, 2013, O'Leary *et al.*, 2018b, Spina *et al.*, 2018, Barault *et al.*, 2018, Agarwal *et al.*, 2019). Circulating tumour DNA (ctDNA) is relatively stable and simple to handle, and its sequence content can be analysed using a variety of approaches (Henaó Diaz *et al.*, 2016). Monitoring subclonal evolution, to date, has focused on evaluating somatic point mutations in ctDNA. Murtaza *et al.* detected an increasing level of mutations private to a subclone, responsible for the progression of a chest wall breast cancer metastasis (Murtaza *et al.*, 2015). O'Leary *et al.* were able to use ctDNA to both predict longer progression-free intervals (O'Leary *et al.*, 2018b) and detect emerging resistant subclones (O'Leary *et al.*, 2018a) in a small proportion of metastatic breast cancer patients, treated with the CDK 4/6 inhibitor, palbociclib. Abbosh *et al.* (Abbosh *et al.*, 2017) could detect ctDNA 10-346 days (median 70 days) prior to clinical detection of relapsed lung cancer.

ctDNA monitoring is principally limited by cost and biases. Abbosh *et al.* (Abbosh *et al.*, 2017) estimated that a limited bespoke monitoring panel would cost USD 1,750 per patient, though clearly this will continue to reduce with declining sequencing costs. Biological biases may be more challenging to overcome. It is

likely that highly vascular and necrotic tumours will contribute more ctDNA than those in cryptic sites, such as the central nervous system (De Mattos-Arruda *et al.*, 2015). Cell-free DNA is predominantly generated by apoptotic nuclease activity of nucleosome-associated DNA (Thierry *et al.*, 2010, Giacona *et al.*, 1998), resulting in distinct chromatin-associated patterns and a genomic bias. Most approaches also introduce technical biases. As exemplified by Abbosh *et al.* (Abbosh *et al.*, 2017), they often only search for mutations already detected in the primary tumour. This, by definition, means that *de novo* mutations that arose subsequent to the sampling of the primary tumour, will not be detectable in circulation.

As an alternative, circulating tumour cells (CTCs) hold prognostic information (Khan *et al.*, 2018, Barault *et al.*, 2018) and can be analysed using single cell sequencing. The rarity of these tumour cells, however, requires significant enrichment and is likely to introduce other biases, resulting in low sensitivity even for clonal tumour populations (Alix-Panabieres and Pantel, 2014, Krebs *et al.*, 2014).

Cancer prevention, screening and stratification

Reducing the burden of cancer deaths and morbidity will be best served by reducing incidence, the proportion of late stage diagnoses, and focusing surveillance of those at highest risk of relapse. Understanding tumour evolution may help achieve this through understanding preventable factors and optimising screening and risk stratification. Central to this effort are methods that recapitulate the early evolution of cancers using sequencing information from late-stage cancers alone (Jolly and Van Loo, 2018). In general terms, these utilise the number of copies of mutations on gained chromosomal segments to infer whether these mutations happened before or after that gain. For example, if a whole chromosome has been duplicated and mutations are found on two copies then it is likely that the mutations occurred first and were duplicated with the chromosomal gain. Wedge *et al.* (Wedge *et al.*, 2018) have been able to retrospectively identify chromosomal changes from sequencing data, that developed earlier in prostate cancer tumorigenesis. These findings, such as the early gain of chromosome 8q, recapitulated those previously found in prostate

intraepithelial neoplasia (PIN), thought to be a precursor of prostate adenocarcinoma (Jung *et al.*, 2016).

These approaches have also been applied to invasive cancers with less well-characterised precursor lesions (Mitchell *et al.*, 2018, Jamal-Hanjani *et al.*, 2017). Recently, the Pan-Cancer Analysis of Whole Genomes (PCAWG) initiative leveraged whole-genome sequencing data to infer evolutionary timelines across cancer types (Gerstung *et al.*, 2018). This reproduced and refined classic models of mutational progression such as colorectal cancer, in which *APC* mutations precede *KRAS* and *TP53* mutations.

Identifying preventable environmental, or infectious predisposing factors for cancer has previously relied on a combination of epidemiological and biological evidence. Mutational signatures can give direct evidence of the impact of some of these preventable factors on the genome (**Introduction 1.1.4**), exploring the impact of known factors and potentially identifying new ones (Nik-Zainal *et al.*, 2012a). Combining signature deconvolution and timing methods can infer their activity throughout tumour evolution. This approach in lung cancer demonstrated a reduction in relative smoking signature activity later in tumour evolution, despite ongoing smoke exposure, because of the increased relative activity of other processes, including the activity of the APOBEC family of cytidine deaminases (Swanton *et al.*, 2015, de Bruin *et al.*, 2014). Conversely, inherited defects in DNA repair, such as deficient mismatch repair seen in Lynch syndrome, can lead to steady and ongoing mutational activity throughout a tumour's lifetime (Campbell *et al.*, 2017). The aetiology of many mutational signatures has not yet been identified but understanding their activity in tumorigenesis may identify new preventable factors.

A deeper and more comprehensive understanding of tumour evolution should allow us to understand why and how a cancer or pre-invasive lesion will behave in the future. This has specific implications for both screening programmes and the risk stratification of established cancers. To date, risk stratification has relied almost exclusively on histological staging and grading. Overtreatment of lesions, unlikely to cause morbidity results in unnecessary cost, harm and anxiety (Esserman *et al.*, 2013). A number of different studies have suggested that

features of intratumour heterogeneity or the evolutionary trajectory may carry significant prognostic information, and thereby aid in this decision making (Turajlic *et al.*, 2018b, Jamal-Hanjani *et al.*, 2017, Turajlic *et al.*, 2018a, Karlsson *et al.*, 2018). More refined and individualised cancer predictions will require complex computational tools and models (Caravagna *et al.*, 2018, Gerhauser *et al.*, 2018). Much as weather forecasting models require vast amounts of measured data from the real world, cancer evolution models will require many more cancers to have their evolution profiled by sequencing. Ultimately, this will allow these forecasts to guide the optimal management for each patient.

1.3. Bone tumours as a case study of the patterns of mutation in tumour evolution

Much of this thesis is based on the study of the genomics of rare primary tumours of bone which posed clinically important questions amenable to genomic and evolutionary analyses. Osteblastoma (**Chapter 3**) is a rare benign bone tumour of which the genomic cause is unexplored. Occasionally it poses a diagnostic challenge in distinction from another osteoblastic tumour, osteosarcoma. It provides an example of a benign disease, which may have a simple mutational basis, and identifying this may be of clinical benefit. H3.3-mutated bone tumours already have an identified simple and pervasive mutation, however the rare phenomena of metastasising 'benign' tumours or true malignant progression are not understood (**Chapter 4**). The genomic features of this evolution are unknown and might provide useful clinical prognostic information.

Disease	Malignancy	Genetics (% affected)
Chondrogenic		
Osteochondroma	Benign	EXT 1/2 mutation (Bovee <i>et al.</i> , 1999)
Chondroma:	Benign	IDH 1/2 mutation (Amary <i>et al.</i> , 2011)
Enchondroma		
Periosteal chondroma		
Osteochondromyxoma	Benign	Familial syndrome, Carney complex – PPKAR1A mutation (Carney <i>et al.</i> , 2001)
Subungual exostosis	Benign	COL12A1 and COL4A5 rearrangement (Storlazzi <i>et al.</i> , 2006)
Bizarre parosteal osteochondromatous proliferation	Benign	Unknown, t(1;17) reported (Nilsson <i>et al.</i> , 2004)
Synovial chondromatosis	Benign	FN1-ACVR2A fusion (Amary <i>et al.</i> , 2019b)
Chondromyxoid fibroma	Benign	
Atypical cartilaginous tumour / chondrosarcoma grade I	Intermediate	IDH 1/2 mutation (Amary <i>et al.</i> , 2011)
Chondroblastoma	Intermediate	
Malignant	Malignant	H3.3 pK36M (Amary <i>et al.</i> , 2016)
Chondrosarcoma		IDH 1/2 (59%), COL2A1 (37%), RB1 pathway (33%) (Tarpey <i>et al.</i> , 2013).
Grade II, grade III		Peripheral chondrosarcoma: EXT 1/2 mutation (Bovee <i>et al.</i> , 1999)
Dedifferentiated chondrosarcoma	Malignant	As for malignant chondrosarcoma
Mesenchymal chondrosarcoma	Malignant	HEY1-NCOA2 fusion (Wang <i>et al.</i> , 2012)
Clear cell chondrosarcoma	Malignant	H3.3 p.K36M (rare) (Amary <i>et al.</i> , 2016)
Fibrogenic		
Desmoplastic fibroma of bone	Intermediate	CTNNB1 mutation (Song <i>et al.</i> , 2018)
Fibrosarcoma of bone	Malignant	NTRK3 fusion (Yamazaki <i>et al.</i> , 2019)
Fibrohistiocytic		
Benign fibrous histiocytoma / non-ossifying fibroma	Benign	KRAS (64%), FGFR1 (14%), or NF1 (3%) mutations (Baumhoer <i>et al.</i> , 2019)
Myogenic, Lipogenic or Epithelial		
Lipoma of bone	Benign	LPP-HMGA-2 fusion (Petit <i>et al.</i> , 1998)
Liposarcoma of bone	Malignant	MDM2 amplification (Szuhai <i>et al.</i> , 2007)
Leiomyosarcoma of bone	Malignant	RB1 deletion (Verelst <i>et al.</i> , 2004)
Adamantinoma	Malignant	Unknown (Taylor <i>et al.</i> , 2012)
Notochordal		
Benign notochordal tumour	Benign	Unknown (Du <i>et al.</i> , 2019)
Chordoma	Malignant	T (brachyury) duplication germline or sporadic (27%) (Tarpey <i>et al.</i> , 2017)

Disease	Malignancy	Genetics (% affected)
Osteoclastic		
Giant cell tumours of bone	<i>Intermediate</i>	<i>H3.3 G34 mutations (Behjati et al., 2013)</i>
Giant cell tumours of the small bones	<i>Intermediate</i>	<i>H3.3 G34 mutations (Behjati et al., 2013)</i>
Malignancy in giant cell tumours of bone	<i>Malignant</i>	<i>Unknown</i>
Osteogenic		
Malignant		
Osteoma	<i>Benign</i>	<i>Unknown</i>
Osteoid Osteoma	<i>Benign</i>	<i>Unknown</i>
Osteblastoma	<i>Intermediate</i>	<i>Unknown</i>
Conventional Osteosarcoma:	<i>Malignant</i>	<i>Complex and unexplained. TP53 (40%), IGF amplifications (7-14%), MDM2 amplification (Behjati et al., 2017, Kovac et al., 2015)</i>
Chondroblastic		
Fibroblastic		
Osteoblastic		
Low-grade central osteosarcoma		
Telangiectatic osteosarcoma		
Small cell osteosarcoma		
Secondary osteosarcoma		
Parosteal osteosarcoma		
Periosteal osteosarcoma		
High-grade surface osteosarcoma		
Vascular		
Haemangioma	<i>Benign</i>	<i>Unknown (van Ijzendoorn and Bovee, 2017)</i>
Epithelioid Haemangioma	<i>Intermediate</i>	<i>FOS and FOSB rearrangements (Antonescu et al., 2014, Huang et al., 2015)</i>
Epithelioid haemangioendothelioma	<i>Malignant</i>	<i>WWTR1-CAMTA1 fusion (Errani et al., 2011)</i>
Angiosarcoma	<i>Malignant</i>	<i>PTRB (26%) (Behjati et al., 2014)</i>
Undefined or miscellaneous		
Aneurysmal bone cyst	<i>Intermediate</i>	<i>USP6 fusions (Panagopoulos et al., 2008)</i>
Langerhans cell histiocytosis	<i>Intermediate</i>	<i>BRAF.pV600E or other MAPK mutation (Abla et al., 2019)</i>
Erdheim-Chester disease	<i>Intermediate</i>	<i>BRAF.pV600E (60%) (Cohen-Aubart et al., 2018)</i>
Ewing sarcoma	<i>Malignant</i>	<i>EWRS1-ETS fusions (Anderson et al., 2018)</i>

Table 1 Classification of bone tumours. Adapted from the WHO classification of bone tumours (Fletcher *et al.*, 2013). Haematopoietic tumours have been excluded. The most comprehensive and recent characterisation of genetically ‘unknown’ diseases are referenced.

Osteoblastoma, and H3.3-mutant bone tumours are examples of a wide spectrum of primary bone tumours classified broadly by their morphological appearances but specific subtypes have become refined by molecular characteristics (**Table 1**). Bone tumours, as with soft tissue tumours are frequently defined by highly recurrent mutations. Many of these mutations are gene fusions, assumed to have

a simple genetic basis. The finding of chromoplexy underlying many of the canonical *EWRS1* fusions in Ewing sarcoma (**Results 5.1.3**), suggests that more detailed genomic analysis is warranted. This may identify mechanisms generating these rearrangements, which may subsequently be amenable to better prediction or prevention.

1.4. Thesis aims and objectives

The aim of this thesis is to explore the changes that occur in the tumour genome, their complexity and how they define tumour evolution. Bone tumours exhibit many of the mutational patterns seen across other cancer types, though there are still subtypes that remain unexplored. They also present a number of unmet clinical needs, ranging from diagnostic challenges to a paucity of therapeutic options in advanced disease.

The specific objectives are:

- Evaluate the genomic landscape of a previously unexplored disease, osteoblastoma, to identify simple genomic changes underpinning the earliest stages in tumour evolution.
- Explore the genomic patterns underlying clinical features of metastatic and/or malignant progression of histone H3.3 mutated bone tumours.
- Delineate the role of the complex mutational pattern of chromoplexy across cancer evolution.

Chapter 2. Materials & Methods

2.1. Chapter specific analysis flows

Experiments and analysis conducted on samples are represented in the following flow diagrams (Figure 6 - Figure 11). The paragraph references for accounts of the detailed methodology are shown in parentheses. Unless otherwise stated all analysis was conducted by me alone.

2.1.1. Chapter 3 methods

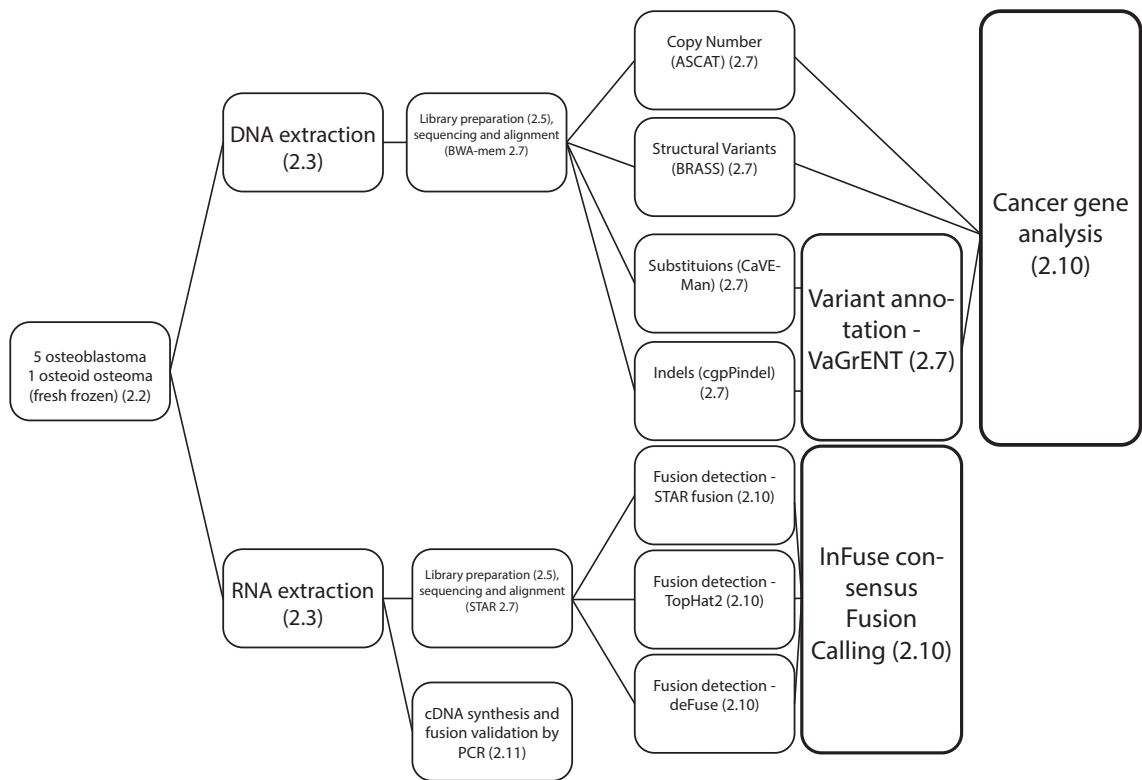


Figure 6 Chapter 3 discovery cohort. The discovery cohort of samples were subjected to DNA and RNA sequencing with workflows as described above. In addition, the RNA was used to create cDNA for fusion validation.

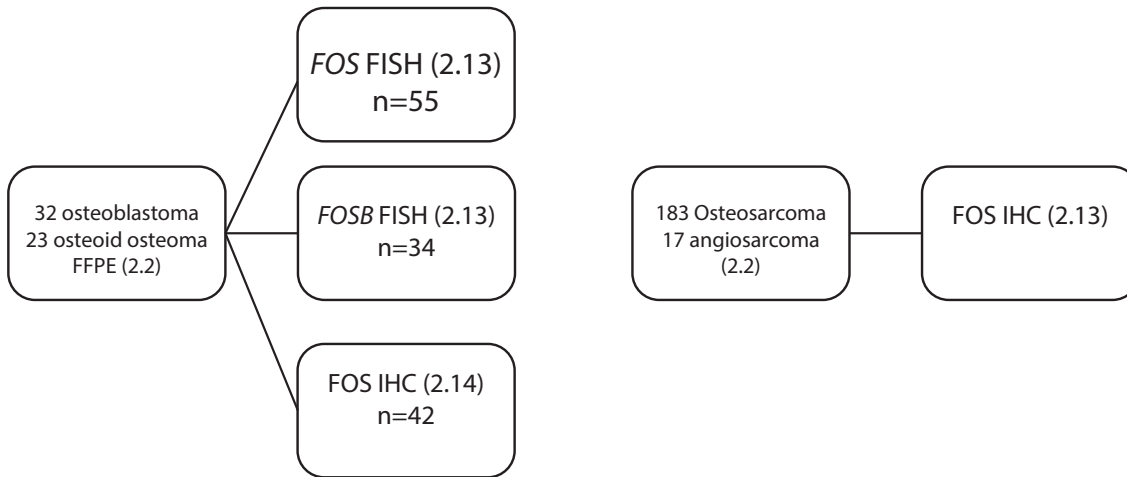


Figure 7 Validation cohort analysis. Validation involved two steps: validation of the findings in a broader group of osteoid osteoma/osteoblastoma (left panel) involving Fluorescence In-Situ Hybridisation (FISH) and Immunohistochemistry (IHC) – note not all samples were available for all analyses (**Appendix 7.1.7**), and in related malignant sarcomas (right panel).

2.1.2. Chapter 4 methods

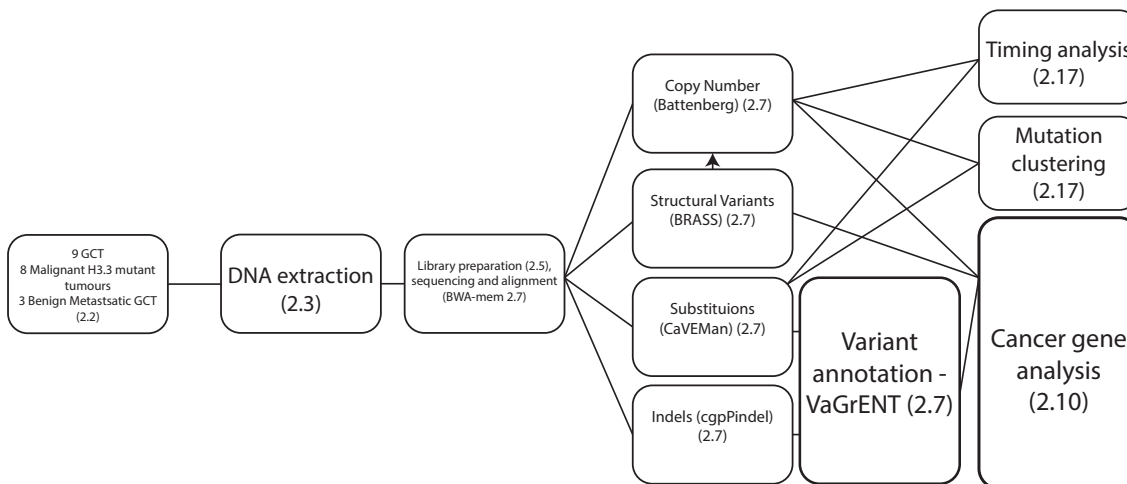


Figure 8 H3.3 mutated tumour sequencing analysis. Whole genome sequencing alone was conducted for the initial cohort of H3.3 mutated tumour.

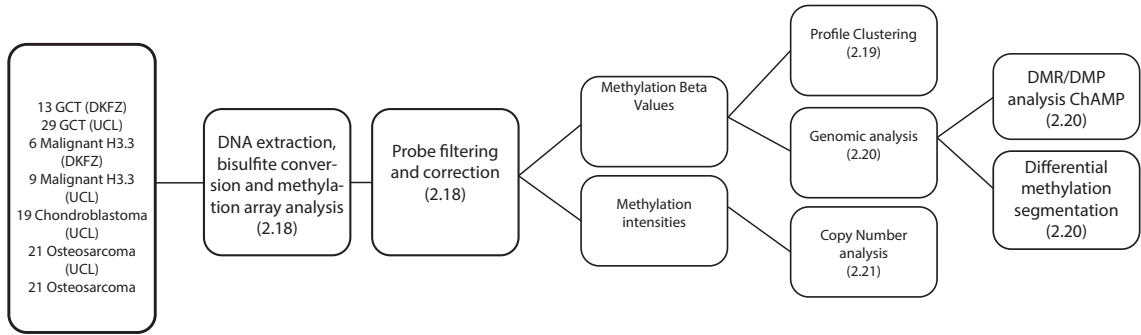


Figure 9 Methylation analysis. Methylation array analysis involved a broader panel of tumours including chondroblastoma, as a K27M mutated tumour, and osteosarcoma, as a malignant bone tumour without a histone mutation. Initial analysis for all arrays was in common.

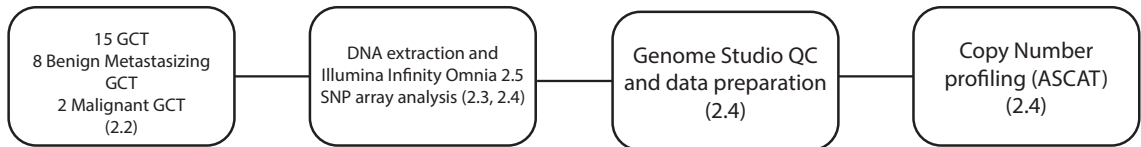


Figure 10 SNP array analysis. Limited numbers of tumours were subjected to SNP array analysis to confirm the trends seen in copy number profiles between benign and malignant tumours. Note the samples of benign metastasizing tumours were all taken from primary tumours.

2.1.3. Chapter 5 methods

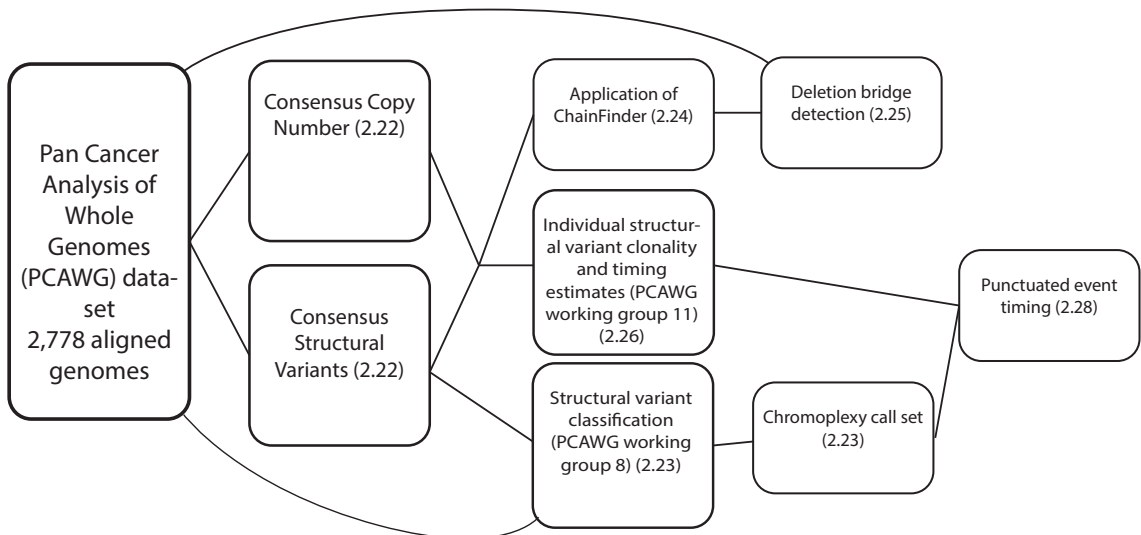


Figure 11 Chromoplexy workflow. The complex structure of this workflow demonstrates the reliance of each step on collaborative working groups. The aligned genomic sequences were utilised for both deletion bridge detection and PCAWG structural variant classification.

2.2. Patient samples

Patients provided their written and informed consent to provide samples for the work detailed in chapters 3 and 4 via the UCL Musculoskeletal biobank, based at the Royal National Orthopaedic Hospital. This was approved by the National Research Ethics Service (NRES) Committee Yorkshire & The Humber – Leeds East (15/YH/0311). Pathology review of selected cases was conducted by specialist bone pathologists, Adrienne Flanagan, Roberto Tirabosco, and Fernanda Amary.

2.3. DNA and RNA extraction

Fresh frozen tumour samples were embedded in Tissue-Tek OCT and sectioned on a cryostat. Haematoxylin and Eosin (H&E) stained sections were reviewed for tumour type and uniformity, to ensure a tumour content of greater than 50%. DNA was extracted using an automated magnetic bead extraction and purification system according to the manufacturer's protocols (Prepito DNA Tissue10 Kit, Perkin Elmer Ltd, Bucks, UK). Matched normal DNA was acquired from blood using a column-based system (Qiamp DNA Blood Maxi kit, Qiagen, Manchester, UK). DNA concentration and quality were assessed by a fluorometric assay (Picogreen, Thermofisher Scientific, Paisley, UK) and a PCR assay followed by gel electrophoresis. Only DNA that was of suitable concentration (minimum 500 ng total) was used for whole genome sequencing. Total RNA was isolated from frozen tissues using the Zymo Direct Zol RNA isolation kit according to manufacturers' recommendations that included the on-column DNase digestion. All nucleic acid extraction and preparation was performed by staff in the UCL musculoskeletal biobank at the Royal National Orthopaedic Hospital.

2.4. SNP array analysis

Extracted DNA was prepared using the Illumina Array platform pipeline by the UCL genomics core facility. SNP Array analysis was performed on Illumina Infinity Omnia 2.5 chips. Raw data were quality controlled and converted into

normalised LogR and B-Allele Frequency tracks using Illumina Genome Studio (2.0.4). Copy number profiles were produced using ASCAT (2.5.1) (Van Loo *et al.*, 2010). The segmentation parameter was adjusted to 200 to reduce artefactual segmentation noted on comparison with samples also subjected to copy number analysis from whole genome sequencing.

2.5. Sequencing

For whole genome sequencing the Illumina (Illumina, Chesterford, UK) no-PCR library protocol was used to construct short insert 500 bp libraries, prepare flowcells and generate clusters. Whole genome sequencing was performed using the Illumina HiSeq 2000 or 2500 platform, using 100 bp paired-end libraries. Whole genome sequencing for malignant H3.3 mutant bone tumours (**Chapter 4**: PD30981-5, PD37332, PD3788, PD3795, PD38328, PD38329, PD4915, PD4922) was performed using the XTen platform using 150 bp paired-end libraries. Poly-A RNA was sequenced on an Illumina HiSeq 2000 using 75 bp paired-end libraries. All sequencing was performed by the genomics facility at the Wellcome Trust Sanger Institute.

2.6. Data analysis

General data analysis was performed in R (3.5.3 and 3.6.0) in RStudio (1.1.383), with bespoke scripts.

2.7. Whole genome sequencing alignment and variant detection

The algorithms used for alignment and variant detection, with standard settings, are detailed in the table below.

Function	Name	Ch. 3	Ch. 4	Reference
WGS Alignment	Burrows Wheeler Aligner (BWA mem)	2.0.54	2.0.54	(Li and Durbin, 2009)
Substitutions	Cancer Variants through Expectation Maximisation (CaVEMan)	1.11.0	1.11.0	(Jones <i>et al.</i> , 2016)
Indels	cgpPindel	2.1.0	2.2.4	(Ye <i>et al.</i> , 2009)
Copy Number	Allele Specific Copy Number Analysis of Tumours (ASCAT) NGS	4.0.0		(Van Loo <i>et al.</i> , 2010)
Copy Number	Battenberg (with SVs from BRASS)		2.2.8	(Dentro <i>et al.</i> , 2017, Nik-Zainal <i>et al.</i> , 2012b)
Structural Variants	Breakpoints via Assembly (BRASS)	5.3.3	5.3.2	https://github.com/cancerit/BRASS
RNAseq Alignment and Fusion calling	Spliced Transcripts Alignment to a Reference (STAR)	2.0.42		(Dobin <i>et al.</i> , 2013)
Fusion Calling	TopHat2	2.1.0		(Kim <i>et al.</i> , 2013)
Fusion Calling	deFuse	0.7.0		(McPherson <i>et al.</i> , 2011)
RNAseq counts	HTSeq	0.6.1		(Anders <i>et al.</i> , 2015)

Table 2 Sequencing algorithms.

2.8. Variant validation

The precision of *Cancer Genome Project* (Wellcome Trust Sanger Institute) variant calling pipeline has been determined in multiple studies (Nik-Zainal *et al.*, 2016). I confirmed this through manual inspection of raw sequencing reads for up to 100 variants of each type from all samples. The precision of all variants was >95% in all cases. Additional post-processing filters were applied to substitutions to achieve this precision: median alignment score (ASMD) of variant reads >90 (100bp paired end) and >140 (150bp paired end) and median number of clipped bases in variant supporting reads (CLPM) of 0. Sample PD37332 (**Chapter 4**) was noted to have a large number of structural variants supported by minimal

numbers of locally discordantly mapping reads, therefore for this sample only, structural variants were only considered if they were possible to map to base-pair resolution. All copy number profiles were manually scrutinised for the requirement of refitting, however in all cases the first solution was optimal.

2.9. Analysis of mutations in cancer genes

Variants were analysed using a defined strategy. Variants were considered as potential drivers if they presented in established cancer genes (Chapter 3 COSMIC v82, Chapter 4 COSMIC v85). Tumour suppressor coding variants were considered if they were annotated as functionally deleterious by the VAGrENT algorithm (<http://cancerit.github.io/VAGrENT/>). Disruptive rearrangements or homozygous deletions of tumour suppressors were also considered. Additionally, homozygous deletions were required to be focal (<1 Mb in size). Mutations in oncogenes were considered driver events if they were located at previously reported hot spots (point mutations) or amplified the intact gene. Amplifications also had to be focal (<1 Mb) and result in at least 5 copies in diploid genomes, or 4 copies more than the modal major copy number in genome duplicated samples.

2.10. Fusion detection (Chapter 3)

Rearrangements in *FOS* and *FOSB* were analysed using the DNA structural rearrangement caller, BRASS and the RNA fusion detection algorithms deFuse, TopHat2 and STAR fusion. Fusions were considered if breakpoints and orientations were supported by both BRASS and at least one RNA-based algorithm. All DNA and RNA reads supporting the breakpoints were manually inspected. In sample PD13482, in which neither algorithm identified the fusion, both split reads and discordant read pairs spanning the fusion were identified in the DNA- and RNA-Seq data. In PD7525, additional rearrangements were identified mapping to sequences with homology across chromosome 16. The breakpoint location was selected on the basis of the greatest number of supporting discordant and split reads after manual local assembly.

FOS fusion partner breakpoints were all intergenic or intronic and therefore not normally represented in RNA sequencing libraries. The per-base coverage in these regions therefore reveals a clear peak, present only in that tumour sample, demonstrating expression of aberrant transcripts (normalised by the mean of HTSeq counts $\times 10^3$; **Figure 15**). For schematic purposes, horizontal line segments are plotted to reflect the mean normalised coverage: the ‘mate transcript segment’ is between the breakpoint (grey vertical dashed line) and the poly-adenylation cleavage site; surrounding segments are the mean sequencing coverage over a genomic range of equal length to the ‘mate transcript segment’. The end of the transcript (**Figure 19**) was considered to be immediately downstream of the cleavage and poly-adenylation signal (“AATAAA”) with the greatest drop in coverage in the surrounding 200bp.

2.11. *FOS* fusion validation

To validate *FOS* fusions Dr Annelien Verfaillie synthesized cDNA from 1 μ g of total RNA from each sample using the ProtoScript® II First-Strand cDNA Synthesis Kit (NEB). PCR was performed with Phusion high-fidelity PCR master mix (HF buffer, NEB) with amplification primers (**Table 3**). Amplified products were size selected using gel electrophoresis and then Sanger sequenced using internal primers (**Table 3**). All primers were designed by both Annelien Verfaillie and me.

Sample	Forward Amplification	Reverse Amplification	Sequencing Primer
PD13480	GGTGCTCGAGTGAGAAGCCAAGACTGAGCC	GGTGGGTACCTTTATTAGATAGATGCAGGGAGGGC	AGAGTTCATCCTGGCAGCTC
PD13481	GGTGCTCGAGTGAGAAGCCAAGACTGAGCC	GGTGGGTACCTGCTCTATAAAGCGTTTATTTAATTAATGAGG	AGAGTTCATCCTGGCAGCTC
PD13482	GGTGCTCGAGTGAGAAGCCAAGACTGAGCC	GGTGGGTACCCATGCCTTATTCATCTTTATTTCTGAAAG	AGAGTTCATCCTGGCAGCTC
PD7519	GGTGCTCGAGTGAGAAGCCAAGACTGAGCC	GGTGGGTACCCCTGTTAATCATTTTGTTTTATTTGAC	AGAGTTCATCCTGGCAGCTC
PD7521	GGTGCTCGAGTGAGAAGCCAAGACTGAGCC	GGTGGGTACCTGCATATACATTATTTACTTTATTAATCTTACTAATATTCTTTAAGG	AGAGTTCATCCTGGCAGCTC

Table 3 *FOS* fusion validation primers.

2.12. Allele-specific expression analysis

Allele-specific expression in *FOS* and *FOSB* was analysed using allele counts at heterozygous single nucleotide polymorphisms (SNPs). Heterozygous SNPs were identified from DNA sequencing data. Allele counts were measured from

RNA-Seq reads using GATK ASEReadCounter (McKenna *et al.*, 2010) with Jonas Demeulemeester.

2.13. Fluorescence In-Situ Hybridisation (FISH) for FOS and FOSB

A cohort of 55 informative cases of osteblastoma/osteoid osteoma was examined by FISH for *FOS* breakapart by Dr Hongtao Ye (RNOH). *FOSB* probes were custom-designed with Agilent SureDesign to flank the breakapart region by Dr William Mifsud (GOSH). *FOS* probes and methods have been described previously (Huang *et al.*, 2015) (**Supplementary Data 11**). Deparaffinised sections were pre-treated by pressure cooking for 5 minutes and subsequently incubated in pepsin solution at 37°C for 50 minutes. Probes were applied to tissue sections and denatured at 72°C, followed by hybridisation overnight at 37°C. After hybridisation, the sections were washed and mounted with 4',6-diamidino-2-phenylindole and coverslips.

2.14. Immunohistochemistry

FOS, *FOSB* and Cyclin D1 immunohistochemistry was performed by Dr Elena Miranda (UCL Cancer Institute, CRUK Core Facility). Deparaffinised hydrated tissue sections underwent antigen unmasking in Tris-EDTA pH 9 (DAKO S2367 - Agilent Technologies LDA UK Limited, Cheshire UK) at high pressure for 2 minutes. After washing and quenching, sections were blocked in 2.5% horse serum (Vector ImmPRESS Kit) for 20 minutes at room temperature. Incubation with primary antibodies was for 60 minutes, secondary antibodies for 30 minutes, and Diaminobenzidine (DAB)+ substrate/chromagen (Dako, K3468) for 5 minutes, all at room temperature, prior to counterstaining and mounting. *FOS* antibodies (EMD Millipore ABE457, Rabbit Polyclonal) were used at 1 or 0.5ug/mL with ImmPRESS Horseradish Peroxidase Anti-Rabbit IgG (Peroxidase) Polymer Detection Kit, made in Horse (MP-7401, Vector Laboratories, Peterborough, UK). *FOSB* antibodies (clone 5G4, dilution 1:100, Cell Signaling Technology, Danvers, MA, USA) were used as previously described (Sugita *et al.*, 2016). Cyclin D1

antibody (clone SP4, Rabbit monoclonal clone SP4, Cell Marque, Rocklin, CA, USA) was used at 0.1 ug/mL. H3.3 G34 immunohistochemistry was performed by Mr Fitim Berisha (RNOH) on the Leica Bond III automated immunostaining platform, with peroxidase blocking and detection carried out using the Leica Bond Polymer Refine DAB kit (Leica, DS9800, Leica Microsystems, Milton Keynes, UK) according to manufacturer's instructions. Pre-treatment was with Leica Epitope Retrieval solution 2 (AR9640) for 20 minutes. Primary antibody against H3.3 G34W was with a Rabbit monoclonal antibody (clone RM263; RevMAb Biosciences USA, San Francisco, CA, USA) at a dilution of 1/1500 with Leica Bond Primary Antibody Diluent (Leica, AR9352) for 30 minutes at ambient temperature. Slide interpretation was performed with Adrienne Flanagan.

2.15. Copy Number Scoring (Chapter 4)

A sample was considered Whole Genome Duplicated (WGD) when modal total copy number was >2 (Bielski *et al.*, 2018). The baseline total copy number was considered as 4 for WGD samples and 2 for others. Autosomal copy number segments were then scored as the difference from this baseline; no difference (0), total copy number of 0 (homozygous deletion, 2), total copy number $\geq 3 +$ baseline (amplification, 2), other score not equal to baseline (1). Scores were normalised relative to the length of the chromosome, summed and then divided by the theoretical maximum (44).

2.16. Mutation clustering, purity estimation and phylogenetic reconstruction

The algorithm DPCLust (2.2.6) and its pre-processing pipeline (1.0.8) were used to cluster mutations according to fraction of cancer cells (Cancer Cell Fraction, CCF) in which they were found, as described previously (Nik-Zainal *et al.*, 2012b). Filtered substitutions (**Methods 2.8**) and Battenberg copy number profiles were used as input. For samples with significant copy number aberrations, purity estimates derived from Battenberg were considered accurate. For samples without significant copy number aberration, purity was estimated with an initial

run of DPCLust on only balanced segments (1+1), to derive the VAF of the clonal cluster, which reflects the $2/\text{purity}$. This purity value was then used for a definitive run of DPCLust. All individual samples were run with DPCLust in single-sample mode, while PD38329 was also run as a multi-sample case.

Three-dimensional clustering plots were produced for PD38329 with the R packages `rgl` (0.1) and `htmlwidgets` (1.3). Phylogenetic reconstruction was performed using the pigeon-hole principle as previously described (Nik-Zainal *et al.*, 2012b). In brief, subclones were designated to be nested within a clone or another subclone if their combined CCF exceeded that of their parent.

2.17. Simple timing analysis (Chapter 4)

In brief outline, timing analysis comprised two approaches: estimating the ordering of a mutation relative to a copy number gain at the same locus, and estimating the timing of a whole genome duplication. Timing a mutation relative to its copy number context requires calculating the number of chromosome copies that possess that mutation. If all duplicated copies possess the mutation, the only feasible explanation is that the mutation occurred first and was duplicated with copy number gain. This analysis therefore requires the transformation of individual mutation allele frequencies into mutation copy number. This was performed using the equation:

$$MCN = \frac{VAF(\rho \times TCN + 2(1 - \rho))}{\rho}$$

MCN is Mutation Copy Number, ρ is the sample purity, TCN is the local total copy number.

Timing of whole genome duplication extrapolates this principle to all mutations at a copy number state, when it is simple enough to make a reasonable assumption about the steps through which that copy number change occurred; e.g. If most of a genome has two copies of each parental allele, it is likely this occurred in a single whole genome duplicating event. The proportion of mutations that occurred before and after that gain gives an estimate, in mutational time, of when that duplication event occurred. If only deamination (timing) mutations are considered this proportion can be used, with some adjustments, as a proxy for the timing in

real-time of the duplication. Specifically deamination (clock-like, C>T mutations at CpG dinucleotides) mutations were selected from regions of balanced gain (2+2) or LOH (2+0). A probabilistic approach to WGD timing was taken. Each mutation had a probability assigned for being at each mutation copy number state up to and including the major allele copy number state at that locus. This probability was calculated from a binomial distribution based on the sequencing depth and the measured allele frequency. An estimate for WGD was estimated as:

$$WGD = \sum \frac{pMut_2}{pMut_2 + pMut_1/2}$$

Where $pMut_x$ is the probability of a mutation being at a copy number x . $pMut_1$ is divided by x_{max} (in this example 2) to account for the increased genomic material after genomic duplication. The fractional estimate of WGD is then scaled to the patient's age at diagnosis for a real-time estimate of WGD. Confidence intervals are created by 1,000 bootstrap iterations, resampling the underlying mutations. In PD4922e, regions with major allele copy number of 3 in WGD samples were assumed, by parsimony, to have been acquired by a copy number gain subsequent to WGD. This is because the only alternative explanation would be a deletion after a second local duplication and since only a single genome duplication occurred this was considered unlikely. The timing of this additional gain was computed by calculating the proportion of clock-like mutations (again probabilistically) that were acquired prior to WGD (at MCN 3), between WGD and additional copy number gain (MCN 2) and after this gain (MCN 1).

2.18. Methylation array analysis and data pre-processing

DNA preparation and array handling were conducted by the UCL genomics facility using the Illumina Infinium Methylation pipeline. All samples were analysed with Illumina Infinium HumanMethylation450k or MethylationEPIC arrays. Raw data were quality controlled and pre-processed using the minfi R package (1.30.0) (Aryee *et al.*, 2014). Initial quality control of samples was conducted by Patrick Lombard and Iben Lyskjaer. Probes that were ambiguously mapped or located on the sex chromosomes, contain SNPs, or were not included

on both EPIC/450k platforms, were filtered. Filtered probe lists were provided by Martin Sill at DKFZ, Germany. Raw fluorescence values were corrected for background fluorescence using negative control probes and the minfi background correction function. Dye biases were corrected by scaling all values by a factor required to scale the mean red and green control probes intensities to 10,000. Probe methylation beta values were calculated using the default minfi function.

2.19. Methylation data clustering

For unsupervised clustering, a cohort was assembled with all available GCTs, malignant H3.3 tumours and chondroblastoma. This included data provided by David Jones at DKFZ, as detailed in Appendix 7.2.2. Though more osteosarcoma samples were available, a random sample of equivalent number (42) to the next largest group, GCT, was used so as not to dominate clustering effects. Unsupervised clustering was performed using the 5,000 most variable (by standard deviation) probes across samples. Beta values were transformed to a distance matrix of (1 - correlation values) for hierarchical clustering and dimension reduction analysis. T-distributed stochastic neighbour embedded (t-SNE using the Rtsne package(0.15)) and Principal Component Analysis plots were scrutinised to exclude significant non-biological batch effects introduced by different analysing centres, array platforms or sample types. Hierarchical clustering was then visualised using the packages ape (5.3) and dendextend (1.12.0). Two dimensional projections were produced using multi-dimensional scaling (MDS) in the MASS package (5.3). Samples were assigned methylation clusters (“M”, “G”, “C”, “Os”), with names to reflect the predominant disease type within that cluster, by cutting the hierarchical tree at the level of the 4 principle clades.

2.20. Methylation genomic analysis

Differentially methylated probes and regions were detected using the ChAMP package (2.14.0). Comparison was made between benign and malignant samples with methylation clusters concordant with their diagnoses. Bespoke

analysis of genome wide methylation difference was performed after calculating the signal-noise ratio at each probe position. This was calculated as the difference between the mean beta value for each group divided by the pooled variance across the groups. Segmentation was performed using adapted Circular Binary Segmentation (CBS) functions from the DNACopy package (1.58.0). Segmentation was adapted to require a minimum of 10 probes per segment and a segmentation alpha value of 1×10^{-10} to prevent spurious oversegmentation. Permutation analysis was conducted by permuting the diagnostic labels randomly and repeating segmentation as above. 10,000 permutations were performed and the threshold of significant SNR aberration set from the distribution of segment values (< -3 or > 3 , each $p < 0.003$).

Gene set enrichment analysis was performed using an adapted approach from the ebBayes function in the ChAMP package. In brief, a global test was used to assess the statistical difference between the diagnostic groups across all probes assigned to each gene. This probability value was used as the magnitude of the methylation difference for each gene, while a direction of change was inferred by the mean change in beta value for all gene probes in that gene. Gene set enrichment analysis was performed using the fgsea package (1.10.0) across the following gene sets downloaded from the Broad Mutational Signatures DataBase (MutSigDB) (Subramanian *et al.*, 2005): C2 Reactome, C2 Kegg, C5 Gene Ontology, C6 oncogenic pathways, C7 immunology pathways and the hallmarks pathways.

2.21. Methylation based copy number analysis

Methylation arrays, as a genomic hybridisation technique, provides information about the amount of DNA (copy number) across the genome. This requires normalisation using samples known to be diploid (in this case from a large panel of normal brain tissues) and I have used a approach similar to that implemented in SNP array and sequencing based copy number methods to calculate integer values for total copy number. Clearly without allelic information, allele specific copy number is not available.

Methylation array-based copy number analysis was performed using a binned intensity track generated using the conumee package (1.18.0). Raw data for 119 control male and female diploid samples were downloaded from GEO (GSE109381). Binned intensity values were fitted to integer copy number states using the principles underlying the ASCAT package (Van Loo *et al.*, 2010). In brief, a grid of possible purity and ploidy values was searched to minimise the sum of Euclidean distances between the intensity-based inferred number of copies and integer values.

2.22. Pan Cancer Analysis of Whole Genomes (PCAWG) variant calling

The PCAWG pipeline is described in detail elsewhere (PCAWG, 2020). All sequencing was performed using the Illumina HiSeq platform. Alignment was performed using bwa-mem (0.7.8-r455) (Li and Durbin, 2009) against the human reference hs37d5. Substitutions were detected using the consensus of the following methods: CaVEMan (1.5.1) (Jones *et al.*, 2016), MuTect2 (Cibulskis *et al.*, 2013), MuSE (1.0rc) (Fan *et al.*, 2016) and a bespoke method from DKFZ. Small insertions and deletions were detected with the consensus of: cgpPindel (1.5.7) (Ye *et al.*, 2009), Platypus (0.7.4) (Rimmer *et al.*, 2014), MuTect and SMuFIN (mod. 26/10/2014) (Moncunill *et al.*, 2014). Copy number was called using the consensus of the following methods: ABSOLUTE (Carter *et al.*, 2012), ACEseq (Kleinheinz *et al.*, 2017), Battenberg (Dentro *et al.*, 2017, Nik-Zainal *et al.*, 2012b), cloneHD (Fischer *et al.*, 2014a), JaBbA (Li *et al.*, 2016, Medvedev *et al.*, 2010, Oesper *et al.*, 2012), and Sclust (George *et al.*, 2015, Peifer *et al.*, 2015).

2.23. PCAWG structural variant detection and classification

Structural variants (SVs) were detected using the consensus of the following algorithms: DELLY (0.6.6) (Rausch *et al.*, 2012), BRASS (4.012) (<https://github.com/cancerit/BRASS>), SvABA (Wala *et al.*, 2018), and dRanger (BROAD pipeline, not publically available).

Clustering and classification of structural variants into footprints is as described (Li *et al.*, 2020). This entails the following steps:

1. Breakpoint coordinates re-defined from clipped reads.
2. Rearrangement breakpoints merged with copy number data
3. Individual SVs clustered into clusters and footprints
4. SV clusters and footprints heuristically refined
5. Artefactual fold-back-type SVs filtered if insufficient supported
6. Balanced breakpoints with overlapping breakpoints recaptured from soft clipping patterns.

Chromoplexy events were selected from SV clusters containing at least two balanced footprints.

2.24. Application of ChainFinder

Inputs for ChainFinder (1.0.1) were formatted from PCAWG consensus minor allele copy number and structural variants. Background rearrangements were generated by concatenating all structural variants within each disease group. Default parameters were used, with a deletion threshold set to 0.95, unless the sample had been designated whole-genome duplicated by consensus copy number calling, in which case it was set to 1.9.

2.25. Detection of deletion bridges

In order to detect deletions between pairs of breakpoints flagged by ChainFinder as being adjacent, coverage was evaluated using BEDtools coverage (2.26.0). Coverage was calculated in bins of 100bp if the breakpoint pair were within 1kb or in 10 equal sized bins if more than 1kb. Coverage was also calculated for 10 bins of the same size on either side of the breakpoint pair. Coverage values were compared with a one-sided Wilcoxon signed ranks test. Observed p-values were then compared to a distribution of expected p-values by permutation of SV breakpoints around the genome across all samples.

2.26. PCAWG mutation clustering and timing

Mutation clustering is described in detail elsewhere (Dentro *et al.*, 2018). In brief, variant read counts and total sequencing depth for substitutions were converted to cancer cell fractions (CCF) using purity, and total copy number at the mutation locus. These were then clustered to identify mutations that were co-inherited in the same proportion of cells. PCAWG utilised a consensus of multiple methods to do this, included methods were: BayClone-C (Sengupta *et al.*, 2015), Ccube (Yuan *et al.*, 2018), CliP (<https://github.com/wwylab/CliP>), cloneHD (Fischer *et al.*, 2014b), CTPsingle (Donmez *et al.*, 2017), DPCLust (Nik-Zainal *et al.*, 2012b), PhylogicNDT (Leshchiner *et al.*, 2018), PhyloWGS (Deshwar *et al.*, 2015), PyClone (Roth *et al.*, 2014), Sclust (Cun *et al.*, 2018) and SVclone (Cmero *et al.*, 2017). Note that SVclone computes CCFs from structural variants. The probability for each mutation to be assigned to the clonal cluster was required for punctuated event clonal assessment.

Mutation timing is described in detail by Gerstung *et al.* (Gerstung *et al.*, 2018) but is a generalised expression of the principles described in methods 2.17. This uses the algorithm MutationTime.R (<https://github.com/gerstung-lab/MutationTimeR>). This generated a probability that any mutation was found at single or multiple copy number states. In gained regions, if the mutation copy number is compatible with the major allele copy number, then the mutation occurred prior to the gain and is considered 'clonal early', if lower and clonal it occurred after the gain and is considered 'clonal late'. If a mutation falls in an ungained locus and the mutation is clonal, the mutation is designated 'clonal NA'.

2.27. Defining kataegis and chromothripsis

Full details of methods to identify kataegis, devised by Jonas Demeulemeester, and chromothripsis, devised by Maxime Tarabichi, are detailed elsewhere (PCAWG, 2020). In brief, kataegis was identified by:

1. Clustering mutations by proximity, taking into the mutational burden of the sample
2. Testing clustered mutations for:

- a. Consistency of mutational context within the cluster (based on mutational signatures and trinucleotide contexts)
- b. Strandedness of mutated base (ie the same base change relative to the reference strand)
- c. Phasing of mutations (ensuring they happen on the same parental chromosome and are not sequential). Anti-phased clusters are excluded.

Chromothripsis was identified as the overlap of chromothripsis calls identified by Cortes-Ciriano (Cortés-Ciriano *et al.*, 2018) and the following method, devised by Maxime Tarabichi:

1. Clustering copy number and structural variant breakpoints to identify clusters with an average segment length >3MB with a minimum of 30 breakpoints per chromosome arm.
2. Filter clusters by:
 - a. The distribution of segment lengths within the cluster – the distribution should be non-random (i.e. not exponential)
 - b. The number of copy number states, using a threshold scale by the number of segments
 - c. The number of different rearrangement types, chromothripsis should be a random mixture of different types

Chromothripsis clusters were permitted to span chromosomes or regions if sufficient structural variants linked them.

2.28. Punctuated event timing

The method to define the timing and clonality of punctuated events was devised in collaboration with Jonas Demeulemeester and Maxime Tarabichi and is also described elsewhere (PCAWG, 2020).

For every punctuated event, the probability of being clonal was calculated as the normalized likelihood from the clonal assignment probabilities of the constituent substitutions or SVs. For instance, for an event i involving $N = 1 \dots j$ variants, each with an associated probability $1 - p_{sub,j}$ of being clonal in the tumour sample, the likelihood of being clonal was determined as $\prod_{j=1}^N (1 - p_{sub,j})$ and of

being subclonal as $\prod_{j=1}^N p_{sub,j}$. The likelihoods were normalised to yield probabilities for (sub)clonality of the event ($p_{cl,i}$ and $p_{sub,i}$).

The timing of every event was also computed as the probability of the event being clonal early, clonal late or clonal NA ($p_{early,i}$, $p_{late,i}$ and $p_{NA,i}$, respectively). This used the probabilities as derived by MutationTimer. Normalised likelihoods were computed using the variants in the event: stratified by consensus gain/LOH status (PCAWG consensus copy number), weighted by the fraction of variants in each class, and summed according to the rules for distinguishing clonal early/late/NA (as described above) to obtain the final probabilities.

The odds of observing clonal versus subclonal events of different types were

computed for every cancer type by bootstrapping the ratio $\frac{\frac{\sum p_{cl,i} + 0.5}{\sum p_{cl,sim,i} + 0.5}}{\frac{\sum p_{sub,i} + 0.5}{\sum p_{sub,sim,i} + 0.5}}$ where 0.5

represents a pseudocount (i.e. a single event with $p_{cl,i} = p_{sub,i} = 0.5$) and $p_{cl,sim,i}$ and $p_{sub,sim,i}$ are the clonal and subclonal assignment probabilities of a simulated event matched to observed event i . For every punctuated event observed we simulated 10,000 comparable events. These were simulated by sampling the same number of substitutions or SVs from the background of non-punctuated variants with identical gain/LOH status in that tumour sample. Clonal and subclonal assignment probabilities ($p_{cl,sim,i}$ and $p_{sub,sim,i}$), as well as the probability of being clonal early, late or NA ($p_{early,sim,i}$, $p_{late,sim,i}$ and $p_{na,sim,i}$) were computed for the simulated events as described above. To obtain the median odds ratio and 2.5th/97.5th percentiles, 10,000 bootstrap replicates of the observed events were generated with a different sampled set of events used for each iteration. During the bootstrap, events were sampled with weighting according to $\frac{1}{(\# \text{ events in sample})}$ in order to give equal weight to samples with different numbers of punctuated events. The odds of observing clonal early versus clonal late events were computed similarly by bootstrapping the ratio

$$\frac{\frac{\sum p_{early,i} + 0.5}{\sum p_{early,sim,i} + 0.5}}{\frac{\sum p_{late,i} + 0.5}{\sum p_{late,sim,i} + 0.5}}.$$

BLANK PAGE

Chapter 3. Osteoblastoma is defined by recurrent rearrangements in *FOS/FOSB*

3.1. Introduction

Osteoblastoma is a rare benign bone-forming tumour, the mutational basis of which was unexplored prior to this work.

3.1.1. Clinical and epidemiological features of osteoblastoma

Osteoblastoma typically presents with progressive bone pain in the 2nd and 3rd decade with a 2:1 male to female preponderance (Czerniak and Dorfman, 2016, Yalcinkaya *et al.*, 2014). The most common sites of disease are the medulla of the long bones and the neural arch, which may result in neurological sequelae from extension into and around the vertebral body. Osteoid osteoma is pathologically similar, indistinguishable from osteoblastoma apart from its size of less than 2 cm (Fletcher *et al.*, 2013). Osteoblastoma and osteoid osteoma together are the most common benign bone-forming tumours. Exact incidence is unknown but they are estimated to comprise 15-20% of primary benign bone tumours, which is likely to represent a total of fewer than 100 cases per year in the UK (Hakim *et al.*, 2015, Zhang and Rosenberg, 2017).

Imaging is ordinarily by plain radiograph and computed tomography, revealing areas of lysis, bone formation, and mineralisation which is typically more organised in osteoid osteoma and more irregular in osteoblastoma, potentially with soft tissue extension (Atesok *et al.*, 2011, Yalcinkaya *et al.*, 2014). Treatment is surgical, either with curettage for smaller lesions or resection and reconstruction for those that are larger and potentially structurally comprising (Atesok *et al.*, 2011). Radiofrequency or cyroablation have been attempted for smaller lesions.

3.1.2. Pathology

Histologically, tumour cells appear as plump osteoblasts, with evidence of metabolic activity and occasional normal mitoses (Yalcinkaya *et al.*, 2014, Zhang and Rosenberg, 2017). The neoplastic cell population is low but they are 'busy' highly vascularised lesions with disorganised trabecular and woven bone. The case for osteoid osteoma and osteoblastoma representing one disease has been proposed on the basis of detailed pathological review by Barlow *et al.* but indisputable biological evidence is still lacking (Barlow *et al.*, 2013). Osteoblastoma-like osteosarcoma is a recognised malignant bone-forming tumour which can be difficult to distinguish from osteoblastoma based on histological appearance (Gambarotti *et al.*, 2019). This benign from malignant distinction has significant clinical implications both on prognosis and on the necessity for multi-modal therapy. There were no molecular markers available to aid in this distinction prior to this study. Malignant transformation of osteoid osteoma or osteoblastoma has not been reported.

3.1.3. Genomics

The genomics of osteoblastoma has only previously been explored at the resolution of the chromosome. Cytogenetics analysis observed a number of different rearrangements but only recurrently involving chromosomes 1 and 14. Neither chromosome was rearranged in all cases (Baker *et al.*, 2010, Giannico *et al.*, 2009). Recurrent deletions in the short arm of chromosome 22 have also been reported in a small number of cases (Baruffi *et al.*, 2001). Chromosome 22 deletions were independently reported after the analysis of SNP arrays, but were still confined to the minority of cases (Nord *et al.*, 2013). On the basis of this chromosome 22 deletion, Nord *et al.* had suggested a role for aberrant Wnt signalling in osteoblastoma development. They argued that the presence of two negative regulators of the Wnt pathway, *ZNRF3* and *KREMEN1*, located on the deleted segment on chromosome 22, explain expression differences in Wnt signalling members from expression arrays. The deletions were, however, hemizygous, without evidence of the second mutational hit, required for negative

regulators. Copy number analysis was also performed without matched normal controls therefore these cannot be established as somatic variants. The expression array analysis was also conducted with osteosarcoma as a comparator, which is more likely the origin of the signal.

In summary, osteoblastoma and osteoid osteoma likely represent a single type of benign neoplasm but biological evidence is lacking for this. There is a clinical need for a diagnostic biomarker, to differentiate it from osteoblastic osteosarcoma and the relatively unexplored genomic landscape offers the potential for this.

3.2. Aim

To explore the genomic landscape of osteoblastoma and osteoid osteoma.

3.2.1. Objectives

- Identify a mutation or mutational pattern that defines and potentially unifies the diseases of osteoid osteoma and osteoblastoma
- Define a usable clinical diagnostic test to aid in the differentiation of osteoblastoma from osteoblastoma-like osteosarcoma

3.3. Results

The starting point for our investigation was a discovery cohort of six cases, with the typical clinical characteristics for osteoblastoma and osteoid osteoma (**Table 4**). DNA, extracted from blood for germline sequences and frozen tumour tissue for somatic changes, was subjected to whole genome sequencing. Tumour and blood samples were sequenced to a median depth of 44x (range 39 to 48) and 36x (range 30 to 42) respectively. All classes of somatic variants were catalogued and scrutinised for putative drivers; substitutions, indels, copy number variants and structural rearrangements (**Methods 2.7**). Transcriptome sequencing was conducted to corroborate DNA variants and identify gene fusions.

Sample	Diagnosis	Age	Gender	Anatomical Site	Bone Site	Size (mm)
PD7519	Osteoblastoma	33	M	Femur, right	Cortical	15
PD7521	Osteoblastoma	25	M	Femur, right	Cortical	40
PD7525	Osteoblastoma	3	M	Humerus, left	Medullary	30
PD13480	Osteoblastoma	13	F	Lumbar vertebra, L3	Medullary	35
PD13481	Osteoblastoma	11	F	Cervical vertebra, C5	Medullary	30
PD13482	Osteoid Osteoma	18	M	Lumbar vertebra, L1	NA	12

Table 4 Clinical characteristics of discovery cohort.

3.3.1. Osteoblastoma possess few somatic alterations

There was a scarcity of somatic changes to the osteoblastoma genome across the sequenced discovery cohort. There were a median of 319 substitutions (range 123 to 700) and 28 indels (range 14-50) per genome. This is quiescent when compared with the genomes of malignant tumours: osteosarcoma (**Figure 12**) or a pan-cancer cohort (**Figure 13**). Very few variants affected coding sequences of genes and none were plausible drivers (**Methods 2.9**).

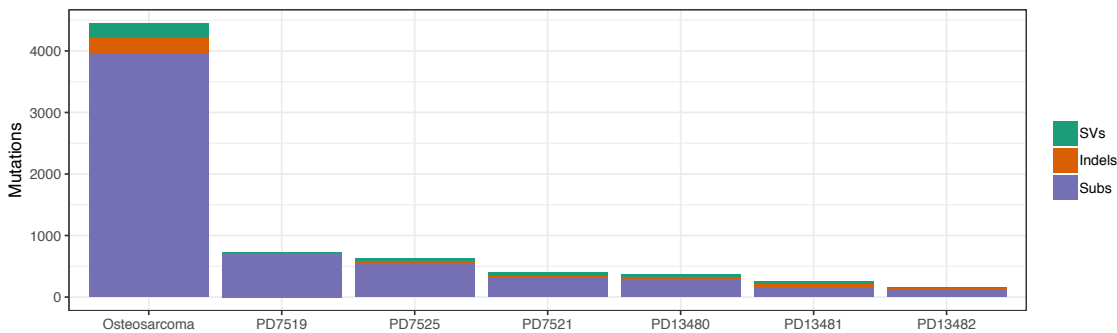


Figure 12 Osteoblastoma mutation burden compared with osteosarcoma. Filtered mutation numbers are shown relative to the median number of mutations in a published cohort of whole genome sequenced osteosarcomas (Behjati *et al.*, 2017)

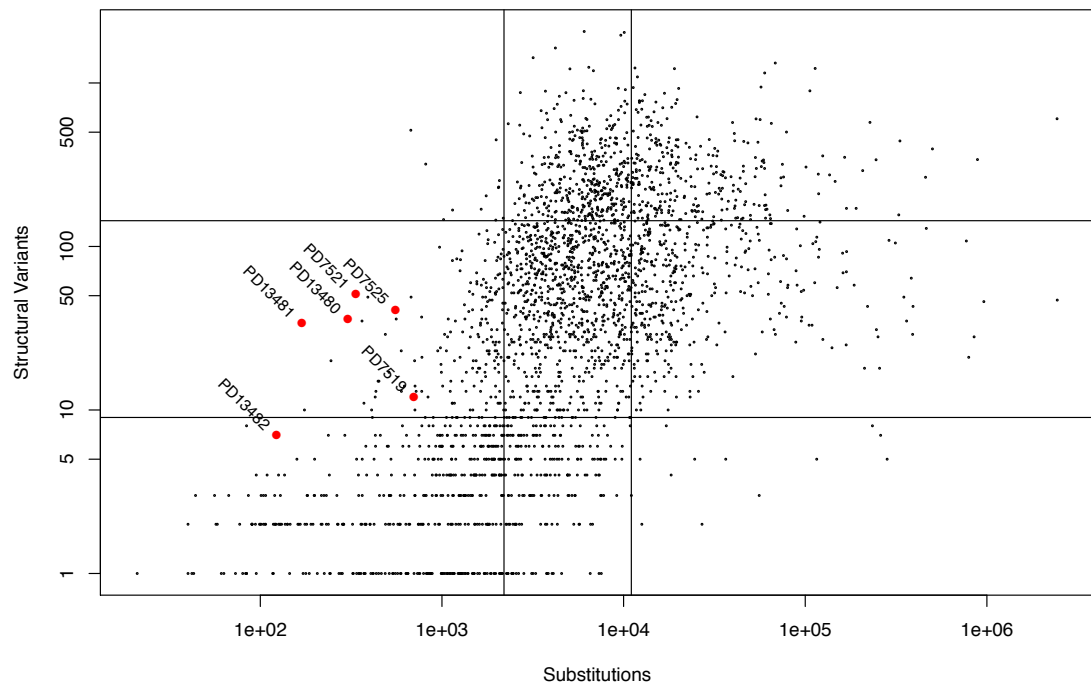


Figure 13 Osteoblastoma mutation burden compared with all cancer types. Raw substitution counts and structural variants compared with mutation burdens across cancer types from the Pan Cancer Analysis of Whole Genomes consortium (**Chapter 5: Identification and timing of chromoplexy across cancer types**). Osteoblastoma samples are superimposed with red points and labels. PCAWG mutation counts are derived from consensus mutation calls and are therefore conservative compared to those presented for osteoblastoma. 25th and 75th centile lines shown as black lines.

Copy number profiles revealed predominantly diploid genomes with few aberrations. Raw data and structural variants were scrutinised for chromosome 22 deletions corresponding with those previously reported, but none were identified (**Figure 14 and Appendix 7.1.5**).

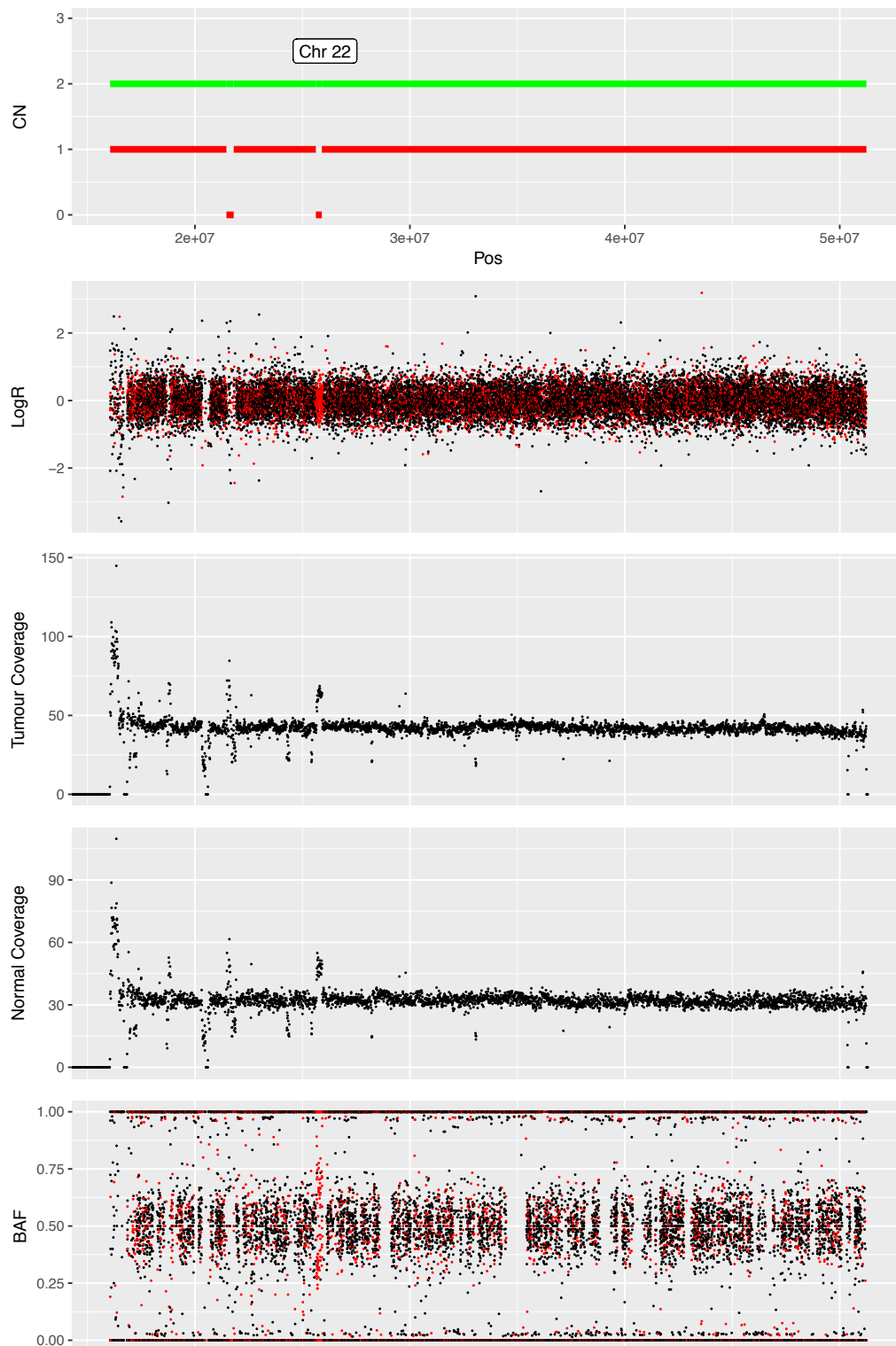


Figure 14 Chromosome 22 copy number data for PD13480. Panels (in descending order): ASCAT derived segmented copy number (green total copy number, red is minor allele copy number), raw LogR with known commonly aberrant germline SNPs coloured in red, tumour and normal raw coverage, raw B Allele Frequency (BAF), again with commonly aberrant SNPs in red. The only aberrations seen in the copy number, between 20-30MB are in known regions of germline aberration and reflect germline duplication.

3.3.2. Recurrent FOS structural rearrangements

On this seemingly barren landscape of somatic changes all samples possessed structural variation in the AP-1 family transcription factor genes, *FOS* or *FOSB*. The structural rearrangement caller, BRASS, was able to identify breakpoint regions in *FOS* in 4/6 samples and exact breakpoints in 2 of these. Very few other structural variants were identified to breakpoint resolution across all samples (**Table 5**). Manual analysis and local assembly confirmed these rearrangements and could identify a *FOS* rearrangement in a further sample, PD13482a (**Appendices 7.1.1, 7.1.2, 7.1.3**).

Sample	Chr1	Pos1	Str1	Gene1	Chr2	Pos2	Str2	Gene2	micro_homology	non_template_insertion	clean_break	sv_type
PD13480a	1	155642647	-	YY1AP1	2	189982974	-	COL5A2	.	GTATACTCAATGAGTATAAAATTA	.	translocation
PD13480a	5	14851060	-	ANKH	14	75747531	+	FOS	AT	.	.	translocation
PD13481a	14	75747834	-	FOS	15	81099417	-	KIAA1199	.	CGAGCCCTTTG	.	translocation
PD13481a	4	54159002	-	SCFD2	4	54160189	+	SCFD2	TTTTCTTTATCCAGT	.	.	inversion
PD7521a	2	199519392	-	AC019330.1	2	206321709	-	PAR3B	TA	.	.	tandem-duplication
PD7521a	2	206318313	-	PAR3B	2	240946624	-	NDUFA10	CTGT	.	.	tandem-duplication
PD7521a	2	208707925	-	PLEKHM3	2	221489402	+	AC067956.1	.	.	Clean_break	inversion
PD7521a	2	208918645	-		2	237995252	-	COPS8	TTT	.	.	tandem-duplication
PD7521a	2	215604968	+	BARD1	2	219961814	-	NHEJ1	A	.	.	inversion
PD7521a	2	216052153	-		2	237617461	-		AAT	.	.	tandem-duplication
PD7521a	2	216398464	-	AC012668.1	2	241270436	+		TACA	.	.	inversion
PD7521a	2	218592386	-		2	240578966	-		.	.	Clean_break	tandem-duplication
PD7521a	2	225750015	-	DOCK10	2	237994578	-	COPS8	CC	.	.	tandem-duplication
PD7521a	8	84196330	+		15	93569737	+	CHD2	C	.	.	translocation
PD7525a	10	53448211	+	PRKG1	10	53549164	+	PRKG1	AGC	.	.	deletion
PD7525a	13	45016456	+	TSC22D1	14	91305920	+		.	AAATAT	.	translocation
PD7525a	13	45016489	-	TSC22D1	19	45971996	-	FOSB	CA	.	.	translocation

Table 5 Structural rearrangements in osteoblastoma genomes identified by the genomic structural variation caller BRASS.

Structural rearrangements defined to breakpoint resolution by BRASS alone are shown. Therefore not only 3/6 FOS/FOSB rearrangements are shown as those in PD13482a, PD7519a and PD7521a were either only called in a breakpoint interval or required some manual detection.

All *FOS* breakpoints possessed either an intronic (3/5 cases) or intergenic (2/5 cases) partner. Expression of the fusion transcript was therefore visible as aberrant spikes in RNA sequencing coverage (Figure 15). cDNA sequencing reads were used to confirm single rearrangements for each sample, suggesting these were all mono-allelic events. In addition, *FOS* rearrangements were all successfully validated by Sanger sequencing (Figure 16).

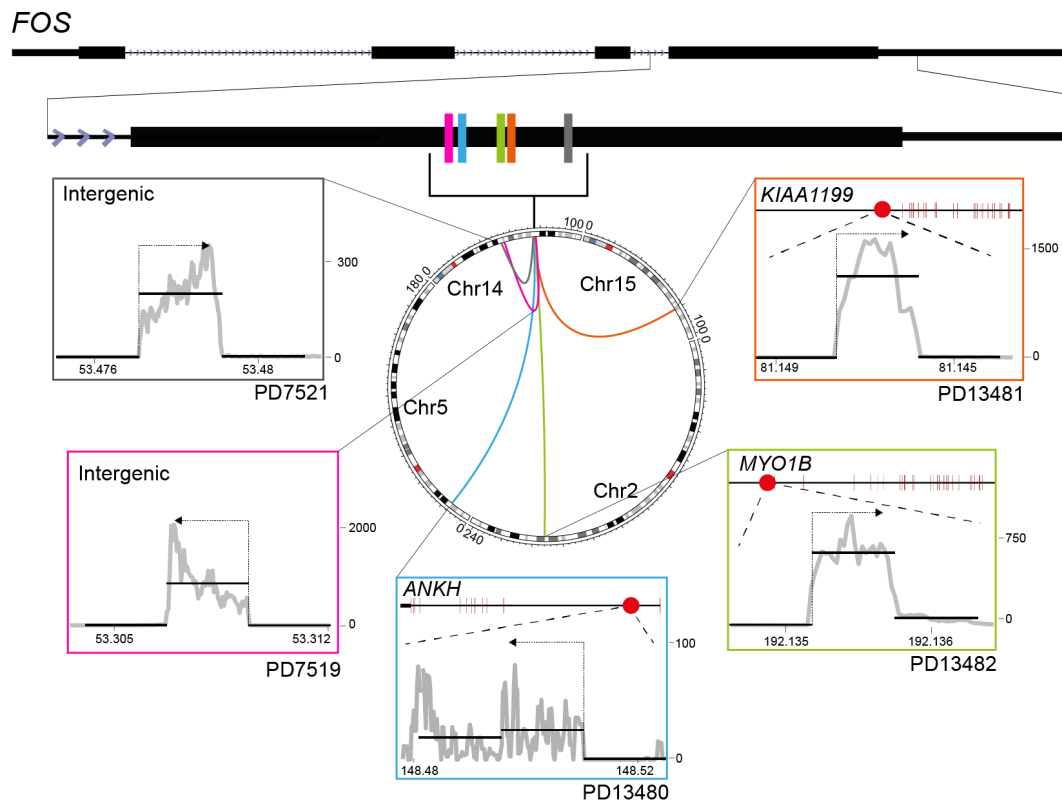


Figure 15 *FOS* breakpoints in osteoblastoma. Schematic view of *FOS* with breakpoints in exon 4 shown as coloured bars. The centre circos plot shows all structural rearrangements in these samples involving these chromosomes. Surrounding plots are of RNA sequencing coverage adjacent to the fusion partner breakpoint. Where relevant the fusion partner gene is shown in schematic above, exons are red vertical segments, red dots are stop codons introduced within intronic sequence. The direction of fusion transcription is indicated by arrows. Black segments are the median coverage outside or inside of the aberrant region of transcription.

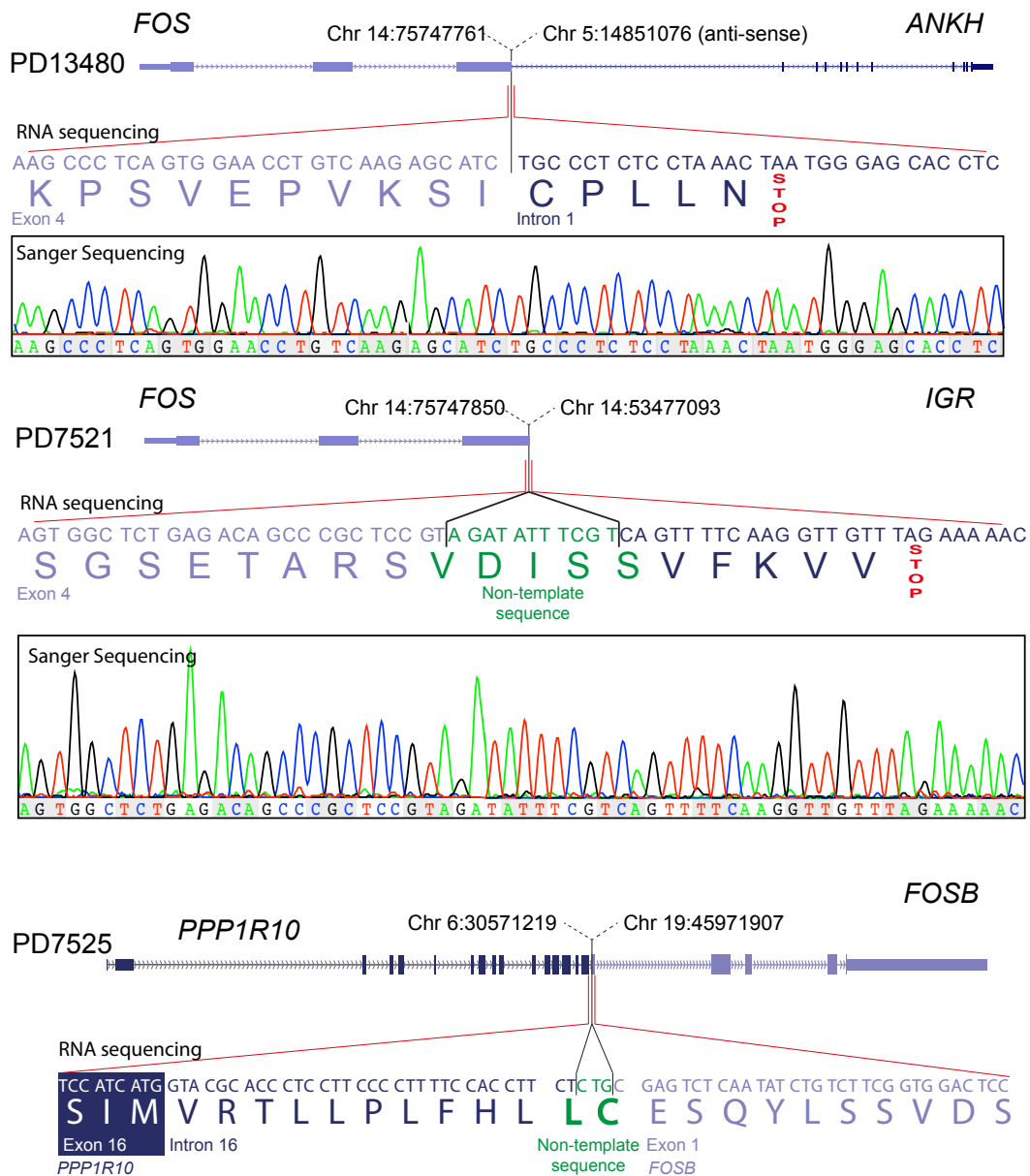


Figure 16 Validation of *FOS/FOSB* rearrangements. Example validation plots for 3/6 samples (see Appendix 7.1.3 for remaining samples). Configuration of *FOS* or *FOSB* rearrangement is shown with RNA sequencing confirmed base and amino acid sequence. Sanger sequencing trace for validation of *FOS* rearrangements is shown beneath.

3.3.3. *FOSB* rearrangement by chromoplexy

One sample, PD7525, possessed a complex rearrangement involving *FOSB* which resembled the chained pattern of rearrangement, termed chromoplexy, as previously reported in prostate cancer (Baca *et al.*, 2013). This *FOSB* rearrangement was detected by both BRASS (Table 5) and RNA fusion callers.

This entailed a complex but complete cycle of rearrangements in which any breakpoint can be traced back to itself via four other chromosomal breakpoints (**Figure 17**). Two rearrangements, required to complete this cycle, that both involved chromosome 16, were not identified by rearrangement callers. This was because they were mapped to a sequence on chromosome that is ambiguously mapping across chromosome 16 because of sequence homology. The site of this rearrangement with the greatest read support was selected. Each chromosome involved contained a pair of breakpoints between 30-140bp of each other, with no intervening change in sequencing coverage, unsurprising considering their close proximity.

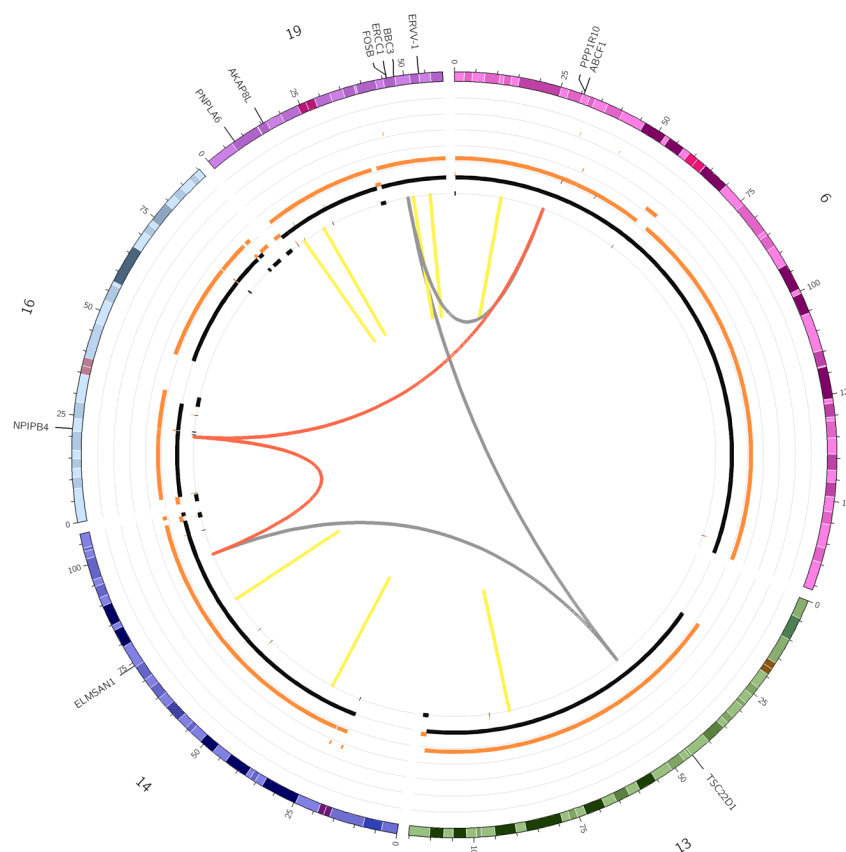


Figure 17 PD7525 Chromoplexy around FOSB. Copy number is shown with the orange (total copy number) and black (minor allele) segments counted from a minimum of zero on the innermost ring. Structural variants are shown as coloured internal links (yellow – inversions, grey – translocations, red – manually detected rearrangements onto an ambiguously mapping region of chromosome 16: the position presented are those best supported by discordant and breakpoint-spanning DNA and cDNA sequencing reads.)

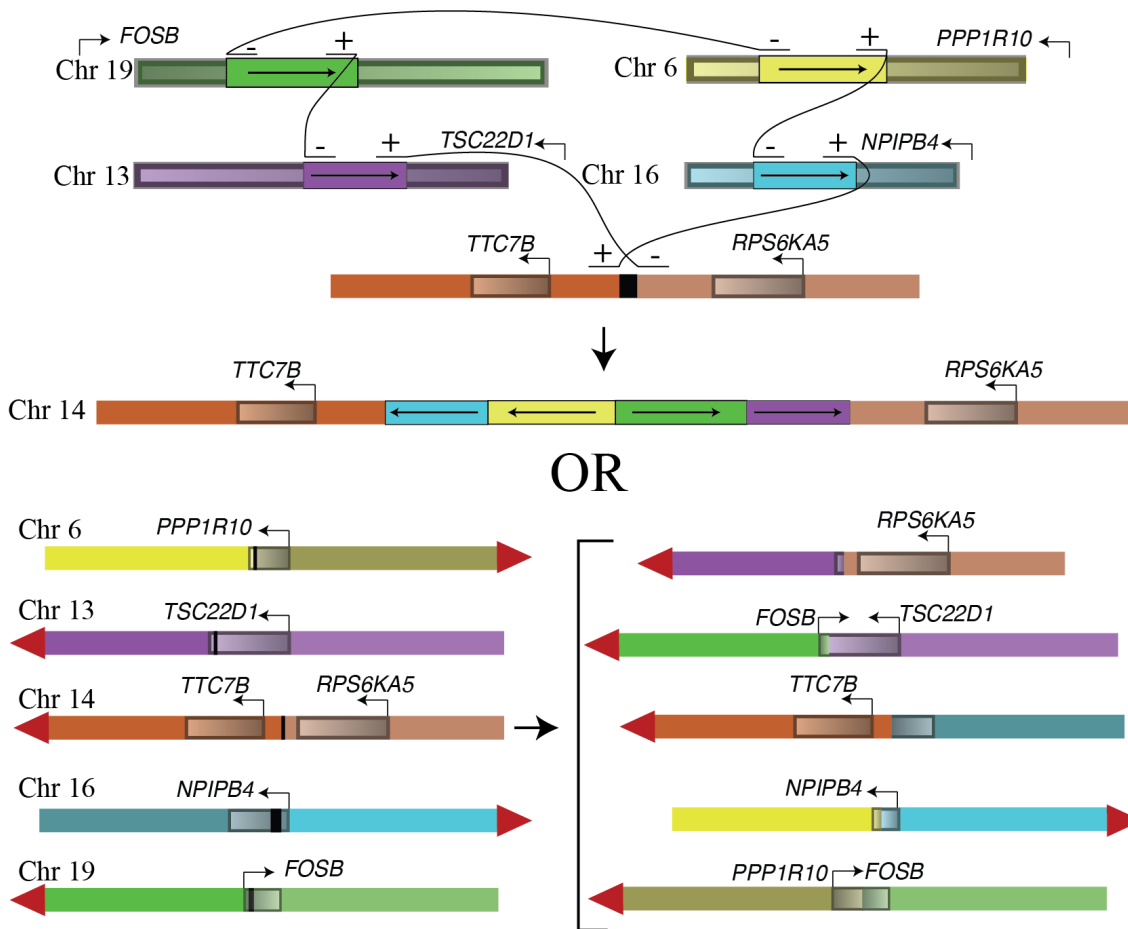


Figure 18 Alternative chromoplexy derivative configurations. Two hypothetical results from complex rearrangement: a) a templated insertion cycle where only the intervening sequence between breakpoint pairs is duplicated onto a receiving chromosome, in this case onto chromosome 14, leaving the original sequence in tact b) a conventional chromoplexy rearrangement with a series of chromosomal translocations (the direction to centromere are shown as red triangles). Note that in both circumstances there is expected to be a gain in genomic material between the breakpoints. Read evidence precludes a).

There was therefore no evidence of spanning deletions, so-called ‘deletion bridges’, or genomic gains. In 4/5 breakpoint pairs the orientation of breakpoints was consistent with the possibility of a ‘templated insertion’ cycle, in which merely the intervening sequence is copied from one fragment to the next (**Figure 18**) (Li *et al.*, 2020). The string of fragments would then be found on a ‘host’ chromosome, in this case chromosome 14. This possibility is precluded by the small distances between the breakpoints as both read pairs and individual sequencing reads would be expected to span multiple fragments. Neither of these were observed in DNA or cDNA. The chromoplexy cycle therefore braids together the five chromosomes into five derivative chromosomes without any large losses of

genomic material and without generating di-centromeric chromosomes. All chromoplexy rearrangements were confirmed with spanning cDNA reads (**Figure 16**).

3.3.4. The impact of *FOS* and *FOSB* rearrangement

Unusually, all *FOS* rearrangements were within an exon, within a narrow 100bp window of the 4th exon. Stop codons were induced at or immediately after each breakpoint (**Figure 16**), therefore no coding sequence was introduced from the gene partners. In two cases, rearrangement partners were with intergenic regions. Transcription cleavage motifs were also observed within 30bp of the end of each fusion transcription spike, supporting the fusion not contributing additional coding sequence to *FOS*. The configuration of the rearrangement would also suggest *FOS* retains its native promoter. These observations are supported by the re-analysis of RNA sequencing data from recently published epithelioid haemangioma, which also harbour *FOS* rearrangements (van IJzendoorn *et al.*, 2015, Huang *et al.*, 2015). Once again, all breakpoints were observed within the same narrow window of exon 4, stop codons were introduced in the vicinity of the breakpoint and no coding sequenced was fused with the *FOS* transcript (**Figure 19**).

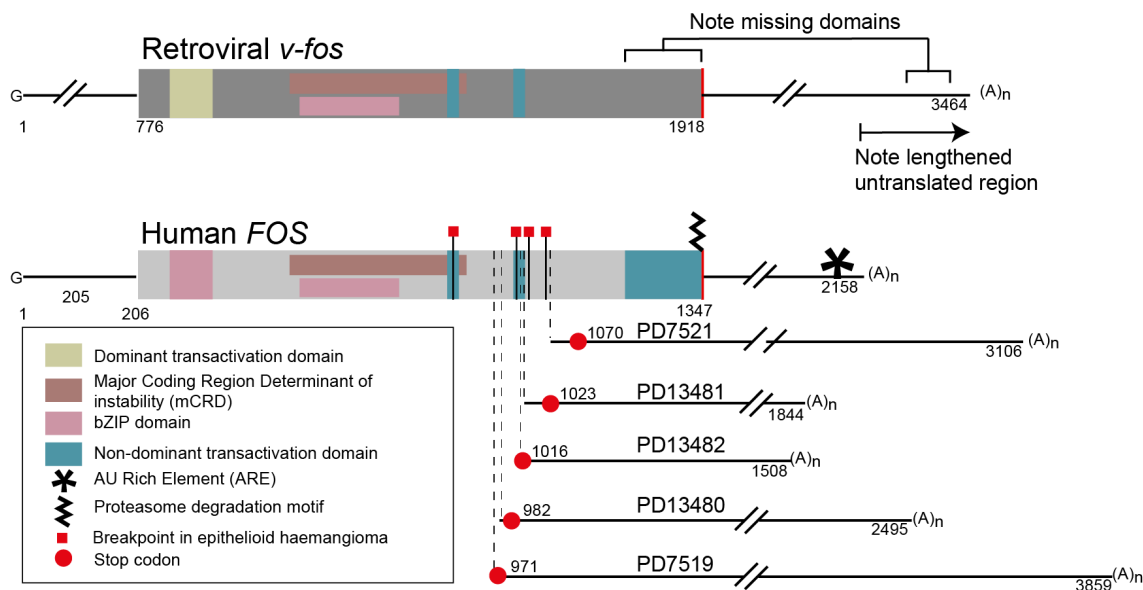


Figure 19 *FOS* rearranged transcripts. Schematics of the transcript of *FOS* in comparison to murine retroviral *v-fos* (**Results 3.4.2**). Key regulatory and functional elements of the gene are shown. Breakpoints from published cases of *FOS* rearranged Epithelioid haemangioma are included. The expected length of the transcript untranslated region is also shown (Methods)

The complex rearrangement of *FOSB* generates a fusion gene of *PPP1R10* in frame with exon 1 of *FOSB*. This brings the expression of *FOSB* under the control of the *PPP1R10* promoter. The benign vascular tumours, pseudomyogenic haemangioendothelioma and epithelioid haemangioma have also been reported to have *FOSB* breakpoints in the same region of exon 1 though with varied fusion partners; *SERPINE1*, *ZPF36*, *WWTR1* and *ACTB* (Walther *et al.*, 2014, Antonescu *et al.*, 2014, Ide *et al.*, 2015, Zhu *et al.*, 2019) (**Figure 20**).

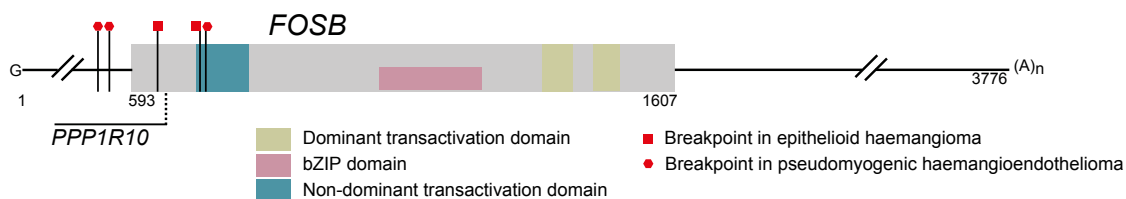


Figure 20 *FOSB* rearrangements. Schematic of *FOSB* transcript with published vascular tumour breakpoints annotated

3.3.5. Validation of *FOS* or *FOSB* alterations

In order to validate these findings a cohort of 55 formalin-fixed paraffin-embedded (FFPE) histological typical osteoblastoma and osteoid osteoma was assembled

(**Appendix 7.1.7**). These were subjected to fluorescence in-situ hybridisation (FISH) with bespoke probes designed to flank the *FOS* and *FOSB* breakpoints. Breakapart signals were observed in 1 and 48 samples in *FOSB* and *FOS* respectively (89% in total; **Figure 21** and **Appendix 7.1.7**).

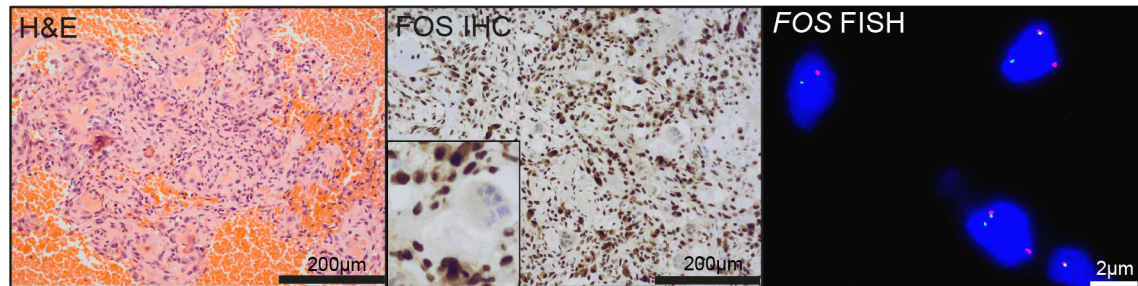


Figure 21 Validation for FOS alteration. H&E appearances of a typical osteoblastoma with extensive vascularisation surrounding a cellular region with intermixed osteoid deposition. FOS immunoreactivity is marked and exceeds that seen in standard positive controls specimens. In the right panel, all fully visible nuclei show a breakapart with clear separation of the green and red probe signals targeted to *FOS*.

Osteoblastoma are however typically of low neoplastic cell content (Czerniak and Dorfman, 2016), rendering FISH insensitive because of the small number of tumour nuclei assessable in any sample. Intrachromosomal rearrangements, as seen in 2/5 cases in the discovery cohort, may also be difficult to identify as the native configuration may either be retained, in tandem duplications, or insufficiently separated. Alternative evidence was therefore sought for alteration in *FOS* using immunohistochemistry. Of the 6 cases, negative by FISH, 3 were suitable for assessment and all 3 demonstrated marked and strong nuclear immunoreactivity for the retained N-terminus of *FOS*. *FOSB* immunohistochemistry was uninformative in osteoblastoma, consistent with experience in decalcified tumours (Hung *et al.*, 2017). In summary, all cases that could be assessed by at least one of FISH or immunohistochemistry demonstrated an alteration in *FOS* or *FOSB*.

3.3.6. *FOS* or *FOSB* rearrangement is unique to osteoblastoma amongst bone-forming tumours

To assess the potential of *FOS* or *FOSB* rearrangement as a diagnostic biomarker of osteoblastoma, their specificity was explored in those sarcoma

types that exhibit some histological overlap with osteblastoma. FOS immunoreactivity was assessed in 183 cases of osteosarcoma, including 97 cases of osteoblastic osteosarcoma, and 17 cases of angiosarcoma. In keeping with previous reports, FOS immunoreactivity was seen in osteosarcoma samples but only one tumour possessed a distribution and intensity of immunoreactivity comparable to that observed in osteoblastomas (Weekes *et al.*, 2016). On FISH testing this sample had no breakapart signals in *FOS* or *FOSB*, but copy number gains were noted (**Appendix 7.1.6**).

To identify any genomic changes involving *FOS* and *FOSB* in osteosarcoma, I then re-examined two published osteosarcoma series, totalling 55 whole genome sequences; none of which contained breakpoints in *FOS* or *FOSB* (Chen *et al.*, 2016, Behjati *et al.*, 2017). Finally, there were no similar rearrangements in the whole genome sequences of 2,652 non-osteoblastoma tumours (Li *et al.*, 2020).

3.4. Discussion

3.4.1. *FOS* and *FOSB* alteration as diagnostic biomarkers of osteoblastoma

This work represents the most comprehensive and detailed genomic exploration of osteoid osteoma and osteoblastoma conducted to date, and reveals a landscape defined by a simple mutational pattern. Amidst a paucity of all classes of somatic rearrangements, there is a highly recurrent rearrangement of the AP-1 transcription factor *FOS* or in a minority of cases its paralogue, *FOSB*. This is consistent with the previous reports of cytogenetic abnormalities involving chromosome 14 (Baker *et al.*, 2010, Giannico *et al.*, 2009). Since 2/6 of our sequenced cases involved intrachromosomal *FOS* rearrangements without copy number changes, it is expected that cytogenetic analysis would have under reported these alterations. I did not identify the chromosome 22 deletions previously reported (Nord *et al.*, 2013, Baruffi *et al.*, 2001). These may occur in a minority of cases, not represented in our small discovery cohort, but considering the ubiquity of *FOS* and *FOSB* rearrangement they are likely to be of secondary importance.

The readily available clinical techniques of FISH and immunohistochemistry were able to detect these alterations in a large cohort of osteoblastoma and distinguish

them from their diagnostic mimic, namely osteoblastic osteosarcoma. *FOS* or *FOSB* alteration therefore offers significant potential as a rapidly translatable diagnostic biomarker (Amary *et al.*, 2019a).

3.4.2. *FOS* rearrangement resembles the murine retroviral *v-fos*

The functional impact of these alterations in *FOS* and *FOSB* can only be hypothesised from these data. Physiologically, *FOS* levels are tightly regulated by rapid degradation at both transcript and protein level. Both translation-dependent mechanisms that ensure rapid mRNA degradation are likely to be disrupted by the observed rearrangements (**Figure 19**). Degradation depends on a length-dependent interaction between the poly-A tail and an exon 3 domain (known as the major Coding Region Determinant of instability) (Grosset *et al.*, 2000), all fusion transcripts are predicted to have a different length to the wild-type transcript. There is also an independent AU-rich element in the 3' untranslated region which is absent from all fusion transcripts (Chen *et al.*, 1994). Wild-type *FOS* protein is also rapidly depleted by ubiquitin-independent proteasomal degradation (Ferrara *et al.*, 2003). The C-terminal truncations seen in osteoblastoma and epithelioid haemangioma have recently been shown to protect *FOS* from this degradation (van Ijzendoorn *et al.*, 2017). It is therefore expected that this rearrangement will have a quantitative impact on *FOS* level, explaining the unprecedented nuclear immunoreactivity observed in osteoblastoma, which required downward titration of the primary antibody used in immunohistochemistry (**Appendix 7.1.6**). It was not possible to establish allele-specific differences in expression resulting from rearrangement because of the lack of heterozygous SNPs in the relevant gene regions. As only rearrangements could abolish both transcriptional and translational levels of regulation, these findings may explain the absence of nonsense mutations mediating *FOS* truncation. This could be validated by introducing wild-type and mutated *FOS* into an expression system. Protein and transcript levels could then be assayed in the presence or absence of proteasome and transcriptional inhibitors. Qualitative differences in *FOS* activity resulting from mutation cannot be excluded. The observed breakpoints disrupt components of the C-terminal transactivation

domain, but this is not required for *in vitro* transformation by the murine retroviral orthologue, *v-fos* (Jooss *et al.*, 1994, Sutherland *et al.*, 1992). It is intriguing that rearranged *FOS* moulds a transcript that so closely resembles *v-fos* (**Figure 19**). It is more than fifty years since the FBJ murine sarcoma virus was discovered (Finkel *et al.*, 1966). It is this retrovirus, able to induce an osteosarcoma-like malignancy in CF-1 laboratory mice, from which *v-fos*, the murine orthologue *c-fos* and eventually the homologue canonical proto-oncogene *FOS* were identified (Miller *et al.*, 1984, van Straaten *et al.*, 1983, Van Beveren *et al.*, 1983). Despite this, until now, mutation in *FOS* in human bone-forming tumours had not been observed.

3.4.3. Mutational recurrence suggests common disease entities and evolutionary origins

A simple highly recurrent mutational pattern provides insights into tumour evolution and origins. Mutations common across a cohort are more likely to have arisen early in evolution. Direct evidence of this requires either longitudinal sampling in time - not possible in osteoblastoma lacking either a precursor lesions or clear evidence of malignant transformation – or copy number gains allowing computational evolutionary analysis (Gerstung *et al.*, 2018, Jolly and Van Loo, 2018, Fittall and Van Loo, 2019). Osteoid osteoma and osteoblastoma both possess the same mutational pattern, which suggests, as previously suspected, that they are manifestations of the same disease.

Epithelioid haemangioma (EH) also commonly possesses a similar *FOS* rearrangement. This benign vascular tumour shares some histological features with osteoblastoma; it can also be found in the bone, whilst osteoblastoma is also marked by significant vascularity. *FOSB* rearrangement is found in a minority of osteoblastoma and epithelioid haemangioma but is almost ubiquitous in pseudomyogenic haemangioendothelioma (PHE) (Sugita *et al.*, 2016, Hung *et al.*, 2017). It is unknown whether rearrangements in *FOSB* frequently occur by chromoplexy. To date no other *FOSB* rearranged samples have been subjected to DNA sequencing.

One can conjecture that benign bone and vascular tumours might share evolutionary origins, possibly segregated by context or differentiation pathway. The mutated paralogue might better define the observed phenotype; *FOSB* alteration resembling PHE, *FOS* alteration resembling osteoblastoma and EH dividing equally between each group. This could be explored by introducing these mutations into pluripotent mesenchymal cell types. Furthermore, it is notable that this mutation seems to preclude malignant transformation, suggesting an evolutionary dead-end. Combining *FOS* or *FOSB* alteration with canonical driver mutations, such as bi-allelic *TP53* loss, might give insights into this, particularly if introduced sequentially.

BLANK PAGE

Chapter 4. H3.3-mutant bone tumours – the pattern of progression

4.1. Introduction

Giant cell tumours of bone (GCT), like osteoblastoma, are rare primary benign bone tumours. A simple recurrent mutation has already been found which defines the GCT genome, occurring in one of the genes of the histone family, *H3F3A*. Predominantly their histological benign appearances are reflected in benign clinical behaviour, albeit with a tendency for local destruction and recurrence. Two rare phenomena exist which remain poorly understood: the occurrence of lung metastasis in the absence of any distinct or malignant histological features, and malignant tumours possessing either the characteristic *H3F3A* mutation or co-existing conventional GCT. The goal of this chapter is to explore the genomics of these phenomena with the aim of identifying mutational patterns that may underlie them.

4.1.1. Conventional giant cell tumours of bone

GCT comprise 20% of primary benign bone tumours, an equivalent proportion to osteoblastoma. It may present with progressive pain, local destruction, swelling and restricted joint movement. The majority of cases affect the meta-epiphyseal region of the long bones, particularly the distal femur and proximal tibia, though rare cases occur throughout the skeleton, including the small bones of the hands and feet (Campanacci *et al.*, 1987, Czerniak and Dorfman, 2016). The most comprehensive study of its epidemiology comes from a national pathology database in the Netherlands (Verschoor *et al.*, 2018). This revealed an overall incidence of 1.7 diagnoses per million per year with a female:male ratio of 1.38:1. The age of onset is bi-modal with the largest peak in incidence between the ages of 20-39 but with a second smaller peak aged 50-59.

Investigation with plain radiographs normally reveal a single well circumscribed lytic lesion without evidence of sclerosis or bone formation (Chakarun *et al.*, 2013). Magnetic Resonance Imaging (MRI) can be a useful adjunct for local

staging, characteristically revealing haemosiderin deposition, related to chronic haemorrhage (Aoki *et al.*, 1996).

Pathologically, GCTs are part of a group of tumours defined by the presence of large multinucleated cells, called giant cells, which are considered to be osteoclasts and referred to as osteoclast-like (Orosz and Athanasou, 2017, van der Heijden *et al.*, 2017). These giant cells are typically larger than normal osteoclasts, potentially containing more than 20 nuclei, and express TRAP, CD51 and CD68 but not the other typical macrophage markers, CD11, CD18, CD14 or HLA-DR (Burstone, 1959). The identity of the tumour cells themselves was only definitively resolved with the identification of the disease defining *H3F3A* mutation. The mutation is possessed by non-osteoclast stromal cells, lacking macrophage markers, and they are often still referred to as stromal. These mononuclear stromal cells can either be spindle shaped or rounded, and may be moderately mitotically active, with up to 20 mitoses per 10 high power fields, but generally without any mitotic atypia (Al-Ibraheemi *et al.*, 2016). The histological appearance typically shows osteoclast-like cells uniformly interspersed amongst the tumour stromal cells, though there is considerable variation within and between tumours. Neoplastic stromal overgrowth, in which there are limited numbers of osteoclast-like cells, can also be seen. This may be associated with the presence of foamy macrophages, haemorrhage, haemosiderin deposition, and reactive bone formation, suggestive that these features occur post infarction. Physiological recruitment of osteoclasts is related to the expression of Receptor Activator of Nuclear Factor kappa B Ligand (RANKL) relative to its decoy ligand, osteoprotegerin and M-CSF, which are expressed by cells of the osteoblastic lineage (Atkins *et al.*, 2001, Atkins *et al.*, 2000, Roux *et al.*, 2002, Huang *et al.*, 2000). Giant cells, as with normal osteoclasts, express the receptor for this ligand, RANK. Therapeutically interfering with this signalling, using the RANKL inhibitor, denosumab, has been trialled (Thomas *et al.*, 2010, Chawla *et al.*, 2013, Gaston *et al.*, 2016, Rutkowski *et al.*, 2015). Denosumab has been demonstrated to cause a dramatic reduction in giant cell infiltration, associated with abundant osteoid deposition. This results in a reduction in tumour size and the scale of surgical resection required (Luengo-Alonso *et al.*, 2019). Denosumab is therefore

currently used for unresectable tumours, those requiring highly morbid surgery, or can be employed for the treatment of patients with metastatic GCT. The impact on prognosis or the development of metastases is still unclear.

Surgical resection of primary tumours, when possible, either by curettage or en-bloc resection is considered optimal management (Orosz and Athanasou, 2017, van der Heijden *et al.*, 2017). Of note, whilst primary giant cell tumours do occur outside of the bone, these are considered to be a separate disease, with distinct genetic causes (Lee *et al.*, 2017a).

4.1.2. The genomics of GCT

GCTs possess few somatic alterations of any type. Whole genome and exome sequencing have been conducted in 5 and 8 tumours respectively (Presneau *et al.*, 2015, Ogura *et al.*, 2017). The only recurrent finding is a missense mutation at the Glycine residue of the histone gene, *H3F3A*, resulting in a p.Gly34Trp (G34W) mutation. Rare variations of a mutation to a Leucine, Methionine or Valine residue at the same position have been observed, though these were mainly in the atypical sites of the small bones of the hands and feet or the axial skeleton (Amary *et al.*, 2017). G34 mutations were initially discovered by targeted sequencing (Behjati *et al.*, 2013) but have been subsequently validated and found to be present in 94-96% of GCT cases (Presneau *et al.*, 2015, Ogura *et al.*, 2017, Fellenberg *et al.*, 2019). Immunohistochemistry has good sensitivity for the mutated histone protein and is now a part of the diagnostic workup for these tumours (Amary *et al.*, 2017, Luke *et al.*, 2017, Yamamoto *et al.*, 2018). The exome and targeted sequencing of GCTs carried out by Ogura *et al.* suggested possible isolated mutations in additional epigenetic regulators (Ogura *et al.*, 2017). The significance of these mutations is unclear for a number of reasons. No mutation was recurrent and they were only present in a small minority of tumours. Furthermore, the sample collection is reported as containing only conventional GCTs, however the rate of metastasis (13%) and death (6.5%) is higher than would be expected for conventional GCTs, which could be explained by a misclassification. Indeed, in the whole genome sequenced cohort, which were independently pathologically reviewed, no additional mutations were

identified (Presneau *et al.*, 2015). In the study by Ogura and colleagues the association between these additional mutations and clinical outcome was not described therefore the possibility remains that they were associated with a malignant histological or clinical phenotype.

A single study reported Isocitrate Dehydrogenase (*IDH*) 2 mutations in 16/20 samples by Sanger sequencing, however this finding has not been reproduced and all samples were wild-type for *H3F3A* calling into question the experimental methodology (Kato Kaneko *et al.*, 2014).

4.1.3. Histone Coding Variants

Histone proteins form the core of the repeating pattern that packs DNA into chromatin (Olins and Olins, 2003). The structure of chromatin is a fundamental mechanism for the control of processes that access DNA, including repair, DNA synthesis and transcription (Klemm *et al.*, 2019). Histone mutations might therefore be expected to lead to an alteration of the structure of chromatin thereby changing gene expression. Through altered gene expression, epigenetic modifications can control the activity, type and fate of the cell. Somatic mutations in histone genes are increasingly recognised as widespread in tumorigenesis (Nacev *et al.*, 2019).

Histone coding variants have now been reported in a number of histone genes and are best recognised in the two members of the H3.3 subfamily, *H3F3A* and *H3F3B* (**Table 6**). These two genes, found on chromosome 1 and 17 respectively, encode identical histone core proteins. The less commonly mutated H3.1 gene, *HIST1H3B* only differs from H3.3 by four amino acid substitutions. Three hotspots for mutation are noted; the Lysine residues at K27 and K36 and the Glycine residue at G34. The disease specific occurrence of mutations is striking: K27M and G34R/V in paediatric central nervous system tumours (Schwartzentruber *et al.*, 2012, Sturm *et al.*, 2012, Wu *et al.*, 2012, Gessi *et al.*, 2013), G34W in GCT and K36M in chondroblastoma (Behjati *et al.*, 2013).

Two of the sites commonly mutated, K27 and K36, are important sites for post-translational modification of histones. This suggests that the mechanism of pathogenic mutation is by manipulation of these post-translational modifications.

As epigenetic states are critical in defining cell type and status it is not surprising that mutations are highly disease-specific. Each mutation may only be effective at inducing tumorigenesis in a particular cell state at a particular phase of differentiation. Mutations must also have a dominant effect as they occur without the loss of the wild-type allele or the plethora of highly homologous histone genes. This dominant effect is why they are commonly referred to as ‘oncohistones’ (Nacev *et al.*, 2019).

Gene	Variant	Number	Disease
<i>H3F3A</i>	K27M	417	Paediatric Glioblastoma Multiforme (pGBM)
		70	Anaplastic Glioma
		65	Diffuse Intrinsic Pontine Glioma (DIPG)
		36	Low Grade Glioma
		28	Other Central Nervous System tumours (CNS)
		8	Primitive Neurectodermal Tumour (PNET)
		2	Other
	K27E/N/R	6	Lymphoproliferative/Other
	G34W	115	GCT
		5	Osteosarcoma
	G34R	12	Other (incl. Pheochromocytoma)
		82	Paediatric Glioblastoma Multiforme (pGBM)
		2	Osteosarcoma
	G34V	7	Primitive Neurectodermal Tumour (PNET)
		8	Other CNS
		10	Paediatric Glioblastoma Multiforme
		4	GCT
1		Other CNS	
G34E/L/X	8	GCT and other CNS	
<i>H3F3B</i>	K36M	74	Chondroblastoma
		1	Chondrosarcoma
<i>HIST1H2AA</i>	R11C	28	Pancreas Adenocarcinoma
		1	Other Carcinoma
<i>HIST1H3B</i>	K27M	40	Paediatric Glioblastoma Multiforme (pGBM)
		22	Diffuse Intrinsic Pontine Glioma (DIPG)
		9	Other CNS
		4	Lymphoproliferative

Table 6 Cosmic histone mutations. Cosmic mutations (v89) in histone codons with more than 30 reported mutations.

K27M Mechanism

The direct mechanism of histone mutation has been best explored in paediatric CNS tumours. K27M mutation causes widespread loss of K27 trimethylation (H3K27me₃) and dimethylation (H3K27me₂) even in unmutated histones (Bender *et al.*, 2013). The methyltransferase responsible for these methylation changes, the Polycomb Repression Complex 2 (PRC2), is still able to bind histones in dense CpG rich islands (CGIs). The presence of H3K27M mutated histones, however, prevents the migration of this complex along the genome from CGIs (Harutyunyan *et al.*, 2019, Lewis *et al.*, 2013). This prevents the more widespread distribution of these methylation marks, which are normally repressive of transcription, and indicative of cellular differentiation. There is a reciprocal increase of histone acetylation, which in combination may increase the transcription of normally untranscribed repeat elements (Krug *et al.*, 2019). This effect may be open to therapeutic enhancement by increasing DNA methylation and inhibiting histone deacetylases (Krug *et al.*, 2019).

G34R/V Mechanism

G34R/V mutations have no impact on H3K27me₃ but instead reduces H3K36 methylation (Lewis *et al.*, 2013). In part this may be via inhibition of the methyltransferase, SETD2, reducing tri-methylation, but this is thought to only occur on the mutated histone itself. More recently introducing G34R mutations into embryonic cell line models, suggests the more global impact of these mutations may be mediated by inhibition of the KDM4 family of demethylases, which normally demethylate both K9 and K36 residues. Contrary to K27 methylation, K9/36 demethylation is the repressive mark, while trimethylation is a marker of active chromatin (Voon *et al.*, 2018). G34R/V mutations in paediatric CNS tumours, frequently co-exist with loss of function mutations of chromatin remodelling chaperones, ATRX and DAXX (Schwartzentruber *et al.*, 2012). *In vitro* exploration of this combination of mutations has not been conducted.

K36M Mechanism

K36M mutation in chondroblastoma causes the sequestration of the methyltransferases NSD1/2 and SETD2, causing a reduction in H3K36me₃ and

concomitant increase in H3K27me₃, most notably in intergenic regions (Lu *et al.*, 2016). This intergenic shift in epigenetic markers is postulated to activate genes responsible for the differentiation blockade by 'diluting' Polycomb Repression Complex 1 (PRC1) activity across these intergenic regions. Only K36M mutated histones, and not G34W/L or K27M/R can cause the blockade of chondrocyte differentiation of mesenchymal progenitor cells (Lu *et al.*, 2016). This reiterates the cell specificity of these histone mutations.

G34W Mechanism

Similar epigenetic exploration of G34W mutation is underway, suggesting that it reduces in-cis H3K36Me₃, but contrary to G34R mutation also results in increased H3K36Me₂ (Deshmukh *et al.*, 2018). These results remain unpublished and the more global epigenetic landscape of G34W mutated tumours is unknown. To date, the only published work confirms the proliferative effect of the G34W mutation but also suggests it may result in an alteration to the RNA processing machinery, resulting in a number of alternative splice isoforms (Fellenberg *et al.*, 2019, Lim *et al.*, 2017).

4.1.4. Metastatic giant cell tumours of bone

The earliest and clearest description of the incidence of metastases from giant cell tumours of bone was given by the hugely influential American bone pathologist, Henry L. Jaffe in a lecture to the Royal College of Surgeons in London in 1953 (Jaffe, 1953). He detailed the pathological findings and treatment of 60 GCTs, including four cases with lung metastases. Jaffe hypothesised this was related to embolization of tumour that could be seen within the primary site. In all four cases the pulmonary metastasis retained the histological appearance of the primary tumour. For several decades controversy remained, with some suggesting that the phenomenon was a manifestation of undiagnosed malignant transformation (Sanerkin, 1980). Over several decades, however, isolated case reports were collected into series which established the occurrence of isolated pulmonary metastases that retain the histological appearances of a benign conventional GCT (Jewell and Bush, 1964, Rock *et al.*, 1984). The incidence

across these collated retrospective series appears to be 2-3% of GCTs (Rock *et al.*, 1984, Muheremu and Niu, 2014). As surveillance is not standard and metastases may be both asymptomatic and clinically undetectable, the true incidence may be higher. Indeed, in a series in which surveillance was standard, either by chest radiograph or CT, the reported incidence of lung metastases was three times higher (7.5%) (Rosario *et al.*, 2017). Attempts have been made to identify risk factors for metastasis, however the only consistent factor is local recurrence (Chan *et al.*, 2015, Wang *et al.*, 2017).

The standard of care for benign lung metastases is unclear. Metastasectomy is recommended for isolated lesions, and denosumab has been trialled when they are more widespread (Luo *et al.*, 2018). Pulmonary lesions generally are indolent and slowly growing, however they do have an associated mortality. Considering the young age of onset, disease specific and overall survival are equivalent and have been estimated at 88% at a median of 6.9 years from a pooled case series (Itkin *et al.*, 2018). It is likely therefore that untreated pulmonary metastases are life limiting, however longer term follow up data are not available. Biological determinants of metastatic behaviour are lacking. Beyond evidence of the persistence of G34W mutation (Yamamoto *et al.*, 2018), nothing is known about the genomics of metastatic GCT.

4.1.5. Malignant H3.3 bone tumours

The concept of malignancy in GCT is controversial. Once again Henry Jaffe provides one of the earliest descriptions of a conventional GCT that changed both in clinical behaviour, with the onset of widespread fatal metastatic disease and the presence of distinct histological features that resembled sarcoma (Jaffe, 1953). Subsequently malignant GCT became defined as sarcoma in the presence (primary) or with a history of (secondary) GCT at the same site (Hutter *et al.*, 1962, Dahlin *et al.*, 1970). The most comprehensive case series comes from the Rizzoli institute in Bologna, Italy, reporting 5 primary and 12 secondary malignant GCTs, representing 1.8% of all GCTs seen in the same time period (Bertoni *et al.*, 2003). The malignant component of the majority of cases exhibited an osteosarcoma-type appearance and the outcome was poor with 10/17 dying

of disease despite multi-modal therapy. The secondary malignant GCTs were split evenly between those that had received primary surgery or primary radiotherapy for the initial GCT. No difference in radiographic or histologic finding was noted between these groups. An analysis of SEER cancer registry data in the USA, suggested that the overall incidence of malignant GCT was 1.6 per 10 million per year (Beebe-Dimmer *et al.*, 2009). This is higher than would be expected from the overall incidence of GCT and this figure is likely plagued with the inaccuracies of pathological diagnosis in registry data. This study also reported a 5-year survival of 85%, far higher than observed in the Rizzoli cohort, raising further questions about the reliability of the data.

The biological factors underlying malignant GCT are not well understood. Isolated reports have suggested potential associations with *TP53* alteration, and *CCND1* and *MET* amplification but the methodologies used included only single gene analysis, array-CGH and FISH (Saada *et al.*, 2011, Saito *et al.*, 2011a, Saito *et al.*, 2011b, Okubo *et al.*, 2013). Limited studies have assessed the association of G34W mutation with malignant progression. Case reports have described either the retention or loss of G34W mutation in malignant progression (Tsukamoto *et al.*, 2017, Tsukamoto *et al.*, 2018, Yoshida *et al.*, 2019). Conversely, four series have reported the occurrence of *H3F3A* mutation in sarcoma not initially reported to have a GCT component (Amary *et al.*, 2017, Koelsche *et al.*, 2017, Righi *et al.*, 2017, Yoshida *et al.*, 2019). The majority of these cases were diagnosed as osteosarcoma, though often rich in osteoclasts (**Table 6**). Methylation array analysis in one series suggested closer similarity of malignant H3.3 mutant osteosarcoma to GCT than H3.3 wild-type osteosarcoma (Koelsche *et al.*, 2017). These observations suggest the theory that malignant H3.3 mutant tumours are true malignant GCTs that lack the histological evidence of previous or existing conventional GCT.

4.2. Aim

To explore the genomic landscape of H3.3 mutant bone tumours

4.2.1. Objectives

- Explore the common origin of H3.3 mutant bone tumours
- To identify genomic patterns that may elucidate the process of metastatic spread in conventional GCT
- To identify mutational events that distinguish benign from malignant H3.3 mutant tumours

4.3. Results

4.3.1. The genomic landscape of H3.3 mutant bone tumours

H3.3 mutated bone tumours were identified for this study by immunohistochemical screening of tumours for the mutated protein (Amary *et al.*, 2017, Presneau *et al.*, 2015). This cohort spanned the spectrum of patients affected by benign (n=9) and malignant (n=8) H3.3 tumours but was in keeping with the typical demographics for GCT: a median age of diagnosis of 25 (Range 14-53), a 9:8 female-to-male ratio, and most cases occurring in the tibia and femur (**Table 7**). Many diagnoses were challenging and were reviewed by multiple specialist bone tumour pathologists nationally and internationally. Osteosarcoma was the most common diagnosis of the malignant tumours (4/8) followed by primary malignant giant cell tumours of bone (2/8). All but two malignant tumours had an osteoclast-like component.

Sample	Age	Sex	Site	Diagnosis	Multiple external reviews	Management	Outcome	Follow Up (months)
PD3795d	21	M	Metatarsal, 1st, left	GCT	Yes	Resection only	NED	226
PD30981a	15	F	Tibia, proximal, right	GCT	Yes	Resection only	NED	154
PD21296a	21	F	Tibia, proximal, right	GCT	No	Resection only	NED	97
PD21295a	29	M	Fibula, proximal, right	GCT	Yes	Resection only	NED	91
PD21294a	49	M	Femur, distal, left	GCT	No	Resection only	NED	96
PD21292a	25	M	Tibia, proximal, right	GCT	No	Resection only	NED	101
PD7524a	23	M	Tibia, proximal, right	GCT	Yes	Resection only	NED	106
PD30982a	20	F	Fibula, proximal, right	GCT*	Yes	Surgery, chemotherapy	NED	113
PD38329a			Radius, distal, left	BMGCT-Prim		Resection only		
PD38329d	14	F	Lung	BMGCT-Met	Yes	Surgery, Chemotherapy (no response)	SD	17
PD38329c			Lung	BMGCT-Met		Surgery, Chemotherapy (no response)		
PD13463a	40	F	Tibia, proximal, right	Adamantinoma, osteoclast-rich	No	Resection only	DOD	7
PD38328a	38	F	Femur, distal, left	Malignant GCT	No	Surgery, Chemotherapy (no response)	NED	24
PD3788d	29	F	Femur, proximal, right	Malignant GCT	Yes	Surgery, Chemotherapy	DOD	50
PD37332a	17	F	Femur, distal	Osteosarcoma (low grade fibroblastic) with GCT component	No	Resection only	NED	49
PD30985a	53	M	Tibia, proximal, left	Osteosarcoma, chondroblastic	No	Neoadjuvant chemotherapy (poor response), surgery, multiple lines of chemotherapy at relapse (no response)	PD	64
PD30984a	25	M	Tibia, proximal, left	Osteosarcoma, GCT-like	No	Neoadjuvant chemotherapy (poor response), surgery	NED	79
PD4915d	47	M	Femur, distal, left	Osteosarcoma, osteoclast-rich	No	Surgery, chemotherapy (no response)	DOD	17
PD4922e	37	F	Femur, shaft, left	Pleomorphic Sarcoma	No	Resection only	DOD	11

Table 7 Clinical characteristics of whole genome sequenced H3.3 mutant bone tumours. Abbreviations: GCT (Giant Cell Tumour of bone), BMGCT (Benign Metastasising Giant Cell Tumour of bone), NED (No evidence of disease), SD (Stable Disease), DOD (Died of Disease), PD (Progressive Disease). *This case was initially diagnosed as an osteoclast-rich osteosarcoma but this was amended to GCT on retrospective review

All tumours were resected and some of the patients received chemotherapy before or after surgery but generally with a poor or no response. Four of the eight patients with malignant disease have, to date, died from disease at a median 14.5 months from diagnosis (Range 7-50), while a further patient has, at most recent follow-up, progressive widespread metastatic disease despite third line chemotherapy. This series also includes an individual with metastatic disease from a conventional benign GCT (BMGCT). At diagnosis of lung metastases, 8 months after initial GCT diagnosis, chemotherapy was instigated for suspected malignant progression. As the pulmonary lesions did not respond to chemotherapy, resection was performed and subsequent review suggested all lesions are consistent with conventional GCT and possessed an *H3F3A*:p.G34W mutation.

Genomic DNA was extracted from freshly frozen tumour sections and matched germline sequences were extracted from blood. These were subjected to whole genome sequencing at a median coverage of 39.1x (Range 35-61) and 37.3x (24-68) for tumours and germline sequences respectively. Five GCTs had been sequenced for a previous study but all data was reanalysed consistently in the entire cohort (**Methods 2.7**) (Presneau *et al.*, 2015). Catalogues of somatic variants were compiled for all tumours and precision of all variant calls was found to be >95% by manual review.

In keeping with the previous report (Presneau *et al.*, 2015), conventional GCT possessed a comparable number of somatic variants to osteoblastoma (**Figure 12** and **Figure 22**) with few substitutions (median 670, range 849-1090), indels (median 44.5, range 19-66) or structural variants (median 8, range 4-15). In contrast, all malignant tumours possessed an increased burden of somatic variants, though broadly these could be divided into two groups: 4/8 tumours with only a modest increase in mutations (medians: 1815 substitutions, 86 indels, 21 structural variants) and 4/8 with a mutational burden similar to osteosarcoma (medians: 4177 substitutions, 205 indels, 108 structural variants).

H3.3 mutant bone tumours

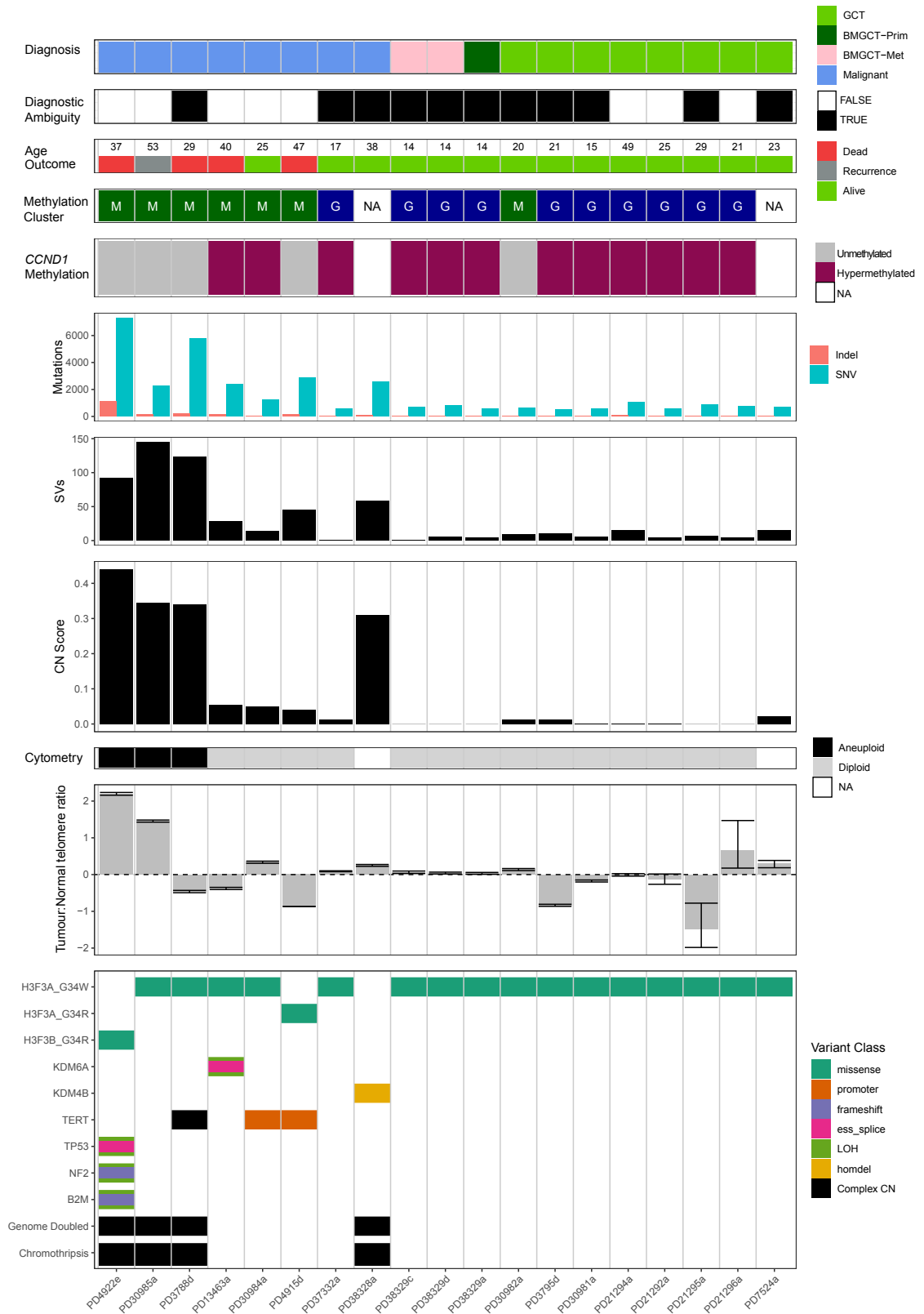


Figure 22 Landscape of H3.3 mutant tumours. From top to bottom: Clinical diagnoses, ambiguity in diagnosis requiring multiple reviews, age at diagnosis and clinical outcome, unsupervised methylation cluster assignment, *CCND1* promoter methylation status (hypermethylation is defined as a mean *CCND1* promoter methylation beta value >0.2), raw small somatic variant (SNV/indel) counts, structural variant counts, Copy Number (CN) score (**Methods 2.15** and **Results 4.3.2**), Tumour:Normal telomere length ratio with error bars reflecting the 5th and 95th centiles from 10,000 bootstrapped estimates sampling the data from the underlying sequencing read groups, tileplot of curated drivers and significant genomic events (genome doubling and chromothripsis).

All variants were scrutinised for putative cancer drivers using a similar strategy as described before (**Methods 2.9** and **Results 3.3.1**). In brief, all coding variants were annotated for their functional impact using the algorithm, VAGrENT (<http://cancerit.github.io/VAGrENT/>). Tumour suppressors were required to have a deleterious variant in a single remaining gene copy or be focally (<1MB) homozygously deleted. Oncogenes were required to have known hotspot mutations or focal amplification, defined as four more copies than the average major allele copy number. Structural variants were all screened against known fusions as well being considered for their disruptive deleterious impact or undetected copy number change. The reads supporting candidate driver variants were manually reviewed. The published literature was reviewed to confirm plausible drivers (**Figure 22**).

All samples possessed known histone variants, most commonly *H3F3A*:p.G34W in sixteen cases and one each of *H3F3A*:p.G34R and *H3F3B*:p.G34R. These G34R variants were only seen in malignant tumours but neither were accompanied with *ATRX/DAXX* mutations normally seen in paediatric GBM harbouring *H3F3A*:p.G34R/V mutation. The only tumour lacking a histone variant was the malignant tumour, PD38328a, which had lost a copy of the *H3F3A* locus, discussed further below (**Results 4.3.3**).

No conventional GCTs, including the metastatic case, possessed an additional driver event. In contrast, all malignant tumours contained at least one additional driver event. The only exception was PD37332a which was a biphasic tumour, possessing both a region of both conventional GCT and a region more closely resembling low-grade osteosarcoma but without *MDM2* amplification as assessed by FISH. On review the sequenced portion of the tumour was purely

conventional GCT, suggesting that the histological appearances reflect the underlying mutational landscape.

Biallelic losses of histone lysine demethylases were seen in two malignant tumours: *KDM4B* by homozygous deletion (**Figure 23**) and *KDM5A* by an essential splice site substitution in the context of loss of heterozygosity of the X chromosome in a female patient.

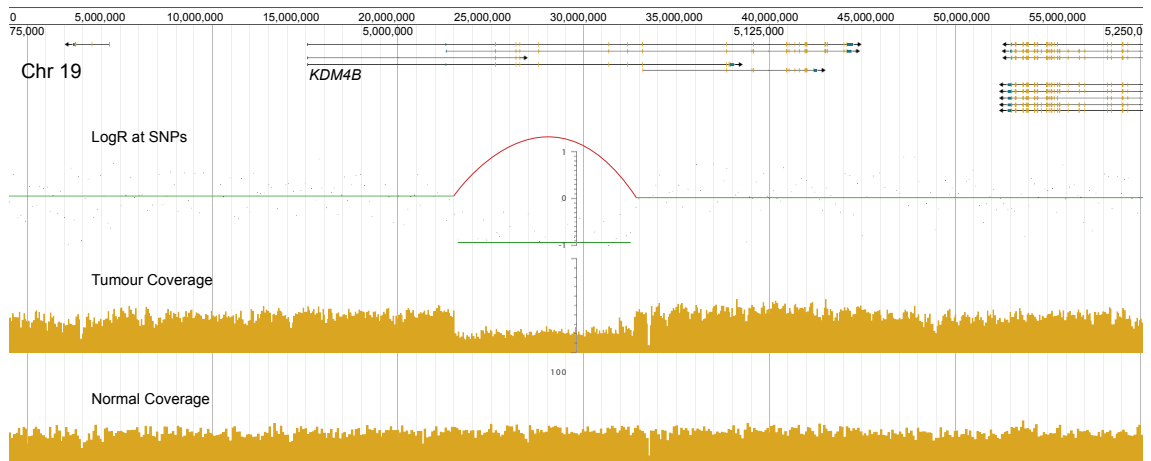


Figure 23 PD38328a *KDM4B* homozygous deletion. The different transcripts for *KDM4B* are shown at the top. LogR represents logged normalised ratios of tumour and normal coverage at SNP positions. The red link is an annotation of a deletion that was orthogonally detected by the structural variant caller BRASS (**Methods 2.7**). The loss of heterozygosity of 19p is not visible at this scale, therefore B-allele frequencies are not shown.

PD4922e, sampled from a pleomorphic sarcoma had biallelic mutations in *TP53*, *NF2* and *B2M*. This sample had the greatest mutational burden and significant evidence of genomic instability (**Figure 24**).

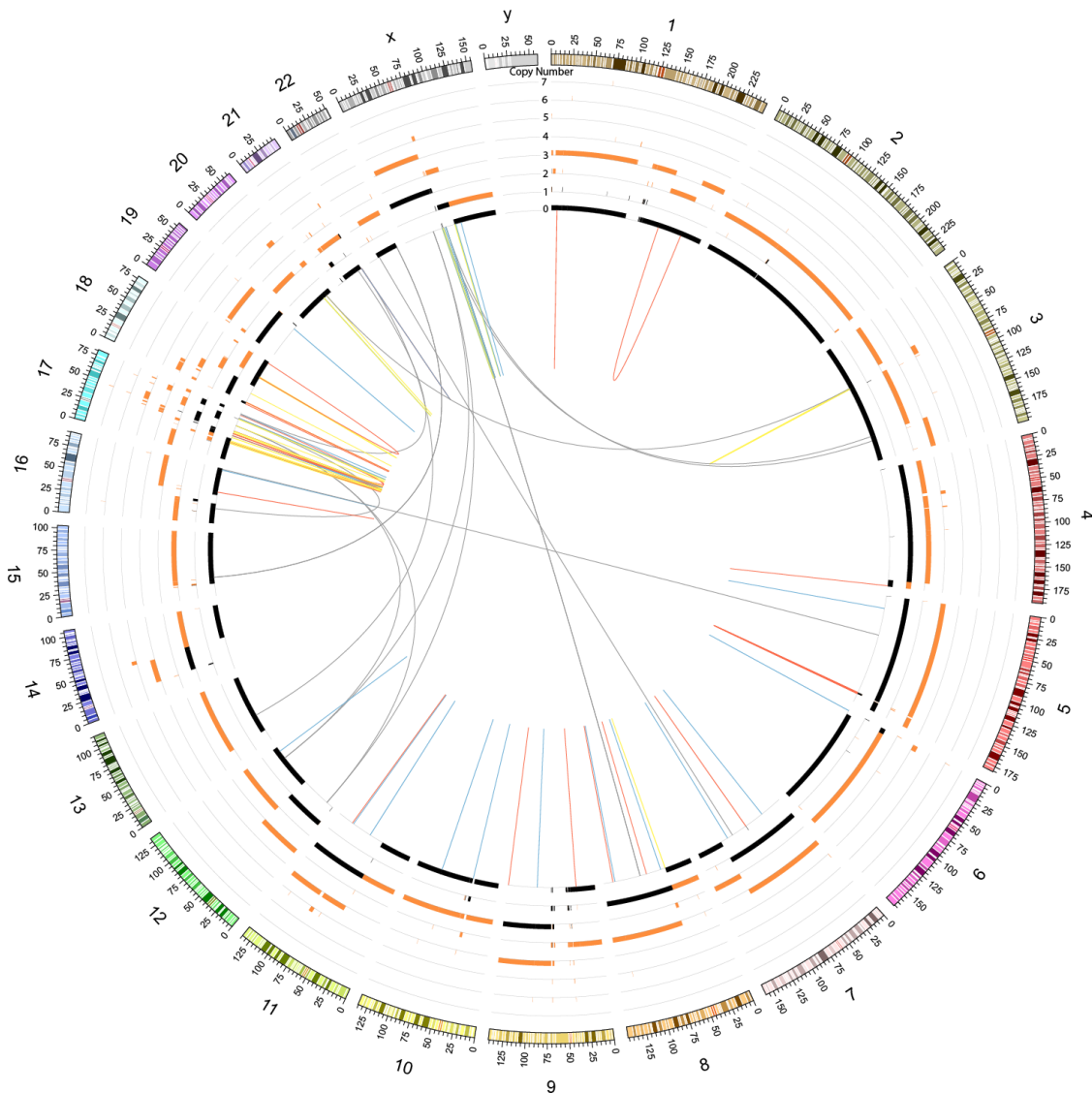
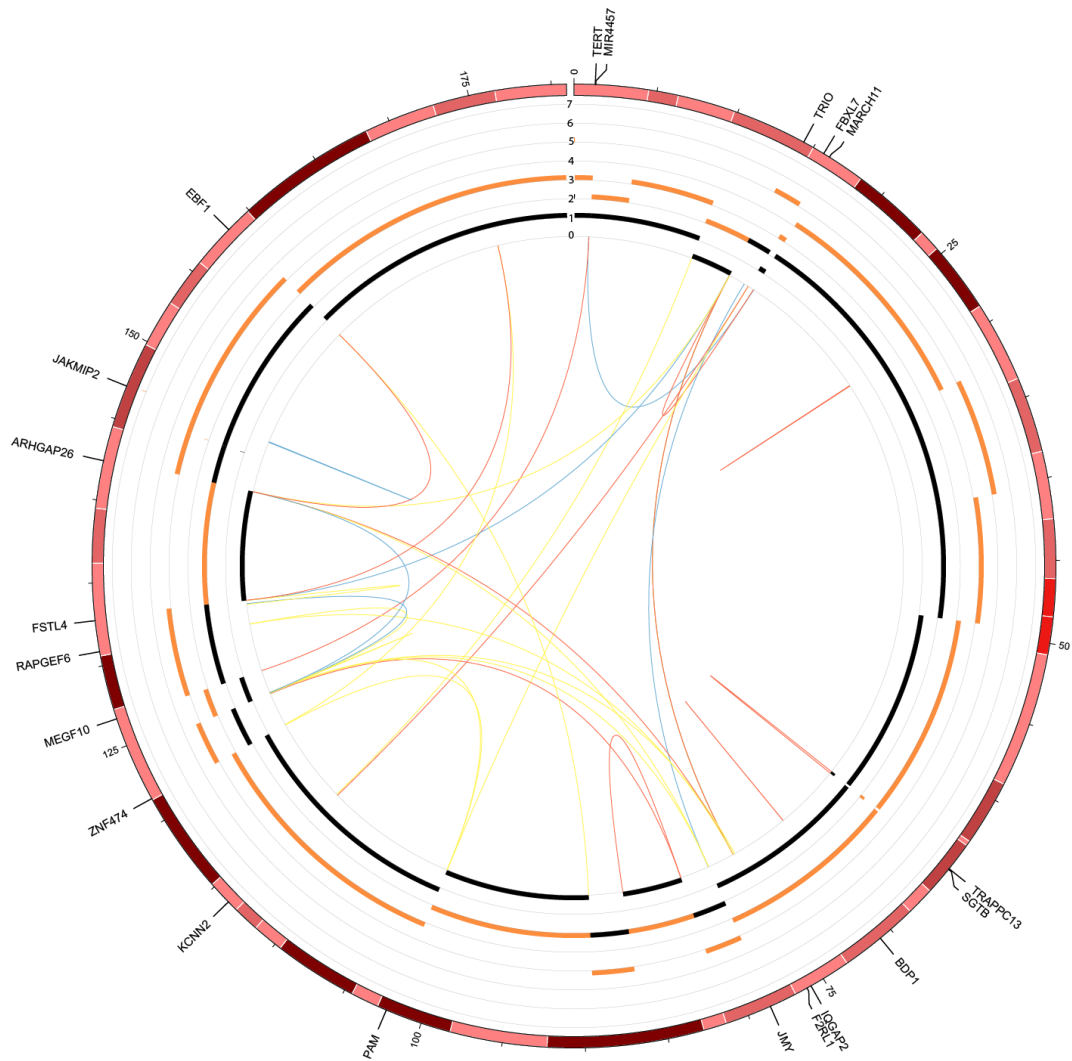


Figure 24 PD4922e Circos plot of genome wide copy number and structural variants. Copy number is shown with the orange (total copy number) and black (minor allele) segments counted from a minimum of zero on the innermost ring. Structural variants are shown as coloured internal links (blue – tandem duplications, red – deletions, yellow – inversions, grey – translocations)

The only recurrent mutation, apart from histone variants, was a canonical promoter variant, *TERT*:n.-124G>A (relative to ATG start site) which is the gene encoding telomerase (*Vinagre et al., 2013*). This mutation is suspected to increase promoter binding, thereby telomerase protein levels and activity (*Horn et al., 2013*). A further sample, PD3788d, had a complex rearrangement event, resembling chromothripsis, encompassing *TERT*. The exact functional consequence can only be surmised in the absence of expression data, however

a structural variant 15kb upstream of *TERT* resulted in the juxtaposition of the gene *MEGF10* with the TERT promoter (**Figure 25**). *MEGF10* is reported to be under the control of a superenhancer in the dbSUPER database, though data are only available for brain tissues (Khan and Zhang, 2016).



5

Figure 25 Circos plot of chromothripsis around TERT in PD3788d. Copy number and structural variants in PD3788d in the same format as Figure 24 restricted to chromosome 5. Genes lying within 10kb of each breakpoint are labelled on the outside. The random configuration of rearrangements and deletions, typical for chromothripsis is seen. Of note, this sample has undergone whole genome duplication.

To identify other methods of achieving replicative immortality, telomere lengths were calculated for each sample. Telomere lengths can be estimated from the number of sequencing reads containing the telomeric hexamer ‘TTAGGG’, using

the algorithm TelSeq (Ding *et al.*, 2014). This revealed two malignant samples, PD4922e and PD30985a, with markedly elongated telomeres. This degree of telomere elongation is consistent with so-called Alternative Lengthening of Telomeres (ALT) which is usually mutually exclusive with *TERT* alteration (Barthel *et al.*, 2017, Bryan *et al.*, 1997, Cesare and Reddel, 2010). A pan-cancer analysis of genome sequencing clustered samples by the features of their telomeric sequence and identified two distinct patterns of ALT (PCAWG, 2020). The first pattern was highly enriched for sarcomas, and similarly to PD4922e and PD30985a also involved higher numbers of genomic breakpoints. This pattern of ALT was commonly associated with biallelic mutations in *RB1* or *ATRX*, however a substantial proportion lacked either, or had only shallow *RB1* deletions. Both PD4922e and PD30985a have lost heterozygosity of chromosome 13, containing *RB1*, though further *RB1*, *ATRX* or *DAXX* (associated with the second pattern of ALT) mutations were not seen. The alternative pattern of ALT was also distinguished by a higher proportion of variant motifs ('TGAGGG' and 'TTCGGG'). Neither of these motifs were increased in these malignant samples with ALT. This suggests that these two malignant samples may have a pattern of ALT consistent with other sarcoma types.

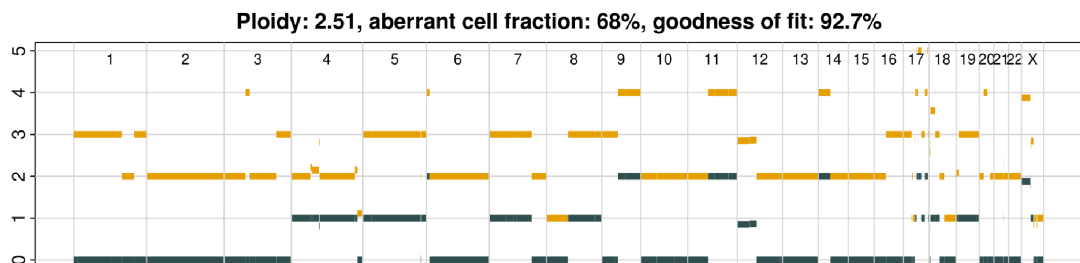
In summary all malignant tumours that were adequately sampled were found to have at least one potential driver mutation in addition to a histone variant. In two tumours this driver was an epigenetic modulator and in the remaining five there was evidence of acquired replicative immortality. Replicative immortality was achieved through *TERT* promoter mutation, ALT or potential enhancer hijacking.

4.3.2. Late aneuploidy and genome duplication mark malignant H3.3 tumour progression

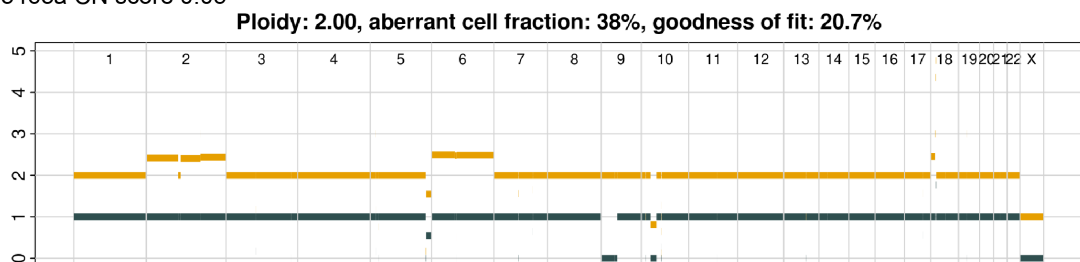
Consistent with the increasing mutational burden, malignant tumours also displayed evidence of significant copy number aberration, including chromothripsis and whole genome duplication in 4/8 cases. To quantify the degree of aberration for each sample I adapted two existing copy number scoring methods: the weighted Genome Instability Index (wGII) (Endesfelder *et al.*, 2014) and somatic copy number aberration level (SCNA level) (Davoli *et al.*, 2017). The

SCNA level is designed for un-normalised logR values for SNP positions across the genome, but allocates each SNP a score based on the degree of aberration. wGII is designed for segmented normalised integer-value copy number and is the proportion of the genome that is aberrated. wGII has no additional weighting for significant aberrations such as homozygous deletions or amplifications. Harnessing the better features for each scoring method, the Copy Number (CN) score reported here is based on a weighted proportion of the genome aberrated, but with increased weighting for homozygous deletions or amplifications (**Methods 2.15, Figure 22, and Figure 26**).

PD4922e CN score 0.44



PD13463a CN score 0.05



PD38329a CN score 0

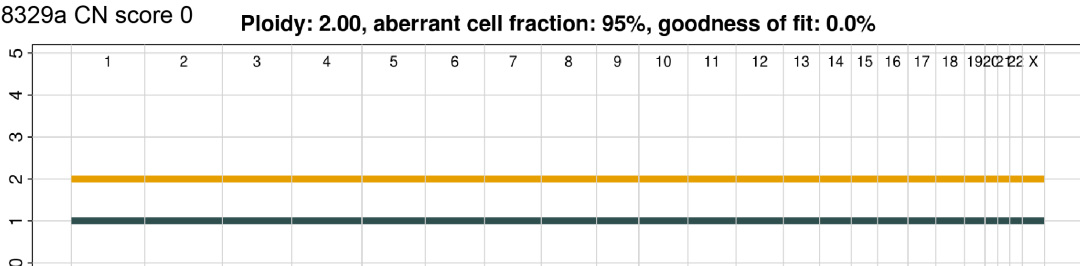


Figure 26 Example copy number profiles and scores. Selected Battenberg copy number profiles for sequenced cases representing the spectrum of copy number alteration and score. Total copy number is shown as orange segments and minor allele copy number in black.

CN scores from sequenced samples were clearly segregated between benign and malignant tumours. Once again malignant tumours separated further into two groups, one with few aberrations and another with a CN score >0.2 . To confirm

the absence of substantial copy number aberrations in conventional GCT, an additional validation cohort of tumours was assembled (**Appendix 7.2.1**). Tumour and germline (where available) DNA were extracted and subjected to SNP array analysis. Copy number profiles were produced using ASCAT (**Methods 2.4**), and CN scores derived (**Figure 27**). Only two samples demonstrated increased CN scores and these were derived from the primary (benign) and recurrent (malignant) tumour from the same patient; S00068941 and S00068945 respectively. While the primary specimen was classified as a conventional GCT, there were already histological features of atypia, which had raised concern about malignancy. Both copy number profiles are highly similar though there are some private changes to each specimen, implying some degree of intratumour heterogeneity (**Figure 28**). The CN scores of both were still substantially less than 0.2.

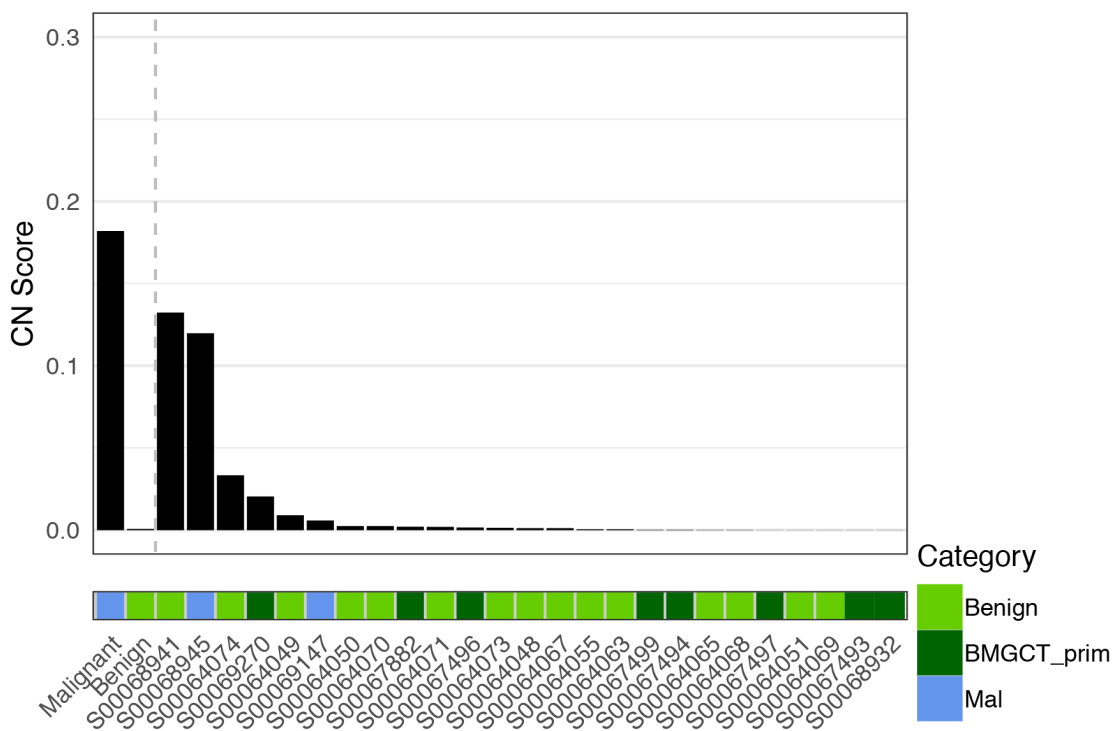
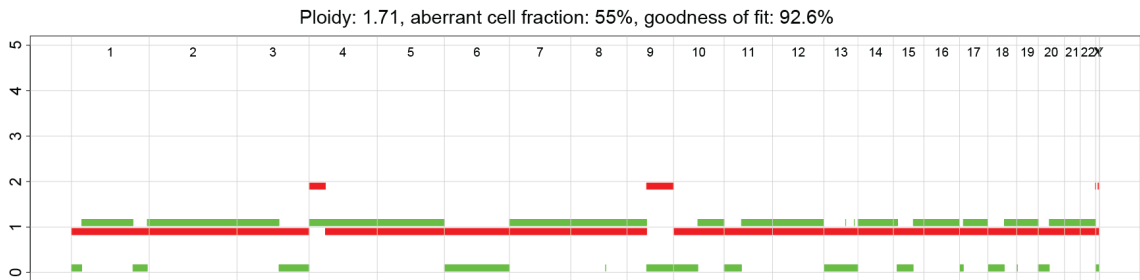


Figure 27 Validation cohort of CN scores. SNP array derived copy number scores for a validation cohort of malignant and benign tumours. Left of the dashed line are the median CN scores for malignant and benign sequenced tumours. All BMGCT specimens were from primaries only

S00068941 - primary (benign)



S00068945 - recurrence (malignant)

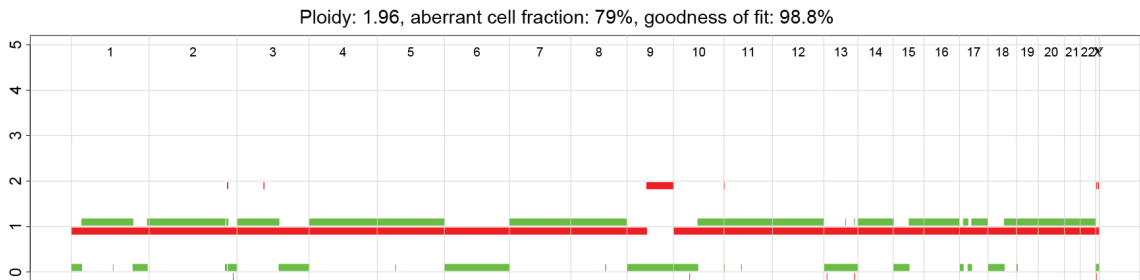


Figure 28 SNP array copy number profiles from a GCT and its recurrence. ASCAT copy number profiles. In contrast to Battenberg-based copy number profiles above, major allele is shown in red and the minor allele is shown in green.

Further validation of the spectrum of CN scores in H3.3 mutant tumours was sought using data available from methylation array profiling (**Appendix 7.2.2**, and **Results 4.3.4**). To generate copy number profiles, the intensity values for each methylation probe were normalised against known diploid samples. Resultant intensity values were binned in genomic regions and segmented. Combinations of ploidy and purity were assessed to find the optimal copy number profile with integer values.

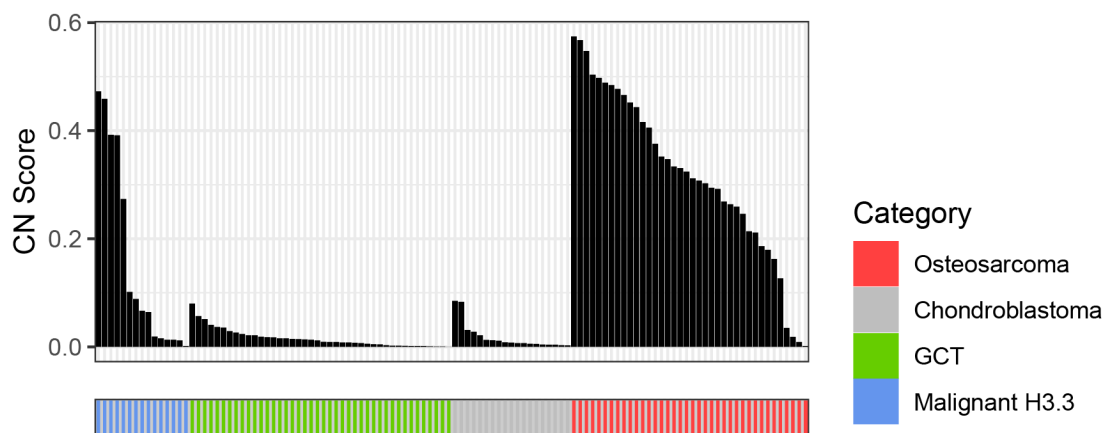


Figure 29 Methylation array based copy number scores

Methylation array-based CN scores recapitulated findings from sequencing and SNP array-based cohorts (**Figure 29**). All GCT samples and chondroblastoma, used as a benign counterpart with H3.3 mutations all had low CN scores of less than 0.1. In contrast, a proportion of malignant H3.3 tumours had CN scores that were greater than 0.2 and comparable with a malignant non-H3.3-mutated counterpart, osteosarcoma.

Copy number gains and genome duplication allow the potential to apply computational methods to calculate both the order and the timing of mutational events. This was possible in three samples. In two of these, PD4922e (**Figure 30**) and PD30985a (**Appendix 7.2.3**), the number of copies of the histone variant was compatible with the local major allelic copy number suggesting that the histone variant was acquired prior to whole genome duplication. *TP53* mutation in PD4922e was also acquired prior to duplication. In the third sample, an unbalanced copy number state (2+1) at the H3F3A locus made the ordering of the mutation ambiguous (**Appendix 7.2.3**).

Extrapolating the principle that the copy number of mutations indicates whether they occurred before or after a broader copy number gain at their locus allows the gain itself to be timed. The proportion of mutations pre-existing before the copy number gain is the relative 'lateness' of the gain. By considering only mutations that are thought to be acquired at a constant rate through life, so called 'clock-like' mutations this relative timing can be converted into real-time relative to the patient's age. This was performed by leveraging all 'clock-like' mutations found on chromosomal segments with balanced copy number gains or loss of heterozygosity (e.g. major allele 2, minor allele 2 or 0). In all three cases (**Appendix 7.2.3**), while genome duplication happened relatively late in the tumour's lifetime, this was still likely to be several years prior to diagnosis. This is demonstrated in Figure 30a using the number of mutations seen in 17p at a copy number 1. These mutations were acquired after genome duplication and appear equivalent in number to those at around 3 copies. Timing calculation, must still take into account the increased probability for a mutation to occur after a copy number gain, because the amount of genomic material on which they can arise has also increased (**Methods 2.17**).

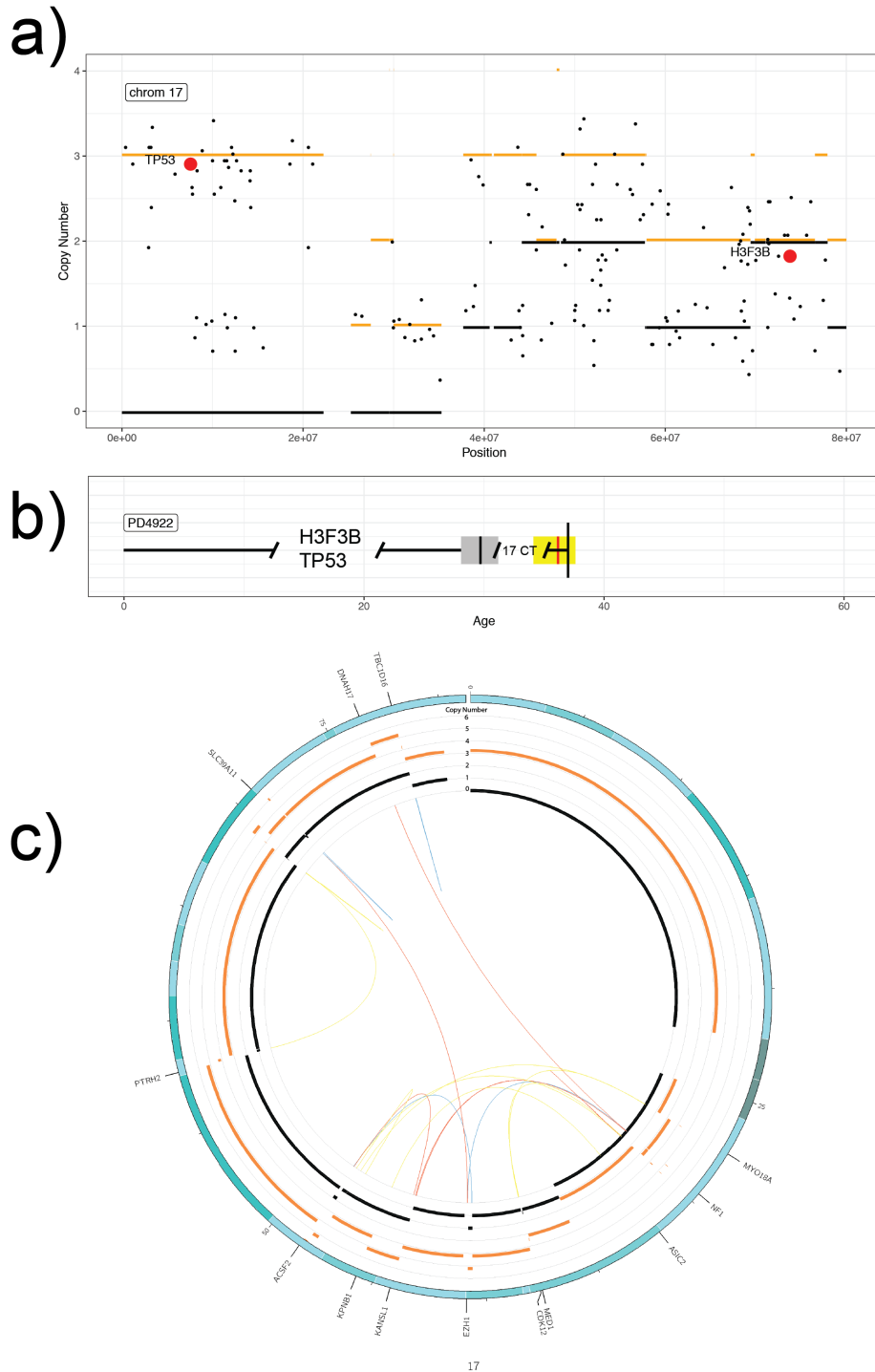


Figure 30 Timing of PD4922e. a) Copy number plot for chromosome 17 (orange segments - total copy number, black segments – minor allelic copy number). Each point is a mutation allele frequency transformed into mutation copy number. Red dots are driver mutations. b) Timeline of driver mutation acquisition and genome duplication. Genome duplication is marked in the grey bar with a confidence interval generated by 10,000 bootstrapped iterations resampling mutations. The red line is the real-time timing for additional gains (>3 major allele) with a confidence interval in yellow. c) The chromosome 17 complex rearrangement event, with many minor allele copy number states >0, therefore this must have occurred after genome duplication.

4.3.3. Case Study – H3.3 mutation potentially lost in malignant progression

One sample in the sequenced cohort, PD38328a, lacked sequencing evidence of an H3.3 mutation, while the initial diagnostic biopsy from the same specimen exhibited evidence of this mutation on immunohistochemistry. The loss of expression of the mutated gene has been reported in the literature but warrants more detailed consideration (Ogura *et al.*, 2017, Tsukamoto *et al.*, 2017).

A 38 year female, with no past medical or family history of note, presented with a 4 month history of left knee pain after a fall. Imaging in a local orthopaedic centre suggested a lytic subarticular lesion consistent with osteosarcoma, therefore she was referred to a regional sarcoma unit for investigation. A core biopsy revealed a fibroblastic proliferation but with a small residual focus of scattered osteoclasts amid scattered mononuclear stromal cells. These stromal cells had scattered nuclear immunoreactivity for a G34W antibody (**Figure 31**). For this reason, a diagnosis of conventional GCT was favoured and the patient proceeded to an en-bloc resection and distal femoral reconstruction.

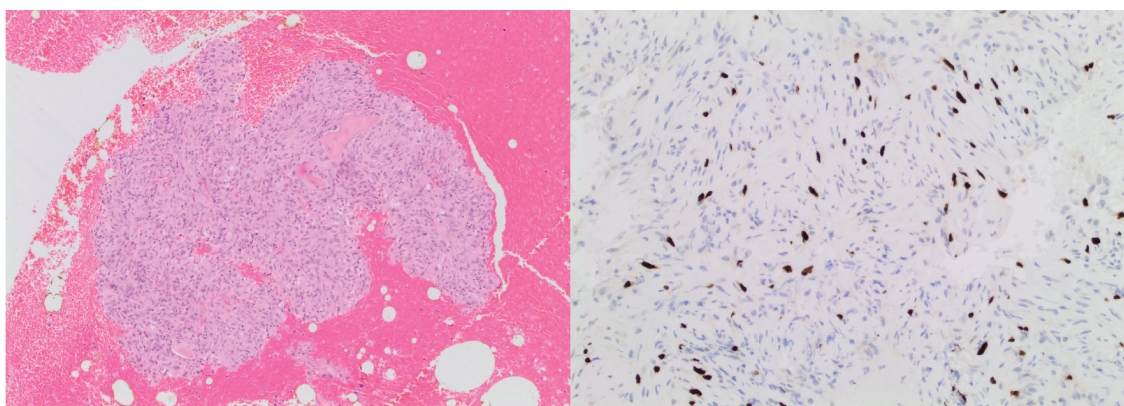


Figure 31 PD38328a biopsy GCT histological images. Left panel is a haematoxylin and eosin (H&E)-stained section of tumour amid cancellous bone. Right panel demonstrates scattered H3.3 G34W immunoreactive nuclei.

The resection specimen, however, showed no G34W immunoreactive cells, and the histological features of GCT, including osteoclasts were absent (**Figure 32**). In contrast the tumour possessed many mitotic figures and nuclear atypia, suggestive of a malignant tumour. The favoured diagnosis was therefore malignant GCT. This resection specimen was subjected to whole genome sequencing as discussed above (**Results 4.3.1**) and no reads were found to

support a *H3F3A/H3F3B* coding variant. Chromosome 1q was however deleted, which could potentially have included the original mutation (**Figure 33**). Of note, this sample had acquired homozygous deletion of *KDM4B*, potentially recapitulating the phenotype of a histone coding variant by another mechanism. The sample otherwise appeared similar genomically to the higher mutation burden malignant H3.3-mutated samples, with whole genome duplication, a high degree of aneuploidy and a number of focal complex rearrangement patterns, compatible with chromothripsis.

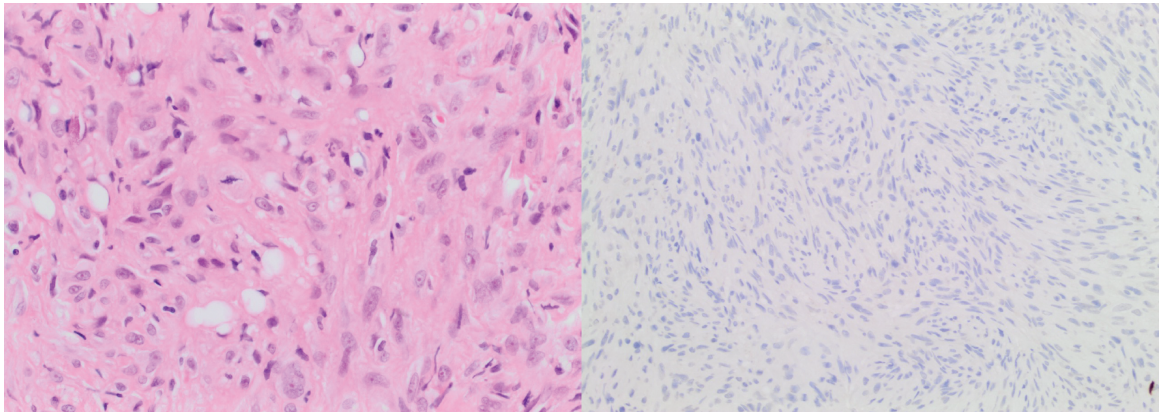


Figure 32 PD38328a resection malignant histology. Left panel is a high-power H&E-stained section of the resection specimen of PD38328a showing mitoses and nuclear atypia. The right panel demonstrates the absence of nuclear immunoreactivity for H3.3 G34W in this sample.

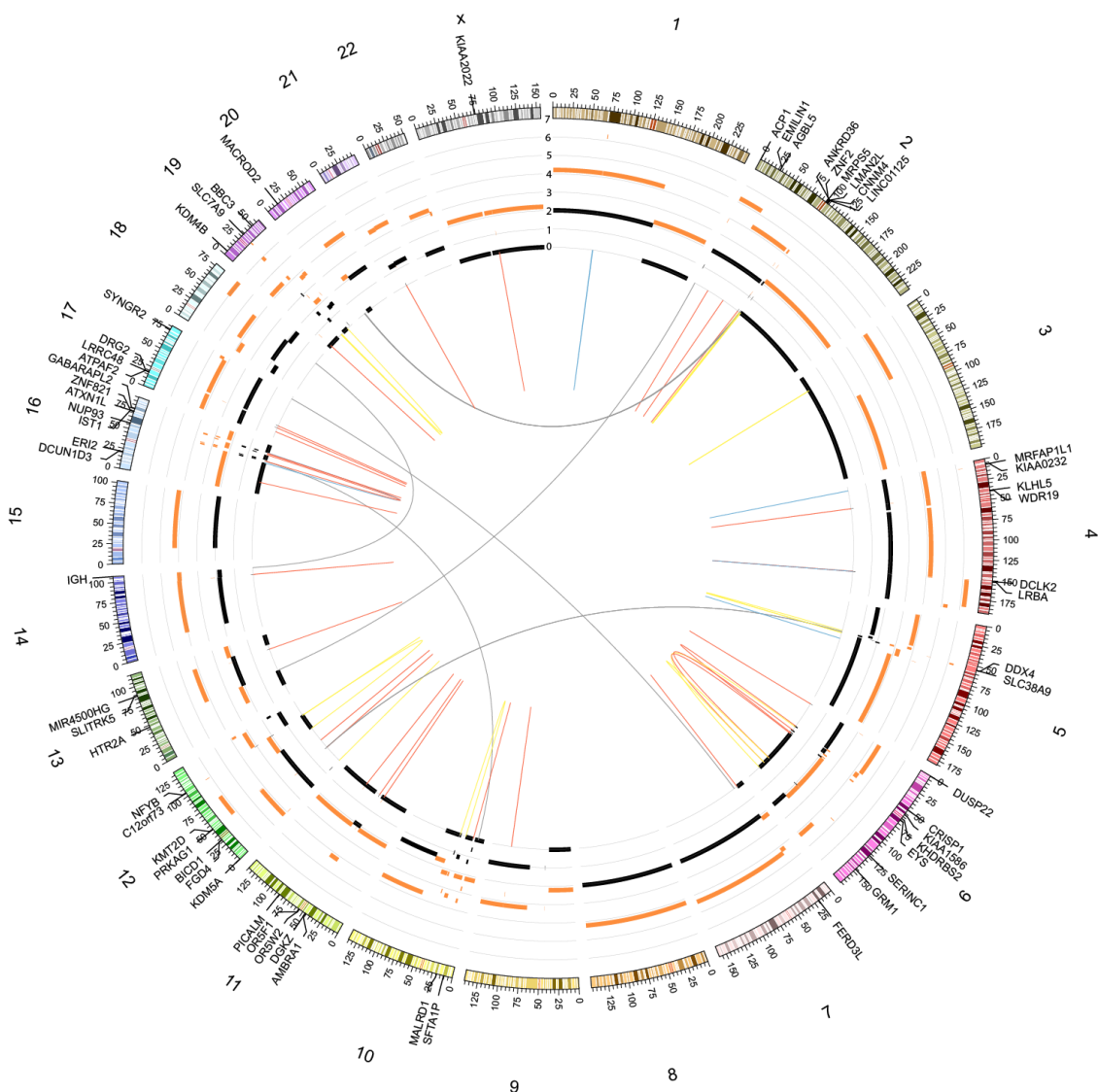


Figure 33 PD38328a circos plot of copy number and structural rearrangements. Format as previously displayed. Note the locus of *H3F3A* at chromosome 1q has lost heterozygosity

Whole genome sequencing copy number profiles for available malignant sarcomas were screened for deletions spanning the *H3F3A* and *H3F3B* loci. Deletions are frequently observed (**Figure 34**), potentially raising the possibility that there are other malignant tumours that have a common origin with GCT and malignant H3.3 mutant bone tumours. Specific gene level loss of heterozygosity (LOH) is, however, clearly associated with aneuploidy in general, so these deletions are not specific to this gene locus.

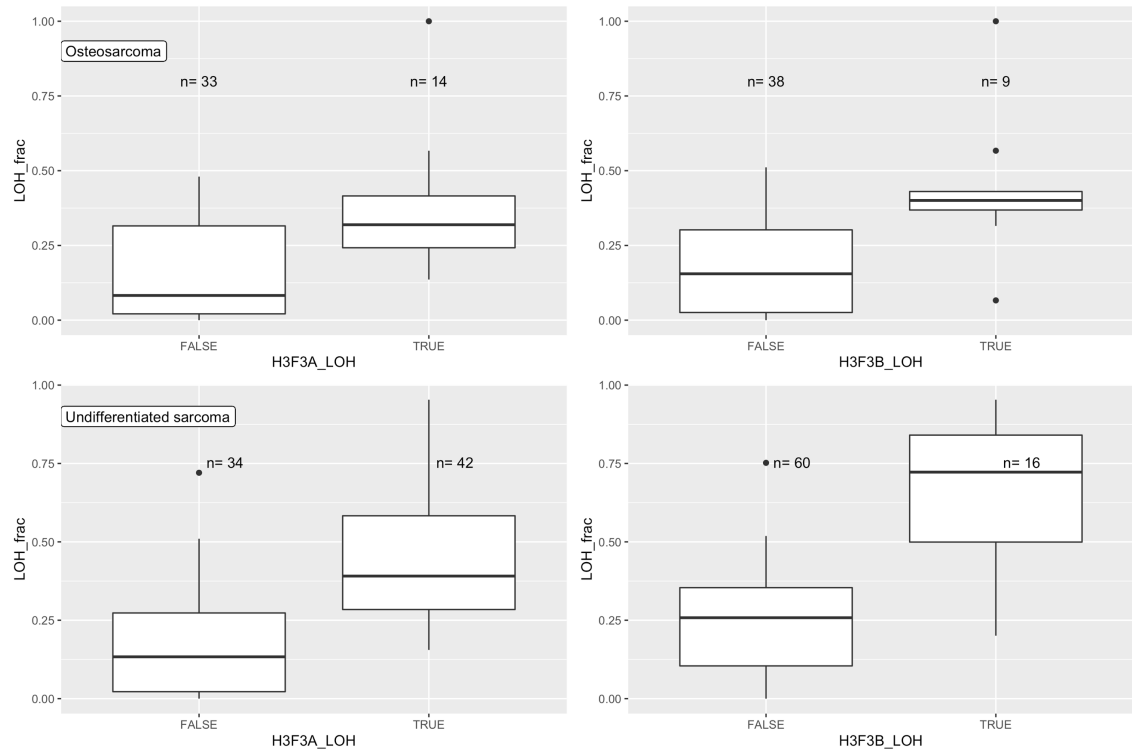


Figure 34 H3.3 gene LOH in osteosarcoma and undifferentiated sarcoma. Top panels related to 47 osteosarcoma whole genome sequencing derived copy number profiles whilst the bottom panels result from 76 whole genome sequencing-derived copy number profiles. LOH_frac is the fraction of the genome with LOH, while gene-specific LOH is shown on the horizontal axis.

4.3.4. Malignant H3.3 bone tumours have a distinct methylation profile

To explore the epigenetic differences between malignant and benign H3.3 mutated bone tumours I subjected a further tumour cohort to DNA methylation analysis. This tumour cohort, in part comprised some of the sequenced samples, as well as osteosarcoma as an H3.3 unmutated malignant counterpart and chondroblastoma as a benign tumour that has an alternative H3.3 mutation; *H3F3B*:p.K27M (**Appendix 7.2.2**). Raw methylation array data from GCTs and Malignant H3.3 mutant tumours were also contributed by a collaborator, some of which were published previously and discussed above (**Results 4.1.5**) (Koelsche *et al.*, 2017). DNA was extracted, predominantly from fresh-frozen tissues, bisulphite converted and analysed on either an Illumina 450k or EPIC array. Data were corrected for dye-biases and background fluorescence, as well as filtered for probes that were ambiguously located, non-autosomal or not common to both array platforms (**Methods 2.18**).

Unsupervised clustering based on the most variable methylation probes easily recapitulated the diagnostic groups (**Figure 35**). Furthermore, while closely related to conventional GCT, the malignant H3.3 mutant tumours also formed a distinct clade. Metastatic samples clustered with the benign GCT group.

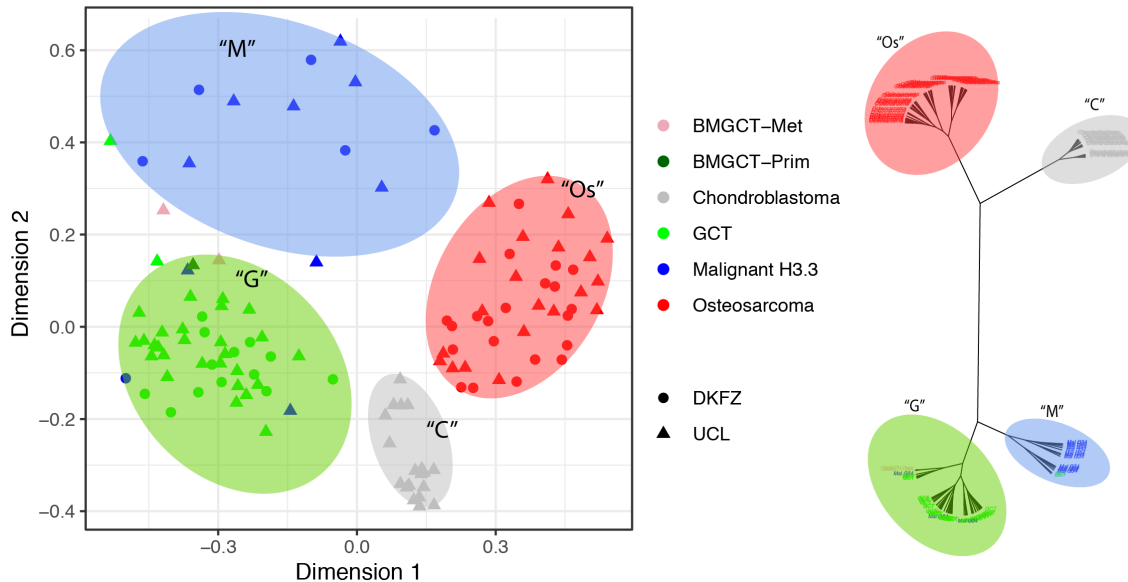


Figure 35 Methylation based unsupervised clustering Left panel is a multi-dimensional scaling (MDS) plot based on the methylation profile for each tumour. Ellipses are schematic for the unsupervised clustering allocation for each sample. Each cluster is named according to the predominant diagnostic group contained within: “G” - GCT. “M” - Malignant H3.3, “Os” - Osteosarcoma, “C” - Chondroblastoma. Unsupervised clustering allocations were derived from hierarchical clustering, shown in the right panel. Note that the MDS does not adequately project all the points at the interface between the “M” and “G” cluster, for example the pink ‘BMGCT-Met’ sample was assigned to the “G” cluster by hierarchical clustering.

To identify any genomic patterns underlying the distinct clustering of benign and malignant methylation profiles, differentially methylated probe and region analysis was conducted. Comparing significant and non-differentially methylated probes (DMPs) there was a significant depletion of regulatory sites amongst differentially methylated probes (**Figure 36**). Specifically, transcription start sites were less likely to be significant DMPs, whilst intergenic regions were enriched. Of these intergenic region probes, those that were significantly differentially methylated were less likely to be at enhancer sites (68% of significant DMPs vs. 91% of those that were non-significant).

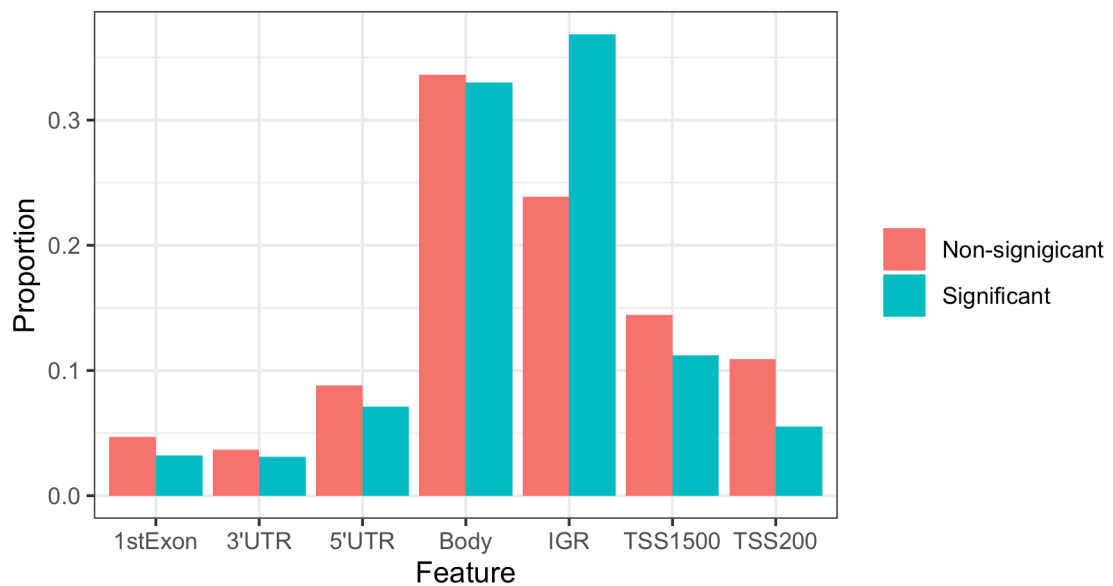


Figure 36 Genomic features of differentially methylated sites. Barplot of genomic features of methylation array probes that are significantly (adjusted p value $< 1 \times 10^{-5}$) or non-differentially methylated between malignant clustered tumours and benign clustered GCTs, using ChAMP (**Methods 2.20**). All differences are highly significant by Pearson's chi-squared test, except for 3'UTRs which is only marginally significant ($p=2 \times 10^{-3}$) and gene bodies which was non-significant ($p=0.21$). Abbreviations: UTR – Untranslated Region, IGR – Intergenic Region, TSS – Transcription Start Site (and either 200 or 1500bp upstream)

As gene regulatory elements were less differentially methylated it was unsurprising that regulatory pathways were not consistently modulated by DNA methylation (**Figure 37**). Of those pathways that did meet statistical significance it was noted that many were comprised of closely related genes found in dense genomic clusters, most notably the olfactory receptor genes (*OR*) and clusters of histone genes. To investigate the significance of these aberrations, bespoke analysis of methylation difference across the genome was conducted (**2.20**). This revealed genomic regions that were significantly differentially methylated between malignant and benign H3.3-mutated samples.

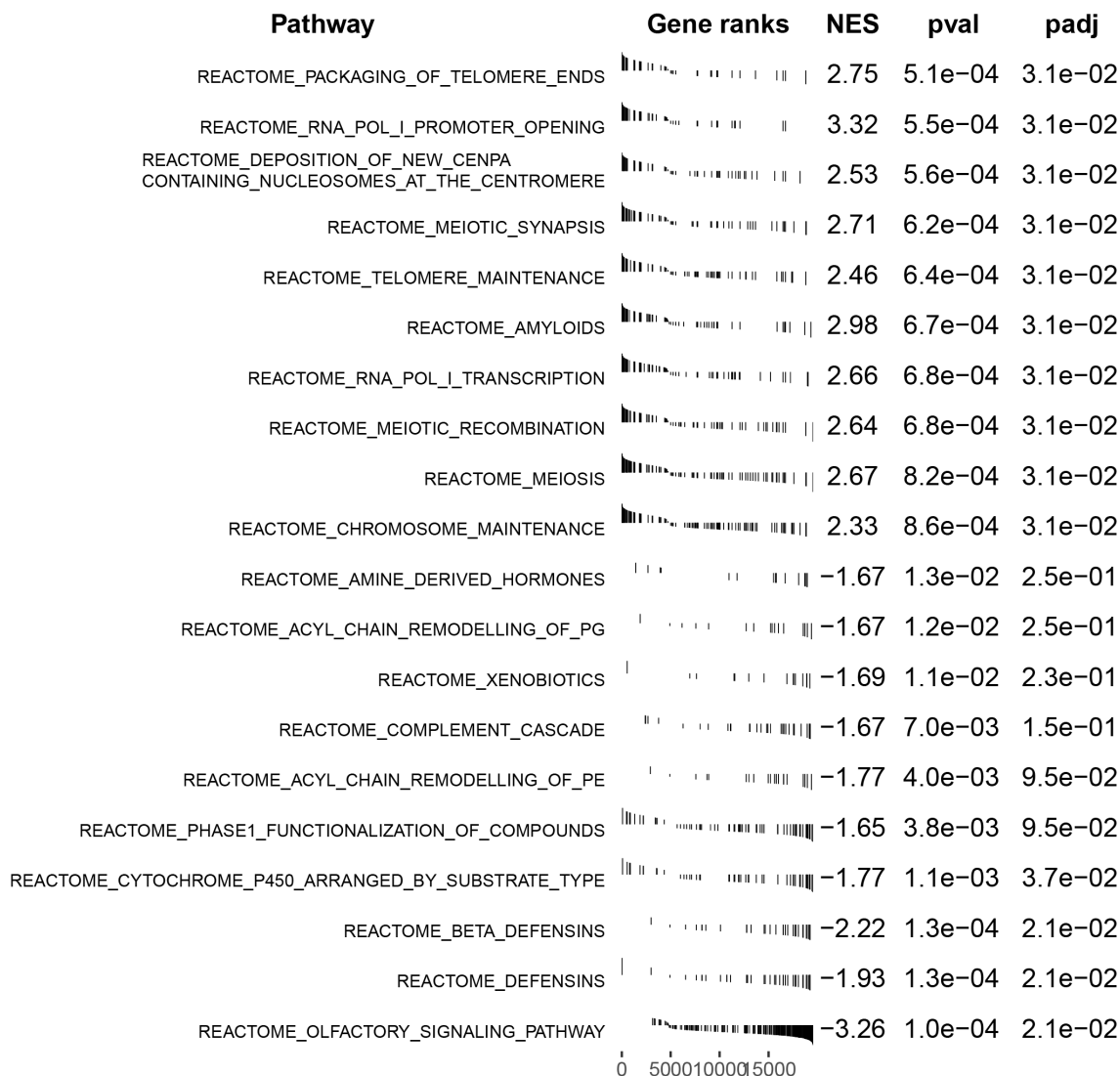


Figure 37 Gene set enrichment analysis for methylation data. Reactome gene sets are shown for methylation differences between benign and malignant H3.3 mutated tumours. NES (Normalised Enrichment Score), padj are Benjamini-Hochberg adjusted p-values

Specifically, when reviewing regions containing clustered histone genes there were short sections of the HIST1 cluster, located on chromosome 6, that were hypermethylated in malignant tumours (**Figure 38**). These seemed exclusively to contain histone genes. Conversely, of the large numbers of olfactory gene clusters, only a proportion were differentially methylated and many were part of larger differentially methylated regions (**Figure 39**). The biological significance of these regional methylation differences was unclear.

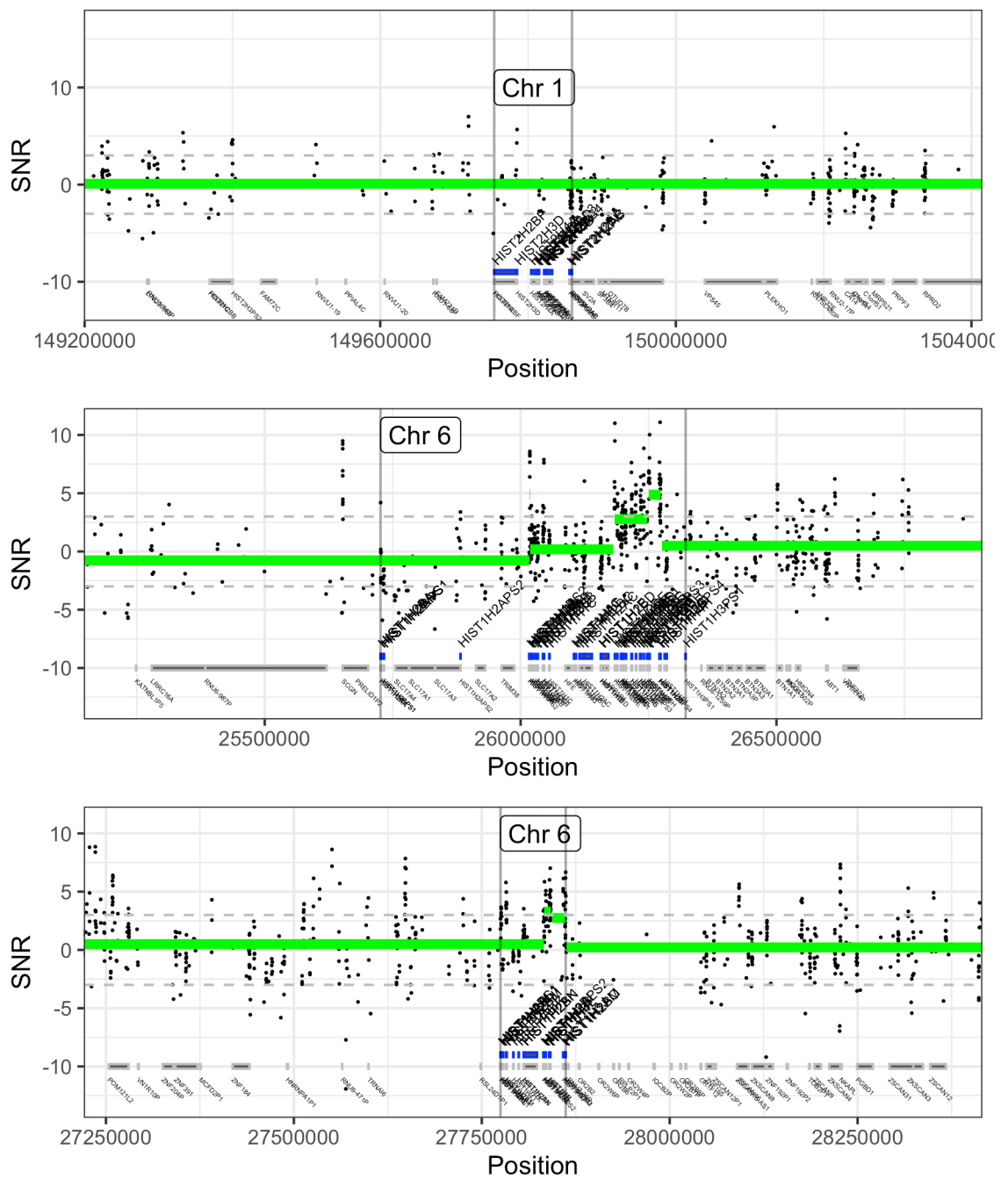


Figure 38 Differential methylation across histone clusters. Genomic regions containing clusters of histone genes (more than 5 genes with less than 200kb between each gene). Point are the signal-noise ratio (SNR) of methylation difference between malignant and benign tumours for each probe (negative values mean relative hypomethylation of malignant tumours). Green segments are segmented SNR values. Grey horizontal dashed lines denote significant SNR levels from permutation analysis. Black vertical lines are the boundaries of the displayed cluster. Blue segments and upward facing gene levels are genes identified by Differentially Methylated Region (DMR) analysis. Grey segments are all genes within the region. NB: One histone cluster on chromosome 6, with no segmented aberration is not shown. Chromosome 1 HIST 2 cluster shows no differential methylation in the top panel whilst section of HIST1 cluster on chromosome 6 show focal hypermethylation in malignant tumours.

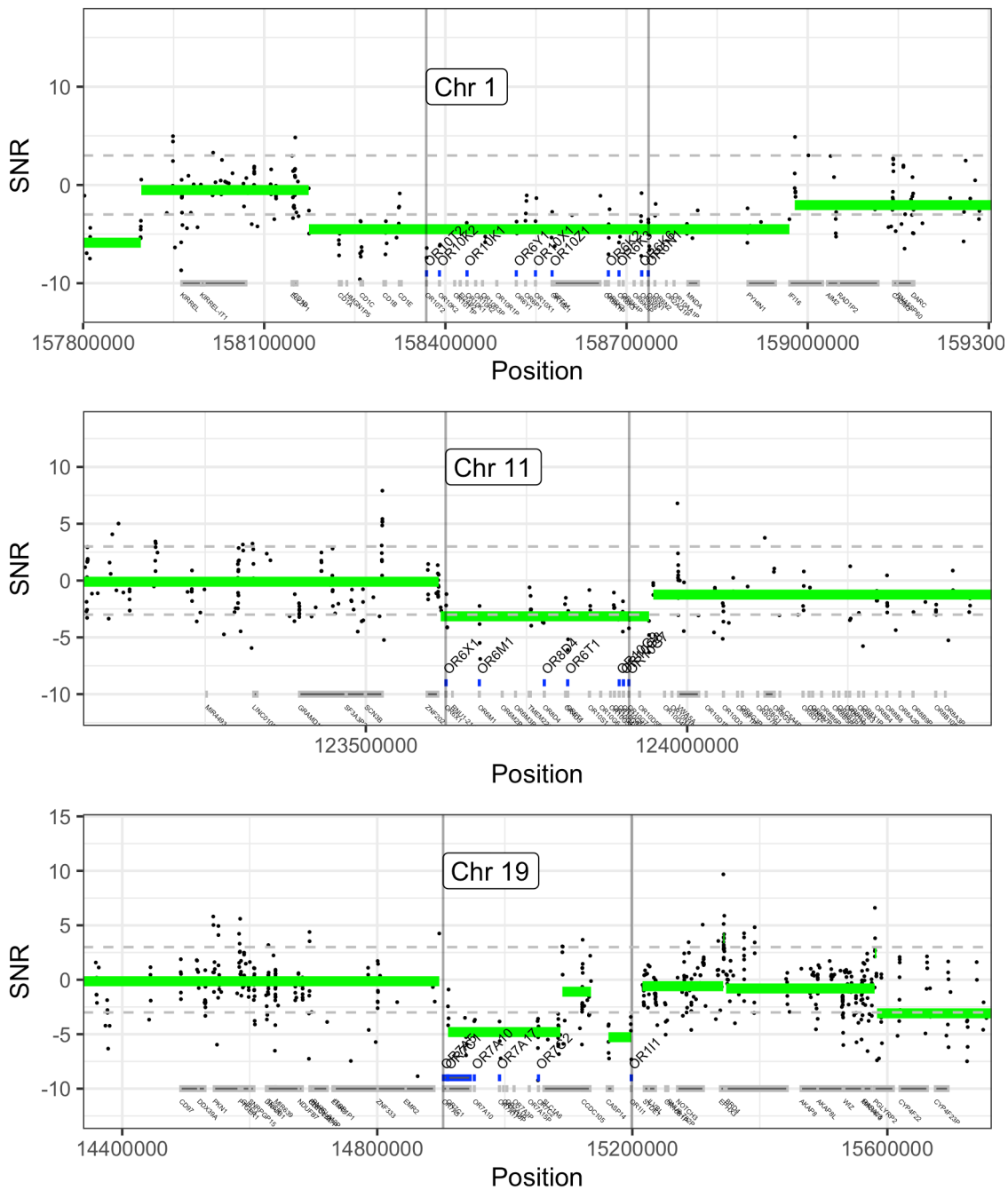


Figure 39 Differentially methylation around olfactory gene clusters. Figure format as for Figure 38. 3/16 representative olfactory gene clusters shown.

4.3.5. Benign H3.3 tumours have a hypermethylated *CCND1* promoter

In contrast to these global and larger regional epigenetic changes between benign and malignant tumours, aggregating methylation probes into differentially

methyated regions (DMRs) identified focal changes in a few specific gene promoters. Of 74 DMRs identified, 56 were identified to be focal around gene transcriptions start sites (**Figure 40**).

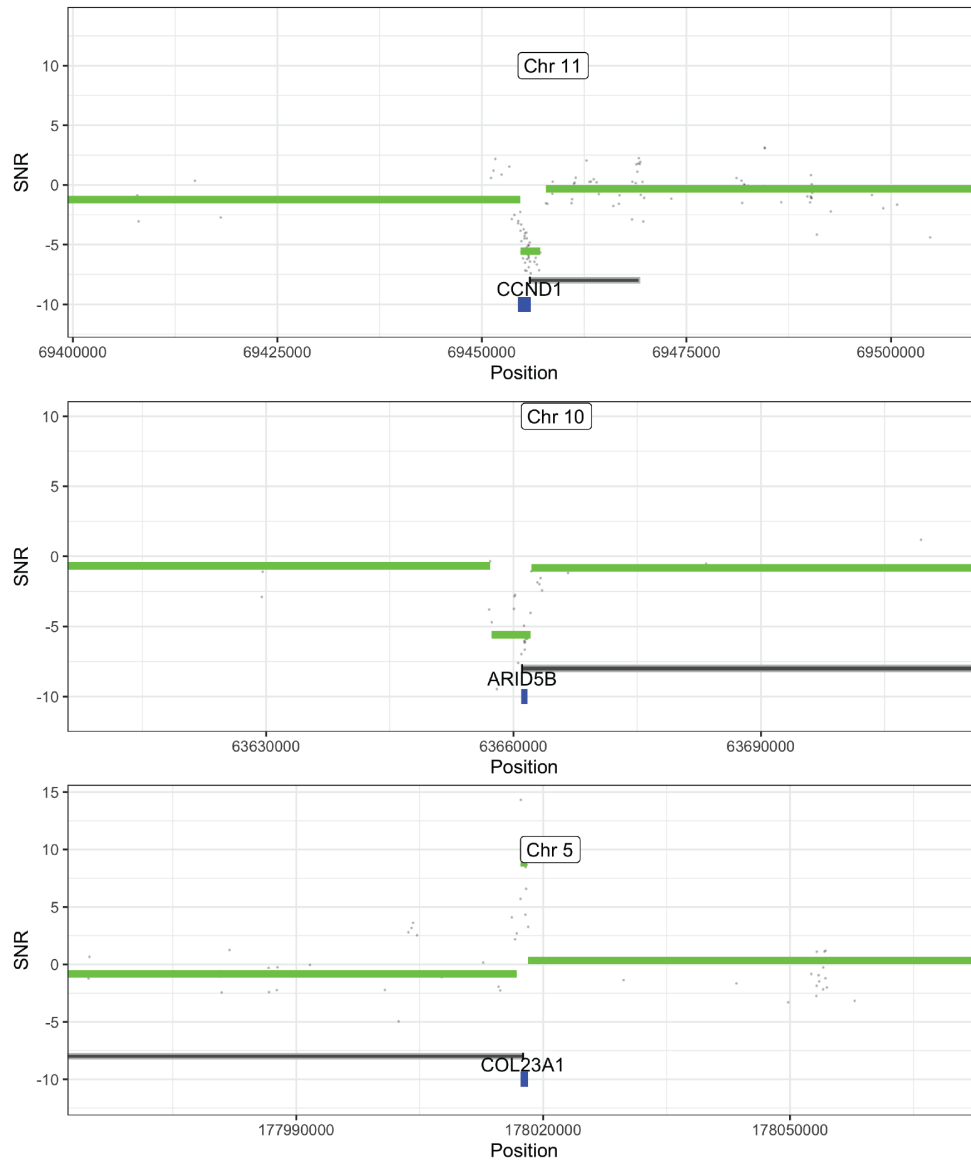


Figure 40 Focal differentially methylated regions (DMRs). Representative examples of the 56 DMRs focally identified to target gene transcription start sites. Grey points are aggregated signal-noise ratios (SNR) of malignant to benign methylation group differences. Green segments are segmented SNR values. Genes are denoted with a grey bar with a black tick identifying the known transcription start site whilst the blue segment is the ChAMP identified DMR.

The most statistically significant DMR was also the only one identified in a plausible cancer driver gene, *CCND1* (**Figure 41**). Differential methylation

spanned a promoter region of 1500bp either side of the transcription start site (TSS).

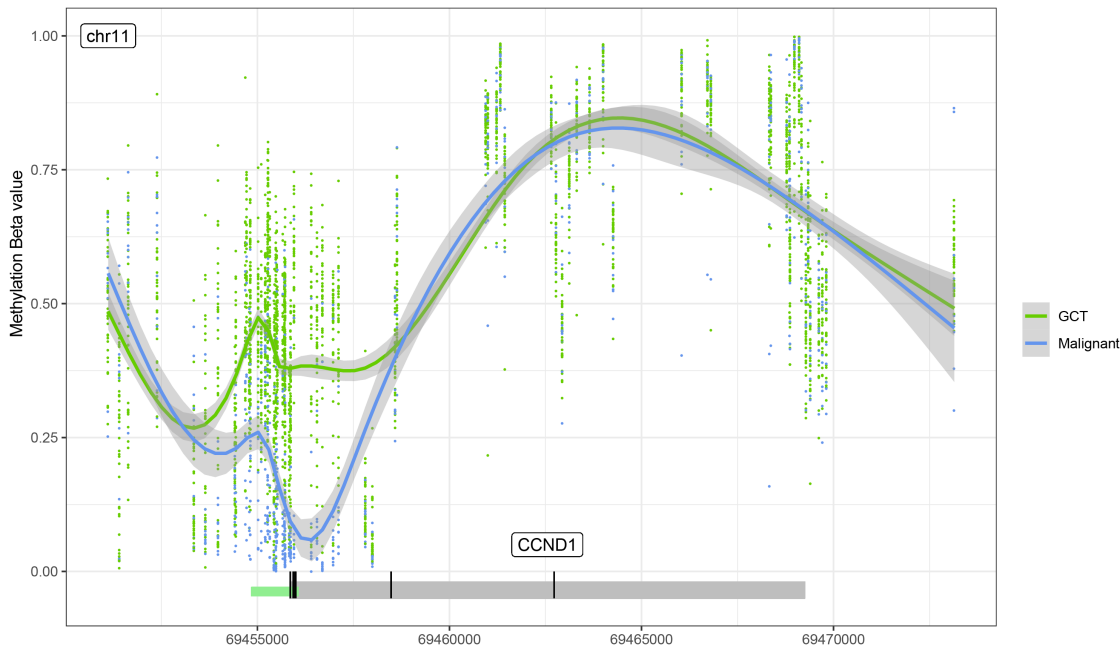


Figure 41 CCND1 differential methylation. Individual methylation probe methylation values for clustered malignant and benign (GCT) H3.3 mutant bone tumours. Lines are smoothed with a generalized additive model. The gene body is shown underneath with black marks highlighting known transcription start sites (TSS) and a promoter region highlighted by the green box (1500bp upstream and 200bp downstream of the 5' TSS)

Comparing the mean methylation level across this promoter region between different bone and soft tissue tumour types revealed that in fact hypermethylation at this site is particular to GCT (**Figure 42**). Malignant histone mutated tumours and chondrosarcoma were the only tumour types with any samples with a similar degree of *CCND1* promoter methylation. *CCND1* promoter methylation was concordant with unsupervised methylation cluster groups (**Figure 22** and **Figure 43**).

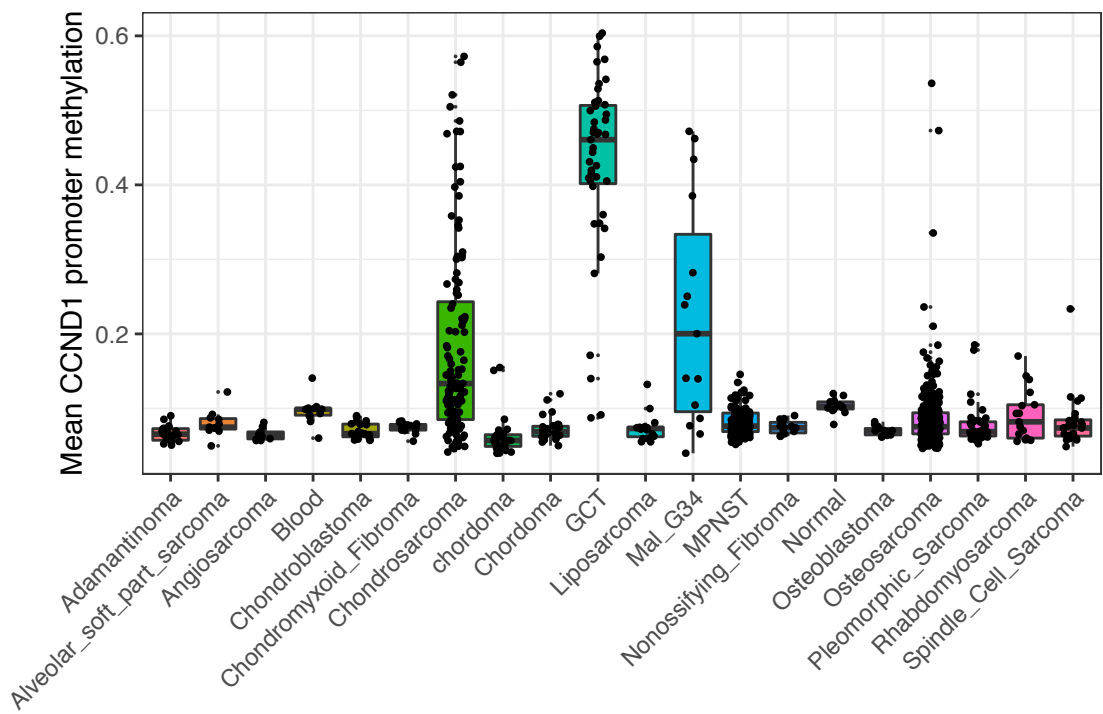


Figure 42 CCND1 promoter methylation across sarcoma types. Boxplot of mean CCND1 promoter methylation for a selection of normal tissues, benign and malignant bone and soft tissue tumours.

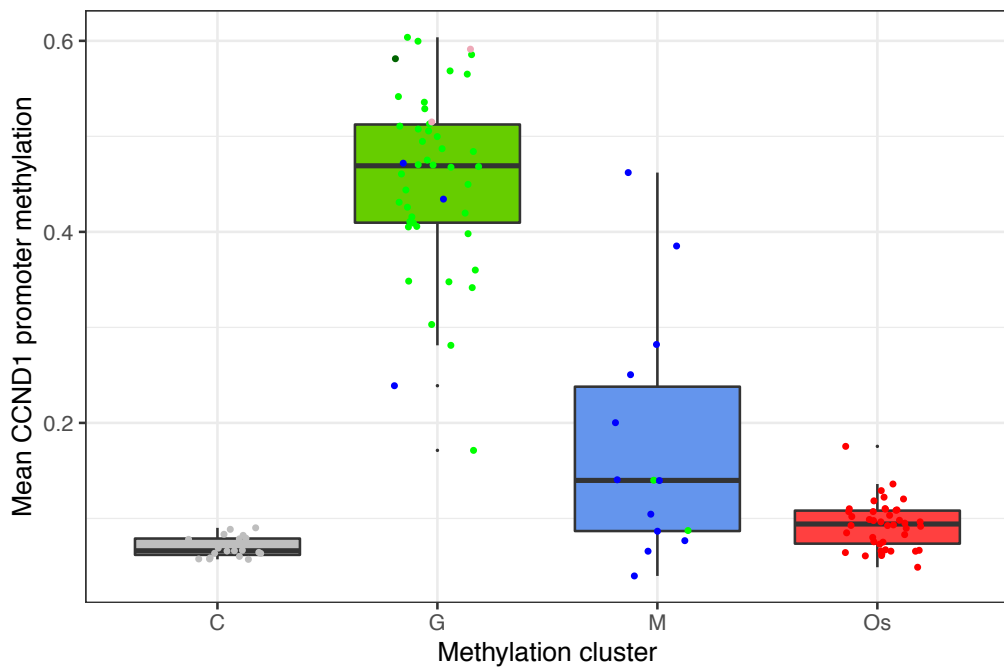


Figure 43 CCND1 promoter methylation in clustered tumours. Boxplot of tumour groups by unsupervised methylation cluster. Tumour diagnosis is shown with the colour for each dot, consistent colouring as from Figure 35

Of the sequenced cohort, only one benign tumour, PD3795d, was found to have a methylation cluster and *CCND1* promoter methylation state discordant with the final histological diagnosis of GCT. This sample, a recurrent GCT, was noted to have areas of focal atypia and osteoid deposition. This had been sufficient for a diagnosis of osteosarcoma by an externally reviewing bone tumour pathologist. Conversely the only malignant sequenced tumour with discordant methylation clustering and *CCND1* promoter status, was PD37332a, which as has already been noted, was a biphasic tumour. In this case, all genomic and methylomic findings suggest that the sampled benign component resembles a conventional GCT and not a malignant tumour which has arisen from it.

4.3.6. GCT metastases are polyclonally seeded

The GCTs with metastases and PD38329, in which the metastases were also subjected to whole genome sequencing and methylation array analysis, have all suggested that these tumours retain the genomic and methylation profile of a conventional GCT. PD38329 offered an unrivalled opportunity to explore the pattern of mutations within metastases. Using the algorithm DPCLust, mutations were clustered by the proportion of cells in each sample possessing them; the cancer cell fraction (CCF) (**Methods 2.16, Figure 44a**). Using this approach, I could confirm that all driver mutations seen in the sequenced samples were clonal. Leveraging the independent sampling across the three tumour samples, increased the power to define these mutation clusters. Mutations common to all cells in a sample can be considered 'clonal' while those in only a subpopulation are 'subclonal'. Both the primary tumour (PD38329a) and the two metastases (PD38329c and PD38329d) possessed a group of mutations private to that tumour sample only. This is a demonstration of heterogeneity across the tumour sites (**Figure 44b**). Furthermore, both metastases showed evidence of polyclonal seeding (**Figure 44c**). A subclonal cluster of mutations (orange cluster in **Figure 44**) was found to be subclonal in all samples. This suggests that the metastases were seeded by both a cell that possessed these mutations and a cell that did not.

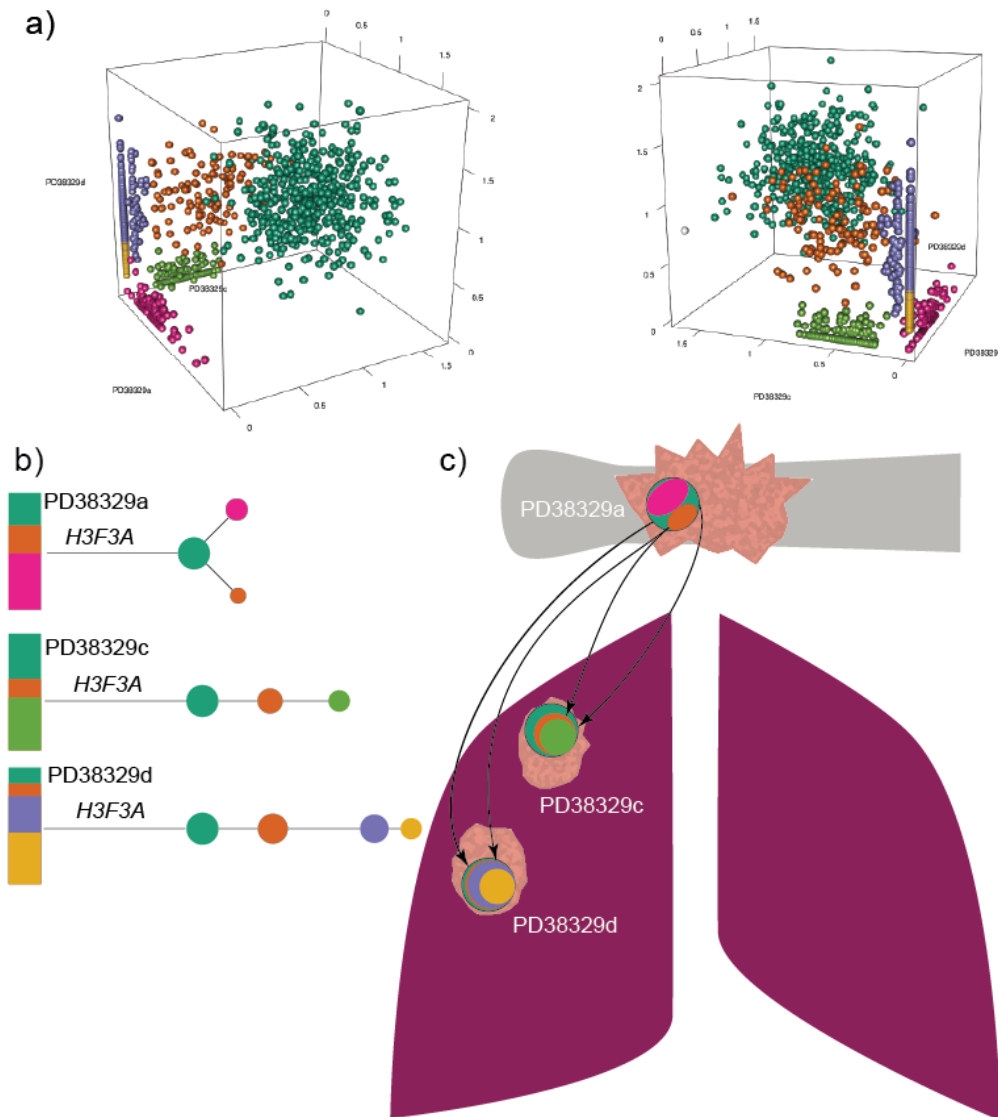


Figure 44 Mutation clustering reveals the pattern of metastatic seeding in benign-metastasising GCT. a) Two-dimensional projections of clustering of mutations by their CCF. Each dimension is one sample, points represent mutations. Points are coloured by the cluster to which they were assigned by DPCLust. Mutations clustering at CCF=1 (green) are clonal in all samples and therefore define mutations found in the most recent common ancestor to all sampled cells. b) clonal phylogenetic trees defined by mutation clustering for each sample. The length of connecting lines denotes the number of mutations acquired, the area of each circle is proportional to the number of cells possessing those mutations. The position of each subclone is defined by the 'pigeon-hole' principle: if the sum of the CCFs of two subclones is greater than their mutual parent, then one must lie within the other. Barplots of shown adjacent with CCFs scaled to the largest (clone) nested according to the phylogenetic relationship c) The pattern of clonal spread from primary tumour to metastases.

4.4. Discussion

4.4.1. Malignant H3.3 tumours acquire additional driver mutations

This study provides the first analysis of the genomic landscape of benign and malignant H3.3-mutated bone tumours. In keeping with previous findings, benign tumours (GCT), have a minimal mutation load, unrearranged genomes and are diploid. The absence of significant copy number aberration was validated in sizeable cohorts, appraised by either SNP or methylation arrays. No additional drivers were identified in GCTs.

In contrast, malignant tumours have an increased mutational burden and a spectrum of genomic rearrangements. The most aberrant genomes, in common with other sarcomas, have often undergone genome duplication and possess complex copy number and rearrangement events resembling chromothripsis (Steele *et al.*, 2019).

In this series of sequenced tumours, all malignant tumours had acquired at least one additional driver mutation. Mutually exclusively malignant tumours acquired an additional epigenetic modulator or replicative immortality. Replicative immortality was achieved either by a *TERT* promoter mutation, *TERT* rearrangement or, in common with other sarcoma types, Alternative Lengthening of Telomeres (ALT). The observed epigenetic modulator mutations were homozygous loss of function mutations in histone specific lysine demethylases (KDM). It is noteworthy that the previous exome study of conventional GCTs also noted isolated epigenetic modulator mutations, including KDM mutations (Ogura *et al.*, 2017). Furthermore, the loss of function of KDM4 family histone demethylases is known to recapitulate the epigenetic changes induced by G34R histone mutation in cell line models. Establishing the clinical outcome of the exome sequenced tumours, reported by Ogura *et al.* (Ogura *et al.*, 2017), could indicate whether they reflect early signs of transformation. If proven, additional driver acquisition could become a clinical predictor of risk of recurrence or progression.

4.4.2. Malignant H3.3 tumours have a distinct methylation profile

The methylation profile of malignant H3.3 tumours is highly similar to GCT but with distinct changes. They remain more closely related to GCT than to the malignant tumours they often mimic, namely osteosarcoma. There are, however, a number of differences between GCT and malignant tumours at the global, regional and local scale. These could, in part, reflect differences in the cell type composition of tumours. Methylation profiles derived from bulk sampling cannot, as yet, be readily purified to represent only a pure tumour profile. Methods to infer cell type composition, such as EpiDISH (Teschendorff *et al.*, 2017), rely on reference profiles for cell types, which do not adequately reflect mesenchymal tissues. Despite this caveat, tumour cells still reflect a substantial proportion of all samples (median sample purity, 49.7% (SD 13.7%) and 46% (SD 14.7%) for malignant and benign tumours respectively; **Methods 2.16**). Consistent changes seen across samples might also more closely reflect tumour cell properties.

Globally, many of the sites in which malignant tumours differ from GCT are non-enhancer sites in intergenic regions. They also have a number of large regional methylation changes that affect clustered genes, including a component of the Histone 1 cluster. Whether this has an additional epigenetic impact, altering the landscape of expressed histones is uncertain. Gene pathways do not seem to be markedly dysregulated by altered methylation patterns, however isolated genes, most notably *CCND1*, exhibit altered promoter methylation patterns. *CCND1* had no evidence of genetic change in any sample. Paradoxically, *CCND1* promoter methylation seems to be unusually high in GCT, while malignant tumours are hypomethylated to a level consistent with most other bone and soft tissue tumours and available normal tissues.

Work to explore the relationship of *CCND1* promoter methylation with expressed Cyclin D1 levels is ongoing. It is noteworthy that Cyclin D1 expression has predominantly been observed in the nuclei of osteoclast-like cells and not stromal cells. This is consistent with the hypothesis that GCTs silence *CCND1* expression by promoter methylation, but this has not been explored since the advent of a reliable marker to visualise tumour cells, namely H3.3 G34 mutation (Matsubayashi *et al.*, 2009). Immunohistochemistry, assaying for both the G34

mutation and Cyclin D1 expression could evaluate Cyclin D1 expression in the tumour cells of both benign and malignant GCTs. Should this in-situ approach not work, the ability to interrogate the effects of *CCND1* promoter methylation is limited by the lack of *in vitro* models. It is not known, as yet, whether introducing the histone mutation *in vitro* recapitulates the *in vivo* methylation findings.

4.4.3. Do malignant H3.3-mutated tumours transform from GCTs?

The malignant tumours assessed here were selected to possess the same driver mutations as conventional GCT. The evidence presented strengthens the argument for their common evolutionary origin. Both tumours are found in the same tissue type and are commonly defined by similar histological features, specifically the presence of giant cells, and present at a subarticular site. They are both defined by a mutation which is highly tumour-specific and likely to rely on a tissue-specific context. Evidence presented here demonstrated the close relationship of their methylation profiles, often considered a reflection of their cell of origins. Finally, where it was possible to assess, histone mutations could be shown to pre-date genome duplication. At one time ancestor cells of these malignant tumours possessed a histone driver mutation in the context of a diploid genome, that more closely resembles GCT.

To complicate the above evidence, I present a further case to add to the seven reported cases in the literature, where malignant progression of a GCT entails the loss of mutated histone expression. For the first time, from sequencing data, I show that it is associated with the deletion of the H3.3 locus. It is conceivable that a proportion of other malignant tumours, with copy number losses at this locus, have developed from H3.3 mutant tumours though clearly many will have acquired these losses by chance.

There is now a body of evidence that malignant H3.3-mutant bone tumours have evolved from GCTs, however it is still indirect. The possibility of the malignant tumour arising in parallel, though extremely unlikely, has not been entirely excluded. Conceivably, this could have occurred because of an undetected cancer predisposition variant in the germline or in other early dividing cells representing mosaicism. Direct evidence would require sufficient and high-quality

DNA from both a GCT and a malignant tumour from the same patient. This is regardless of whether or not the mutated histone was expressed in both. Identifying any shared somatic mutations between the samples would establish a common origin as identical mutations are unlikely to occur twice independently in the same patient. If the histone mutation were deleted, the mutated allele could potentially be phased to nearby SNPs. This would establish whether the mutation was present on the same parental allele that had been deleted. Unfortunately, in the presented case, sufficient DNA was not available from the GCT component and suitable analysis has not been performed on the published cases.

4.4.4. Benign metastatic GCTs resemble conventional GCT and may be polyclonally seeded

The cohorts analysed here suggest that primary and metastatic specimens from benign metastatic GCTs resemble other conventional GCTs. They possess no additional driver mutations and have diploid genomes. They are heterogeneous tumours, with private mutations in both primary and metastatic specimens. In contrast to many malignant tumours, metastases are heterogeneous and not seeded monoclally (Priestley *et al.*, 2019). In the presented example, a population of cells, possessing the same mutations, were found to be subclonal in the primary and both metastases. This demonstrates that the metastases were also seeded by cells lacking these mutations. Whether this occurred by the simultaneous seeding by a tumour embolus, as witnessed by Henry L. Jaffe (Jaffe, 1953), or by successive waves of seeding is unclear. This is however consistent with the concept of passive metastasis, metastatic clones that are embolised mechanically, without necessary the acquisition of a specific metastatic phenotype (Alberghini *et al.*, 2010).

BLANK PAGE

Chapter 5. Identification and timing of chromoplexy across cancer types

5.1. Introduction

Osteoblastoma and malignant giant cell tumours of bone both featured examples of complex patterns of genomic rearrangement. Chromothripsis was seen in malignant giant cell tumours of bone, while chromoplexy generated the *FOSB* rearrangement seen in osteoblastoma. Chromoplexy is relatively unexplored in other human cancers. The purpose of this chapter is to explore methods to identify chromoplexy, whether it generates driver events, and when it occurs during tumour evolution.

5.1.1. A mechanism and definition of chromoplexy

The term ‘chromoplexy’, was first proposed by Baca *et al.* after the observation of series of interconnected structural genomic rearrangements in a cohort of whole genome sequenced prostate cancers (Baca *et al.*, 2013). The term derives from the Greek term, ‘pleko’, meaning ‘to braid’, as several chromosomes were translocated to each other. The observed translocations were balanced, in that there was no loss of genomic material, however they were not directly reciprocal. Each chromosomal breakpoint could be connected to two further points in the genome. Complete ‘cycles’ were observed when the series of rearrangements could be connected back to the original starting point. Often however, the event was an incomplete ‘chain’.

As yet, a clear and unified definition of chromoplexy has not been proposed, particularly in the context of its overlap with chromothripsis. Chromothripsis has been clearly defined by Korbel and Campbell (Korbel and Campbell, 2013) to meet the following criteria:

1. Breakpoints should be clustered, have alternating (head/tail) support, and affect a specific haplotype
2. Copy number states should oscillate regularly between retention and loss of heterozygosity

3. The derivative chromosome should have a random segment order with respect to the reference genome

Few events perfectly meet all criteria because unrelated rearrangements or copy number changes, such as whole genome duplication, alter their appearance. Chromothripsis has at least one observed mechanistic basis, namely chromatid fragmentation within micronuclei during S-phase DNA replication and chromatin bridge rupture after telomere attrition (Zhang *et al.*, 2015).

In contrast, chromoplexy has no clear definition or experimentally validated mechanism. In published reports, it is loosely described as a series of chained or cycling rearrangements linking chromosomes together. When contrasted with chromothripsis it is suggested to have fewer breakpoints but a greater number of involved chromosomes. In general, the copy number around breakpoints is balanced, this is because no genomic material is thought to be lost. Neighbouring breakpoints can be linked by short losses of genomic material, termed 'deletion bridges'. These appearances have led to the suggestion that chromoplexy events result from simultaneous double stranded breaks with misrepair, possibly by non-homologous end joining (NHEJ). Breaks must be acquired simultaneously because if not, genomic losses would be inevitable and as stated above, copy number is generally balanced around breakpoints.

Simultaneous breakage has been proposed to occur in hubs of transcriptionally active genes. In both prostate cancer and Ewing's sarcoma, the genes involved in chromoplexy chains are known to be highly expressed (Anderson *et al.*, 2018, Baca *et al.*, 2013). In addition, in prostate cancer these genes were found, using Hi-C, to be co-localised in the nucleus (Baca *et al.*, 2013). In common with many highly transcriptionally active genes, they are also early replicating. In prostate cancer, these genes are likely to be under the transcriptional control of the Androgen Receptor (AR). Chromoplexy chains occur less in the presence of *CHD1* deletions and transcription of AR dependent genes has been demonstrated to be dependent on *CHD1* (Metzger *et al.*, 2016). *CHD1* is thought to co-localise in a complex with AR and the histone-specific lysine demethylase *KDM1A*, regulating the transcription of AR targets, including ETS family genes.

Deletion of *CHD1* might therefore reduce ETS family transcription and explain the reduced frequency of chromoplexy events.

5.1.2. Chromoplexy in prostate cancer

In prostate cancer, the fusion of *TMPRSS2* and the ETS transcription family gene, *ERG*, which had previously been observed (Berger *et al.*, 2011), frequently occurred in the context of a chromoplexy chain. In addition to ETS fusion events, tumour suppressor genes were commonly disrupted by chromoplexy, including *PTEN*, *NKX3-1*, and *TP53*. The pattern of complex rearrangement was dependent on the ETS fusion status. ETS fusion containing tumours commonly had chromoplexy chains spanning multiple chromosomes, with breakpoints frequently within genes. In contrast, those lacking an ETS fusion but with a *CHD1* deletion, had rearrangement events more closely resembling chromothripsis. These involved fewer chromosomes but with many more rearrangements, more commonly in intergenic regions. *CHD1* is a chromatin modifier known to maintain genome stability, initially thought to explain the genomic differences until the association with AR activity was found. To date, the Baca *et al.* series remains the most comprehensive report of chromoplexy.

5.1.3. Chromoplexy in other cancers

There are only limited reports of chromoplexy in other cancers. Baca *et al.* also applied their algorithm to detect chromoplexy chains, ChainFinder, to breast, lung, head and neck cancers and melanoma, reporting some chained events in all. ChainFinder has also, independently been applied to a small number of mesothelioma tumours and cell lines (Oey *et al.*, 2019). The authors of this study described a highly complex event in one tumour with hundreds of rearrangements across three chromosomes as being chromoplexy. That this event could also represent a chromothriptic event, highlights the ambiguity in the definitions of these events and the methods to detect them. All other reports of chromoplexy-like events have been detected by manual inspection of the constituent structural rearrangements and copy number states. Two extremely rare nuclear protein in testis midline carcinomas (NMC) were found to have the pathognomonic *BRD3/4-*

NUT rearrangement induced by chromoplexy (Lee *et al.*, 2017b). A much larger series of Ewing's sarcoma (Anderson *et al.*, 2018) were also found to have the pathognomonic *EWRS1-FLI1* (an ETS gene) fusion events induced by chromoplexy. Finally, a number of different fusion events, including *EML4-ALK*, previously reported in lung adenocarcinoma were found to be induced by either chromoplexy or chromothripsis (Lee *et al.*, 2019).

5.1.4. Methods to identify chromoplexy

ChainFinder is the only method explicitly designed to identify chromoplexy. This is a MATLAB implementation of a graph theory approach to identify chains. Essentially this works in three steps:

1. Clustering of breakpoints that are thought unlikely to have occurred independently because of their proximity
2. Association or annotation of additional breakpoints by spanning deletions
3. A global assessment of the chain to establish the probability that individual components occurred independently

Two other methodologies that are reported to be applicable to chromoplexy are worthy of note. In both cases they are better suited to other types of complex rearrangements.

CouGaR is an algorithm designed to work directly with aligned sequencing data but connecting together regions of genomic amplification at the same copy number level (Dzamba *et al.*, 2017). Aligned reads are searched for putative rearrangements that are connect these regions. As CouGaR primarily detects amplifications, which are not seen in chromoplexy, this is more suited to either amplified chromothripsis, or chromoanasythesis; a rearrangement event that involves the amplification of chromosomal segments (Liu *et al.*, 2011, Zhang *et al.*, 2015).

Weinreb *et al.*, presented a theoretical framework, though not a functioning algorithm, to detect events involving simultaneous breakpoints (Weinreb *et al.*, 2014). This, again, is better suited for non-chromoplexy events which have greater than two breakpoints in each cluster and more prevalent copy number changes.

All other reports of chromoplexy have involved manual curation of structural variants and their local copy number environment. Chains were identified beginning with known fusions that had been detected by rearrangement callers. These were then extended manually by identifying further, related structural rearrangements. No further attempts have been made to systematically catalogue these events in a pan-cancer setting.

5.2. Aim

To explore the definition, patterns and timing of chromoplexy across cancer types.

5.2.1. Objectives

- Explore methods to define chromoplexy and its underlying rearrangement pattern
- Identify the frequency of chromoplexy and its potential to produce driver events
- Define the timing of chromoplexy in tumour evolution

5.3. Results

5.3.1. The Pan Cancer Analysis of Whole Genomes (PCAWG) dataset

The Pan Cancer Analysis of Whole Genomes (PCAWG) project provided an ideal opportunity to explore chromoplexy across cancer types. PCAWG aimed to collate whole genome sequencing data across cancer types and produce high quality but standardised analyses (PCAWG, 2020). This was to be achieved with a global consortium, organised into working groups based on their areas of scientific interest and expertise. These projects ranged from technical analysis, to driver detection, and relevant to the chapter, the study of structural rearrangement and tumour evolution and heterogeneity.

The PCAWG cohort consisted of 2,658 patient donors, contributing 2,605 primary tumours and 173 metastases or local recurrences. Most cancer types were

represented, though not with a frequency matching population incidence (**Table 8, Appendix 7.3.1**).

Organ	Abbreviation	Included Subtypes	Cases
Neural Crest			
CNS	CNS-GBM	Glioblastoma	41
CNS	CNS-Medullo	Medulloblastoma; Desmoplastic medullo.; Large cell medulloblastoma	146
CNS	CNS-Oligo	Oligodendroglioma	18
CNS	CNS-PiloAstro	Pilocytic astrocytoma	89
Skin	Skin-Melanoma	Malignant melanoma	107
Endoderm			
Biliary	Biliary-AdenoCA	Papillary cholangiocarcinoma	34
Bladder	Bladder-TCC	Transitional cell carcinoma; Papillary transitional cell carcinoma	23
Colon/Rectum	ColoRect-AdenoCA	Adenocarcinoma; Mucinous adeno.	60
Esophagus	Eso-AdenoCA	Adenocarcinoma	98
Liver	Liver-HCC	Hepatocellular carcinoma; Combined HCC/cholangio; Fibrolamellar HCC	317
Lung	Lung-AdenoCA	Adenocarcinoma; Adenocarcinoma <i>in situ</i> ; Mucinous adenocarcinoma	38
Lung	Lung-SCC	Squamous cell carcinoma; Basaloid SCC	48
Pancreas	Panc-AdenoCA	Adenocarcinoma; Acinar cell Ca.; Mucinous adeno.; Adenosquamous Ca.	239
Pancreas	Panc-Endocrine	Neuroendocrine carcinoma	85
Prostate	Prost-AdenoCA	Adenocarcinoma	210
Stomach	Stomach-AdenoCA	Adenocarcinoma; Mucinous adeno.; Papillary adeno.; Tubular adeno.	75
Thyroid	Thy-AdenoCA	Adenocarcinoma; Adeno., columnar cell; Adeno., follicular type	48
Mesoderm			
Bone/Soft Tissue	Bone-Benign	Osteoblastoma; Osteofibrous dysplasia	7
Bone/Soft Tissue	Bone-Benign	Chondroblastoma; Chondromyxoid fibroma	9
Bone/Soft Tissue	Bone-Epith	Adamantinoma; Chordoma	10
Bone/Soft Tissue	Bone-Osteosarc	Osteosarcoma	38
Bone/Soft Tissue	SoftTissue-Leiomyo	Leiomyosarcoma	15
Bone/Soft Tissue	SoftTissue-Liposarc	Liposarcoma	19
Cervix	Cervix-AdenoCA	Adenocarcinoma	2
Cervix	Cervix-SCC	Squamous cell carcinoma	18
Head/Neck	Head-SCC	Squamous cell carcinoma	57
Kidney	Kidney-ChRCC	Adenocarcinoma, chromophobe type	45
Kidney	Kidney-RCC	Adenocarcinoma, clear cell type; Adenocarcinoma, papillary type	144
Lymphoid	Lymph-BNHL	Burkitt; Diffuse large B-cell; Follicular; Marginal zone; Post-transplant	107
Lymphoid	Lymph-CLL	Chronic lymphocytic leukaemia	95
Myeloid	Myeloid-AML	Acute myeloid leukaemia	10
Myeloid	Myeloid-MDS	Chronic myelomonocytic leukaemia; MDS with ring sideroblasts	2
Myeloid	Myeloid-MPN	Essential thrombocythemia; Polycythemia vera; Myelofibrosis	26
Ovary	Ovary-AdenoCA	Adenocarcinoma; Serous cystadenocarcinoma	113
Uterus	Uterus-AdenoCA	Adenocarcinoma, endometrioid; Serous cystadenocarcinoma	51
Ectoderm			
Breast	Breast-AdenoCA	Infiltrating duct carcinoma; Medullary carcinoma; Mucinous adeno.	198
Breast	Breast-DCIS	Duct micropapillary carcinoma	3
Breast	Breast-LobularCA	Lobular carcinoma	13
Total			2658

Table 8 Abbreviated PCAWG cohort. See Appendix 7.3.1 for the full table

Matched normal samples were available for all, predominantly from blood (2,064), but the remainder were from adjacent (87) or distant (507) normal tissues. All samples underwent whole genome sequencing with a minimum average coverage of 30x (modes at 38x and 60x) in the tumour and 25x (mean 39x) in the normal. After alignment and quality control they were analysed with an extensive array of somatic variant callers. These individual variant calls were compiled into definitive consensus catalogues in a variant-type specific manner (**Methods 2.22**). These variants underwent thorough validation to ensure a 80% sensitivity and >95% specificity (PCAWG, 2020).

5.3.2. The ChainFinder algorithm finds clusters of structural variants across all cancer types

In order to detect chromoplexy chains, ChainFinder was employed using high quality structural variant and copy number calls, produced by the respective PCAWG working groups. A total of 10,362 ChainFinder chains were identified and 1,826/2,626 samples contained at least one chain (69.5%, 30 samples were excluded because of unavailable input variant data and 2 because ChainFinder was unable to output results due to the excessive number of variants). Consistent with the prevalence of 88% reported by Baca *et al.*, 182/210 (86.7%) of prostate cancer samples contained at least one chain (**Figure 45**). Most carcinomas contained chains as did at least 90% of glioblastoma, melanoma, leiomyosarcoma, lung squamous cell carcinoma and ovarian serous adenocarcinomas. The only disease group not to have any chains detected were the myeloproliferative diseases, however 26/28 of these were benign conditions (myelofibrosis, polycythaemia and essential thromocythaemia) and the remaining two myeloproliferative cases were a chronic myeloid leukaemia and a myelodysplastic syndrome, neither of which would be expected to possess complex rearrangements.

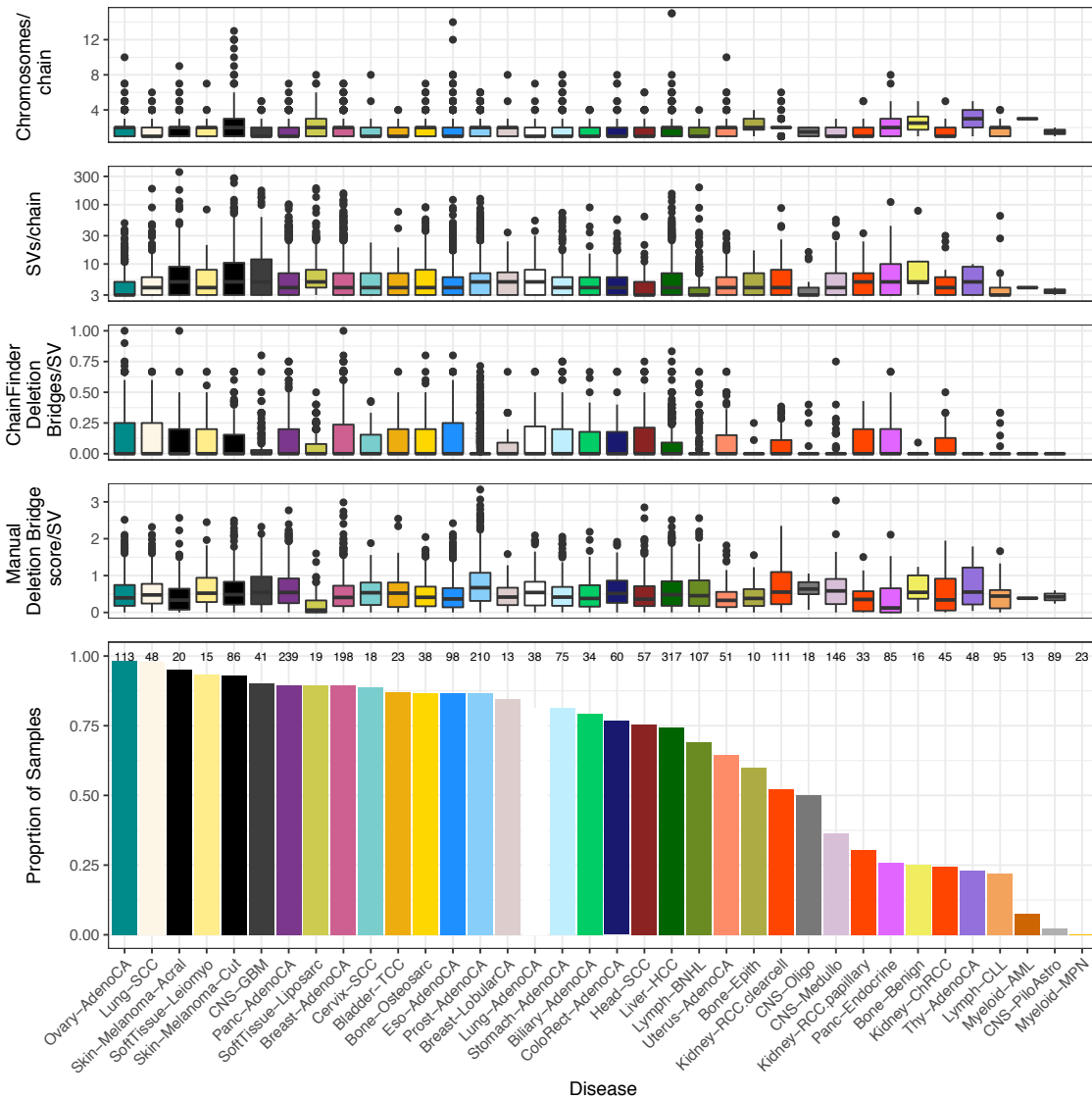


Figure 45 PCAWG ChainFinder chains. Characteristics of ChainFinder chains each disease group. The top four panels represent box and whisker plots, displaying median values, inter-quartile range and outlier values. From top to bottom: The number of chromosomes per chain, the number of structural variants per chain, the number of ChainFinder defined deletion bridges per structural variant in each chain, the computed summed $-\log_{10}$ adjusted pvalue for deletion bridges per structural variant, and a barplot of the proportion of samples with at least one chain.

5.3.3. ChainFinder chains are predominantly short and lack the expected deletion bridges

Detected chains were short (median of 4 rearrangements, range 3-355) but 1385/10362 (13.4%) involved more than 10 structural variants. Chains involved relatively few chromosomes (median 2, range 1-15) with some variability across

cancer types but 4877 (47%) and 3740 (36%) involved only 1 or 2 chromosomes respectively (**Figure 45**).

DNA breaks in chromoplexy are expected to either be repaired conservatively, in which case the two detected breakpoints will be at the same genomic locus or alternatively involve blunt-end attrition in which case a short deletion bridge might be detected. Across all chains, ChainFinder reported 157,044 pairs of significantly adjacent breakpoints. Many breakpoints were adjacent to more than one other breakpoint. The separation of these breakpoints was tri-modal, with peaks at 0, 45 and 230bps respectively (**Figure 46**). ChainFinder detected only 20,491 deletion bridges (13%) between these adjacent pairs but with a marginal excess at higher adjacency distances. It is noteworthy that many longer adjacencies (>10kb) were still frequently not reported as deletions, even though copy number calling methods would be expected to detect them.

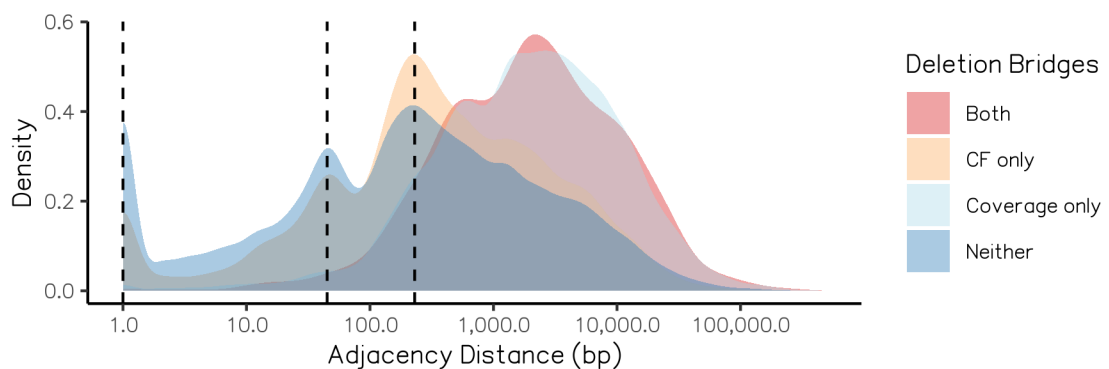


Figure 46 Breakpoint adjacency distance density plot. The adjacency distance for all 157,044 ChainFinder breakpoint pairs. Peaks at 0, 45 and 230bp are shown. Density is shown separately for pairs designated as a deletion by ChainFinder (CF, 18,702), coverage detection (17,021), both (1,789), or neither (119,532).

Many potential deletions were too small to be detected by SNP based copy number calling methods. Coverage was therefore scrutinised to detect evidence for deletions (**Figure 47, Methods 2.25**). This approach detected 17,021 further deletions, however 119,532 breakpoint pairs still had no evidence of deletion, 31,895 (26.7%), of which were >1kb and therefore would be expected to be detectable.

In summary, ChainFinder chains were shorter, involving fewer chromosomes and often lacking the copy number appearances that would be expected from the described pattern of chromoplexy.

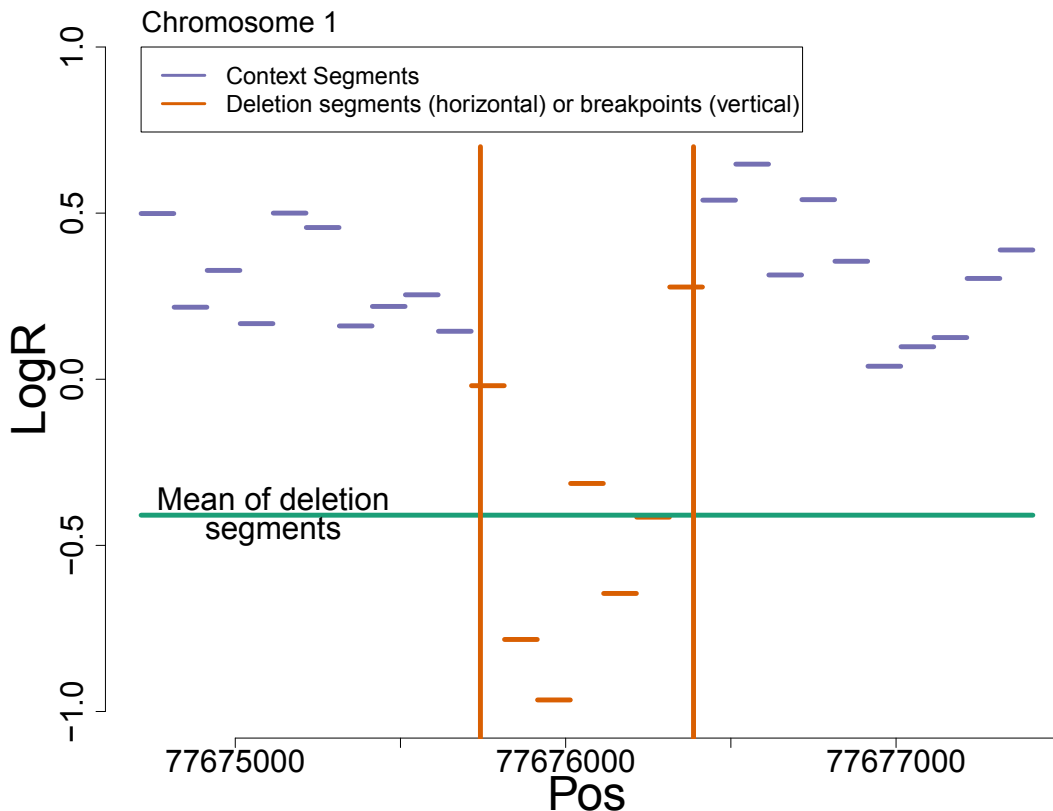


Figure 47 Example deletion bridge. An example breakpoint adjacency checked for a deletion from a pancreatic adenocarcinoma. Segments represent normalised coverage (LogR) over 100bp bins. Orange segments lie between the potential adjacent breakpoints while blue segments are ten bins, of equal size, lying on either side of the breakpoints for comparison. The green line shows the mean LogR for the deletion segments.

5.3.4. Many ChainFinder chains represent chromothripsis or retrotransposon insertions

Manual review of ChainFinder chains highlighted many that were reminiscent of two other known rearrangement phenomena, namely chromothripsis (**Figure 48**) and retrotransposon insertions (**Figure 49**). To systematically evaluate the overlap of ChainFinder events with chromothripsis, all structural variants were annotated for whether they lay within regions of high DNA breakpoint density. These regions were identified as part of a parallel work, by Maxime Tarabichi, to identify chromothripsis in the PCAWG cohort (**Methods 2.27**). This demonstrated that 2,815/10,362 (27.2%) of chains were found to be entirely comprised of

structural variants within these high breakpoint density regions. These chains are therefore more likely to represent chromothripsis events.

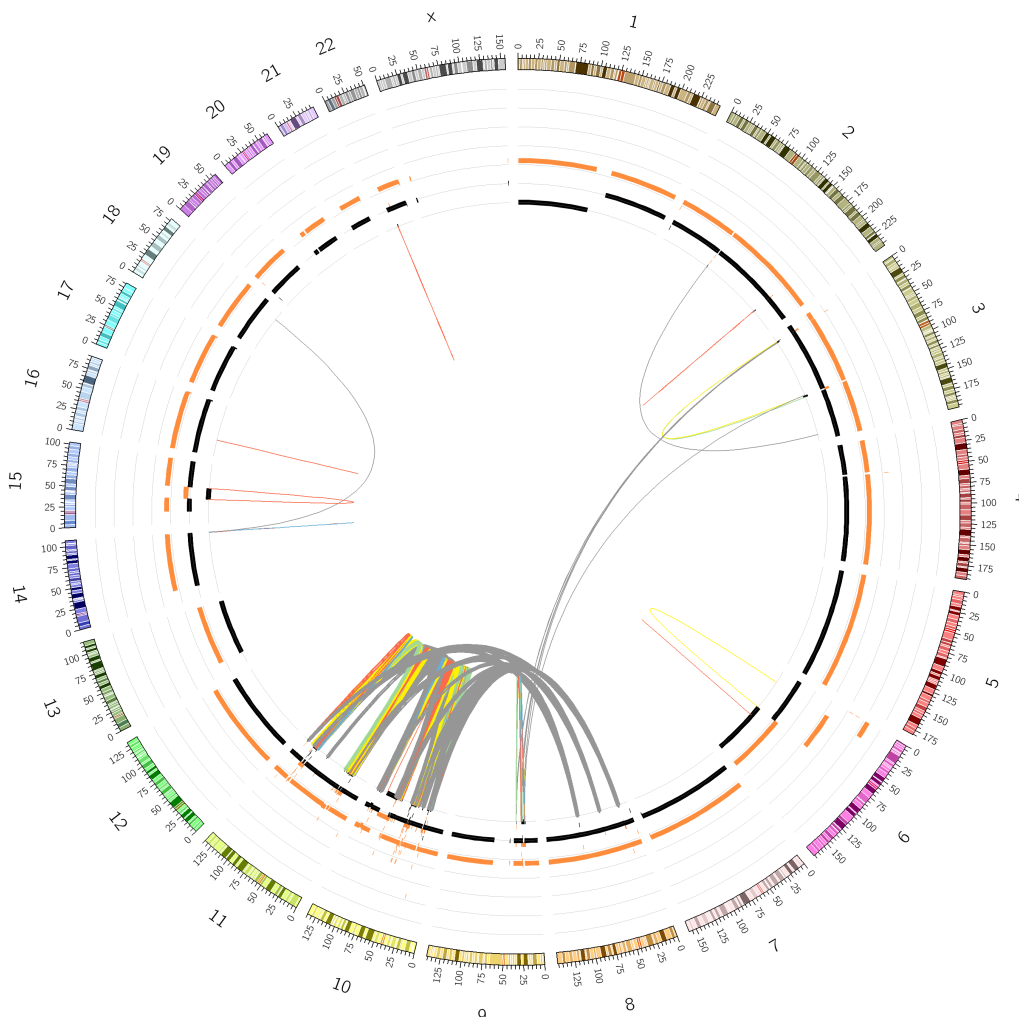


Figure 48 Chromothripsis-like ChainFinder chain. Circos plot of a B-cell non-Hodgkin lymphoma with a structural variants shown as the inner links. Thick links, involving chromosomes 8-11 were highlighted by ChainFinder as being a potential chromoplexy chain. Battenberg derived copy number is shown as segments (black is minor allele, orange is total copy number), with 0 as the inner ring. SV links are coloured as follows: grey – translocations, yellow/green – inversions, blue – gains, red – deletions.

In order to identify ChainFinder events that represent retrotransposon insertions, chains were annotated if they had clusters of breakpoints within close proximity to repeat elements, recorded in the RepeatMasker database. This revealed 988/10,334 (9.5%) chains that contained clusters of at least 5 breakpoints within 1kb of each other that were within 500bp of a repeat element. These chains are

more likely to represent retrotransposon insertions (Rodriguez-Martin *et al.*, 2017).



Figure 49 Retrotransposon-like ChainFinder Chain. Circos plot of retrotransposon event in a uterine adenocarcinoma. Format as for Figure 48.

As ChainFinder chains poorly reflected the expected pattern of chromoplexy and frequently were shown to represent other types of complex rearrangements I sought an alternative definition of chromoplexy with the collaboration of the PCAWG structural variant working group.

5.3.5. Breakpoint and copy number configurations define rearrangement events

The structural variant working group of PCAWG conducted a constructive approach to define clustered structural variant breakpoints (Li *et al.*, 2020). Critically this approach primarily works at the level of breakpoints, of which each structural variant implicitly has two. In brief, the principles of the approach are:

1. Clustering breakpoints into ‘footprints’ that are thought to be statistically unlikely to occur independently. This relies on the analysis of the sample’s breakpoint pattern and density (distinct from ChainFinder which uses a panel of samples but does not consider breakpoint type)
2. Merging footprint breakpoints with coverage-based copy number
3. Cataloguing the breakpoint/copy number configuration of footprints across PCAWG, using these to define event types
4. Incorporating all footprints that are connected by structural variants into clusters, defined by the type of their constituent footprints

This approach is highly conservative. Only highly recurrent and relatively simple footprints will be repetitive and therefore amenable to explanation. Many footprints are almost unique in their exact configuration, chromothripsis exemplifies this, as each event implicitly involves a random configuration of breakpoints. Chromothripsis events are therefore not expressly identified by this approach and are classified as ‘complex unclear’.

Annotating ChainFinder chains by the footprint classes of their constituent structural variants showed that most chains were classified as ‘complex unclear’ (**Table 9**). Chromothripsis and retrotransposon insertions comprise 3517/9016 (39%) of these, but the nature of the remaining complex unclear events remains uncertain.

Type		ChainFinder SVs	All PCAWG SVs
Complex unclear		9016	158,308
Template insertion		499	7,799
chains/cycles			
Deletions		349	57,674
Tandem Duplications		172	50,290
Chromoplexy chains/cycles		138	1987
Other (incl. inverted duplications)		89	16221
Direct Inversions		52	3218
Chromoplexy with template insertions		47	180
Total		10362	295,677

Table 9 PCAWG SV classifications of ChainFinder chains. The classification of structural variants included in ChainFinder chains is shown with the total number of these classes variants across PCAWG shown.

This alternative approach also identified some ChainFinder events with much more simple intra-chromosomal aetiologies, including simple deletions, tandem duplications, and inversions. The difference in these classifications originates from the differing approach to clustering the underlying breakpoints. The PCAWG approach is able to separate some ChainFinder chains into a number of distinct simpler events instead of grouping them together.

5.3.6. The appearances of chromoplexy and templated insertion footprints can overlap

Some chains were classified as chained or cycle events either termed chromoplexy or templated insertions. Templated insertions and chromoplexy ('balanced') footprints are both comprised of a pair of breakpoints, but with the opposite orientation and copy number pattern (**Figure 50**).

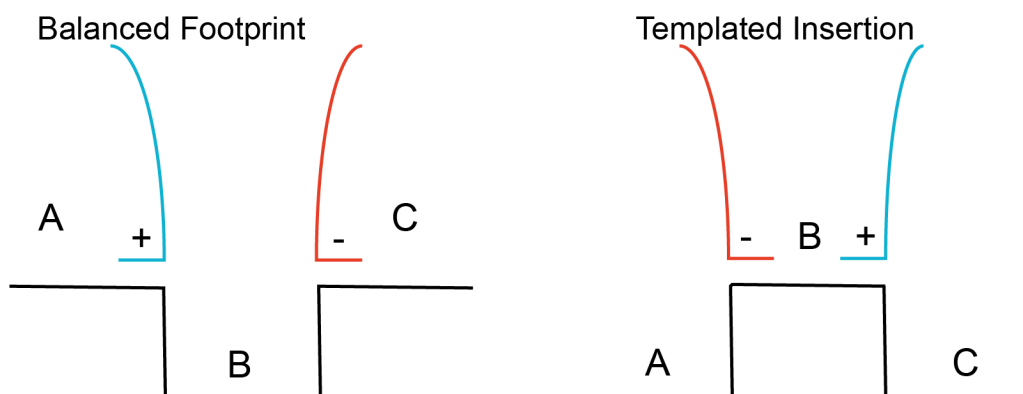


Figure 50 Footprint schematic. Copy number across a footprint is represented by the black segments, B is deleted in balanced footprints and gained in a templated insertion. Breakpoint orientation, which is the direction of reads supporting the breakpoint) are also as shown. Positive orientation implies reads that are directly mapped to the reference strand, while a negative orientation implies that the reads supporting the breakpoint are all reverse-complementary to the reference strand.

The true genomic outcome of a templated insertion and a balanced footprint are opposite. In the schematic in Figure 50, a balanced footprint results in segments A and C being translocated to other genomic regions. In the simplest of cases, this could represent a reciprocal translocation. In contrast, templated insertion is the copying of segment B into another part of the genome, amid a series of template sequences if the event is chained. The native A-B-C sequence would be expected to be left intact, otherwise an additional deletion rearrangement, spanning A and C would be expected.

Footprints with a templated insertion configuration of breakpoints can still represent a balanced breakpoint as demonstrated in the *FOSB* rearrangement observed in a case of osteoblastoma (**Results 3.3.3** and **Figure 18**). The alignment position and clipping pattern of sequencing reads can, when the breakpoints are close together, disentangle this ambiguity, as they did for the *FOSB* rearrangement. Templated insertion breakpoints should be phased to one another, in that the derivative genomic material in segment B should be contiguous with the partners of each breakpoint. Any reads discordantly mapping from one breakpoint partner into segment B should not align to segments A or C. Instead any bases beyond the second breakpoint will represent sequence from the second rearrangement partner, a sequence aligner will typically clip these bases, potentially chimerically aligning them elsewhere (**Figure 51**).

Templated Insertion

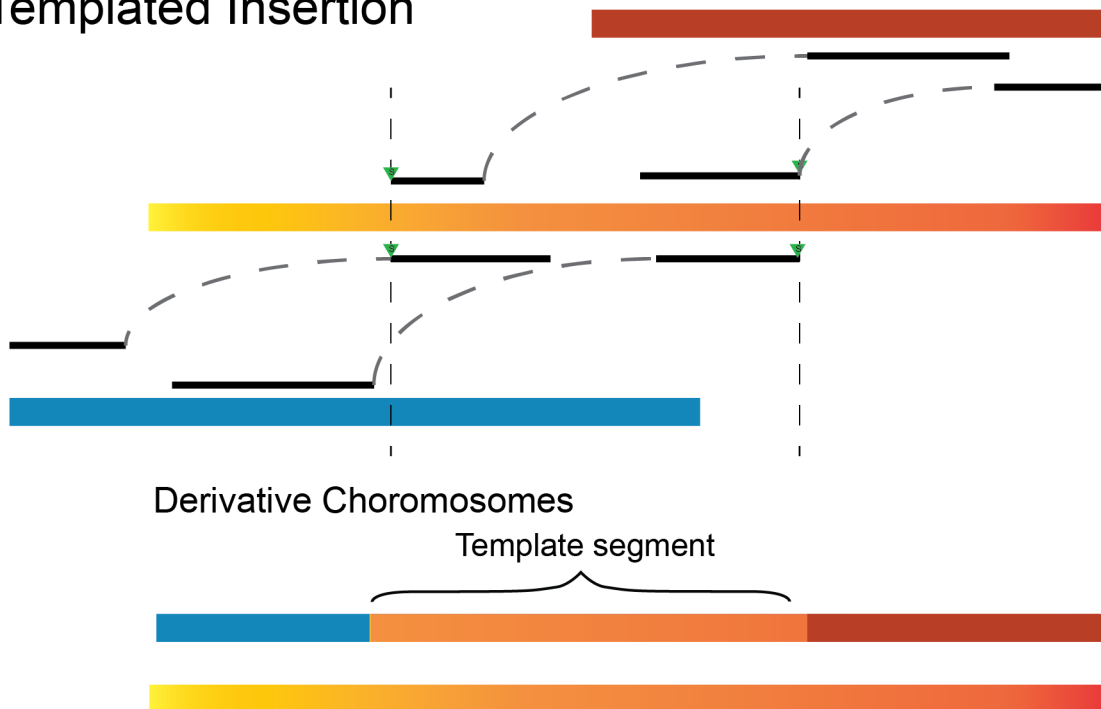


Figure 51 Schematic of templated insertion phasing support. The chromosome to which the template insertion footprint is mapped is shown with the yellow-orange gradient, whilst the rearrangement partner chromosomes are shown in blue and red. Breakpoints are shown as vertical dashed lines. Read pairs are shown as horizontal segments connected by dashed arcs. Read clipping is shown by green triangles. The derivative chromosome is shown underneath, demonstrating the yellow-orange chromosome is also left intact.

Conversely, balanced breakpoints with an inverted orientation can have reads that discordantly map from one breakpoint which extend beyond the second (Figure 52).

Balanced Footprint (chromoplexy)

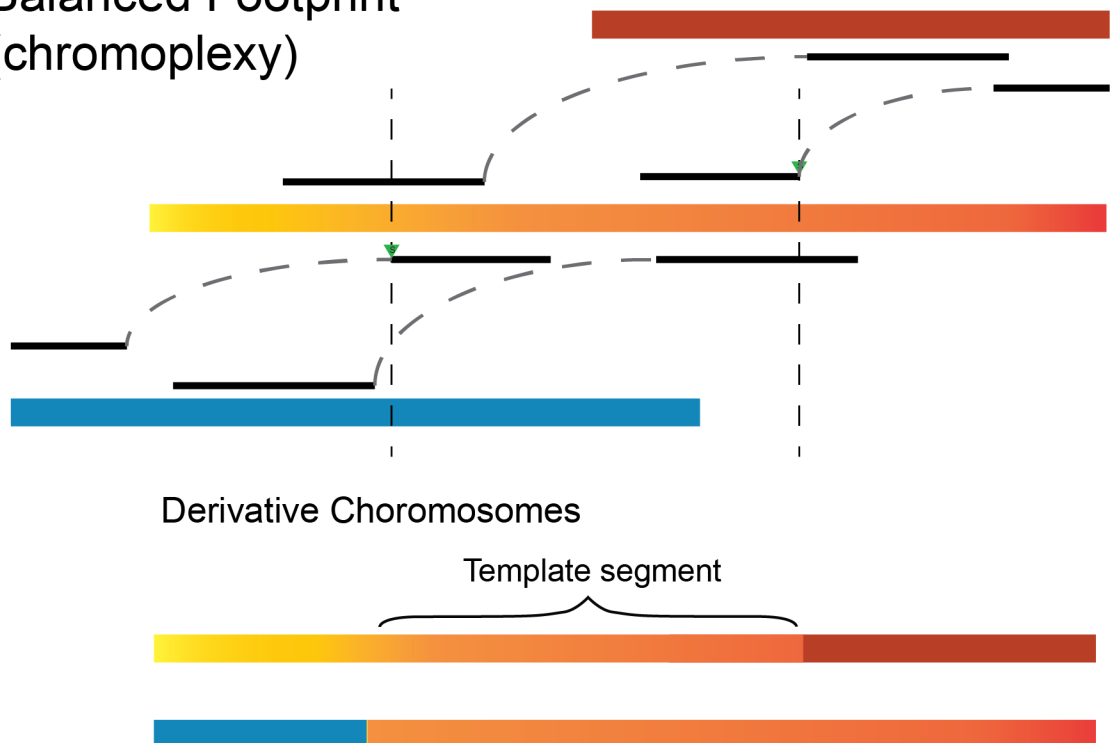


Figure 52 Schematic of inverted orientation balanced breakpoint footprints.

Symbols are all as in Figure 51. The absence of clipping for all discordant reads mapping at the second breakpoint and alignment continuing beyond, are the distinguishing features.

I scrutinised all templated insertion footprints identified in PCAWG for read evidence suggesting that they are true balanced footprints. The search was performed in two phases: footprints of <100bp and footprints 100-500bp). I limited the search to footprints of 500bp because the insert size of most paired-end reads rarely exceeds this. The first phase search, with footprints <100bp, is within the range of a single sequencing read. Templated insertions can therefore be detected when there is evidence of clipping at both ends of an aligned read and this approach had been incorporated by the PCAWG structural variant working group (Li *et al.*, 2020). Clipping evidence suggested 932/1184 (78.7%) of footprints in this range in fact represented balanced breakpoints while the rest were true templated insertions. At a footprint size of 100-500bp, 1605/3351 (47.9%) had insufficient evidence to support either outcome, 225/1746 (12.9%) footprints could be re-classified as balanced. The median separation distance of reclassified footprints, in this range, was 185 but the maximum was 491,

suggesting that this inverted orientation could exist with potentially even longer distances.

5.3.7. Chromoplexy involves recurrent disease specific driver genes

All chains of events involving two or more balanced footprints were therefore considered as chromoplexy events. These were considerably less common than ChainFinder chains (**Figure 53**). Prostate cancer remains one of the cancers with the highest prevalence of chromoplexy. Chromoplexy chains in prostate cancer were also longer, 63/144 (43.8%) of events possessing more than two footprints.

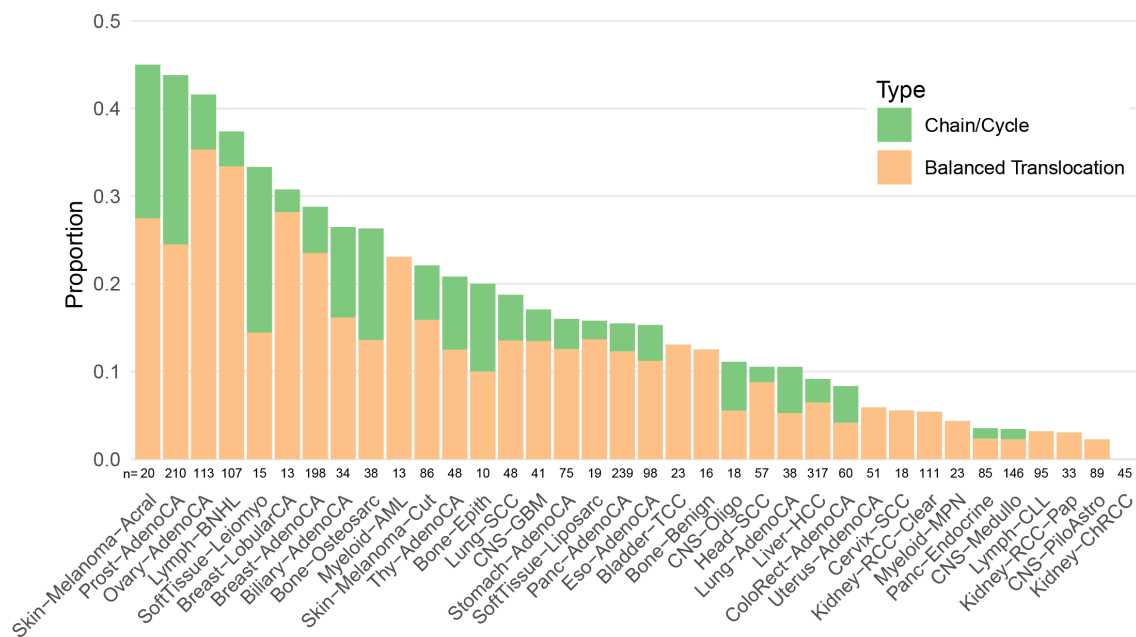


Figure 53 PCAWG Chromoplexy proportions by disease type. Stacked barplot by disease type, segregated by type of chromoplexy chain.

In order to identify recurrent sites for chromoplexy breaks, that could represent either sites prone to recurrent DNA breakage, or selected driver events, I clustered chromoplexy footprints (**Figure 54**). As expected this revealed the reciprocal translocations between the immunoglobulin heavy chain, *IGH*, and *BCL2* (t(14q:18q), dark green), in follicular lymphoma and *RUNX1-RUNX1* (t(8q:21q), maroon) in acute myeloid leukaemia (Weiss *et al.*, 1987, Tighe *et al.*, 1993). The prostate cancer rearrangements involving *TMPRSS2-ERG* (light-blue) were also seen.

The striking and unexpected finding was the number of recurrent chromoplexy footprints in differentiated thyroid cancer (purple). Thyroid cancer chromoplexy chains involved a number of known thyroid cancer fusion genes, including *THADA*, *IGF2BP3* and *BRAF* (Ross *et al.*, 2016), but rearrangements spanning several of these sites simultaneously had not been previously reported.

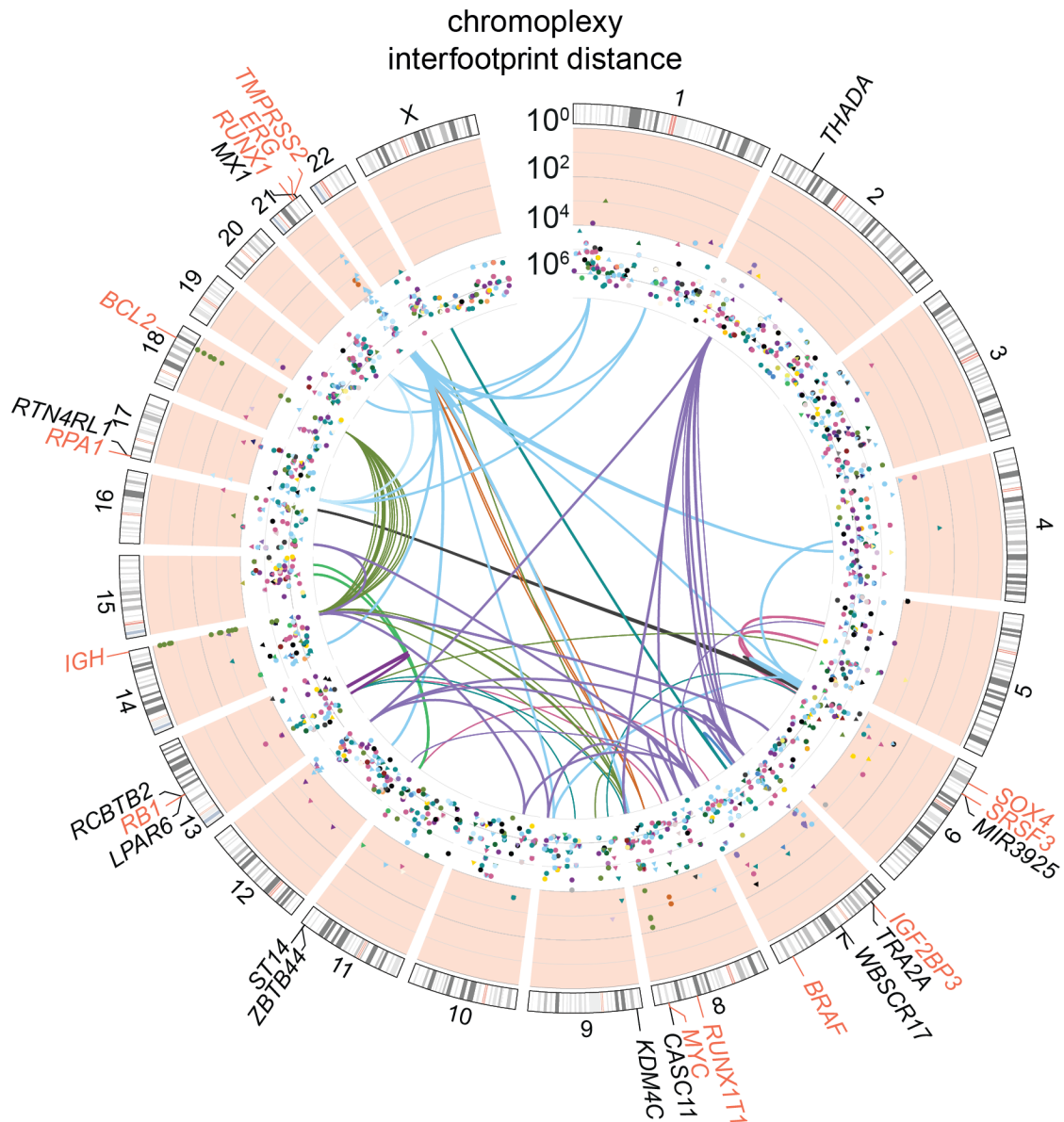


Figure 54 Chromoplexy footprint drivers. The circos scatter plot reflects the distance between each chromoplexy footprint. Clusters of footprints with three or more footprints with an inter-footprint distance of $< 10\text{kb}$ are annotated with genes at that close (or within 30kb) and with structural variants as inner links. Link and point colours match diseases shown in Figure 45.

One particularly remarkable chain, involving four double-stranded DNA breaks, brings *IGF2BP3* (Insulin Like Growth Factor 2 mRNA Binding Protein 3) into the proximity of the *THADA* (Thyroid Adenoma Associated) promoter, truncates *PRKCE* (Protein Receptor Kinase C Epsilon) and disrupts *TG* (thyroglobulin **Figure 55**). *IGF2BP3* enhancer high jacking by *THADA* is a known driver event in thyroid cancer (Panebianco *et al.*, 2017). The truncation of *PRKCE* is also previously reported in thyroid cancer (Knauf *et al.*, 1999). Thyroglobulin is a highly relevant gene in thyroid hormone production and also contains a microRNA known to be downregulated in papillary thyroid cancer (Kolanowska *et al.*, 2017). In total, 10/48 samples contained at least one chromoplexy chain and 4/13 (31%) driver fusion events reported by the PCAWG driver

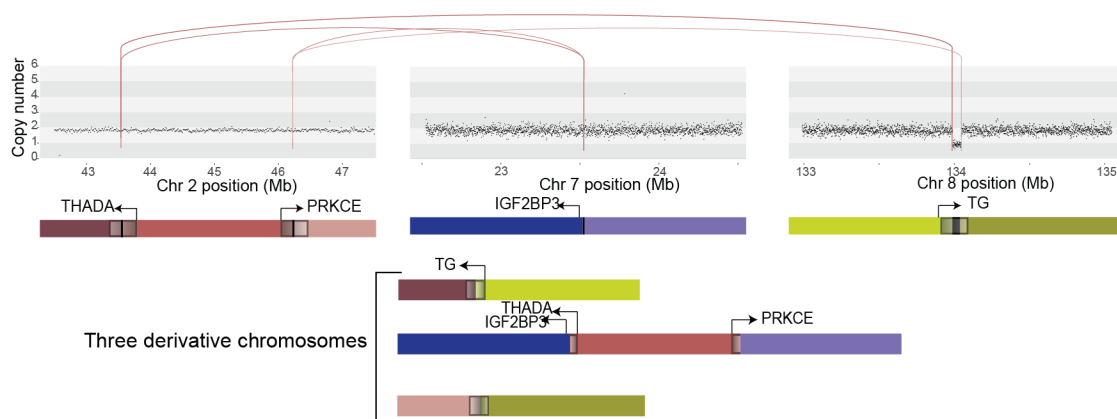


Figure 55 Thyroid cancer chromoplexy chain. Linear schematic of a chromoplexy chain in a follicular thyroid adenocarcinoma. Links above are structural variants. Copy number is generated manually in 1kb bins, normalised by normal coverage and transformed using purity and ploidy from formal copy number methods. Schematics of reference and derivative chromosomes are shown below.

5.3.8. Differentiated thyroid cancer driver rearrangements are predominantly produced by complex rearrangements and chromoplexy

Considering the number of chromoplexy events seen in thyroid cancer, the remaining breakpoints in all PCAWG thyroid cancer samples (n=48) were scrutinised. This revealed a number of foci with highly clustered breakpoints (**Figure 56**), many of which were in genes implicated in thyroid cancer or thyroid function including: *ALK* (Chou *et al.*, 2015), *TDO2* (Finn *et al.*, 2007), *DIO2* (Casula and Bianco, 2012), *BRAF* (Ciampi *et al.*, 2005, Cohen *et al.*, 2003) and

two clusters of ion channel genes *CLCA1/2* and *CACNA1C*. The well recognised *RET-CCDC6* rearrangement were also often part of complex rearrangements (Celestino *et al.*, 2012). Many of the rearrangement clusters were classified by PCAWG as complex events, however 25/48 (52%) of their constituent footprints were of balanced/chromoplexy type. Complex rearrangements, involving thyroid cancer specific drivers appears to be a feature of thyroid cancer, though distinct to previously reported chromoplexy events because they do not involve ETS transcription factor genes.

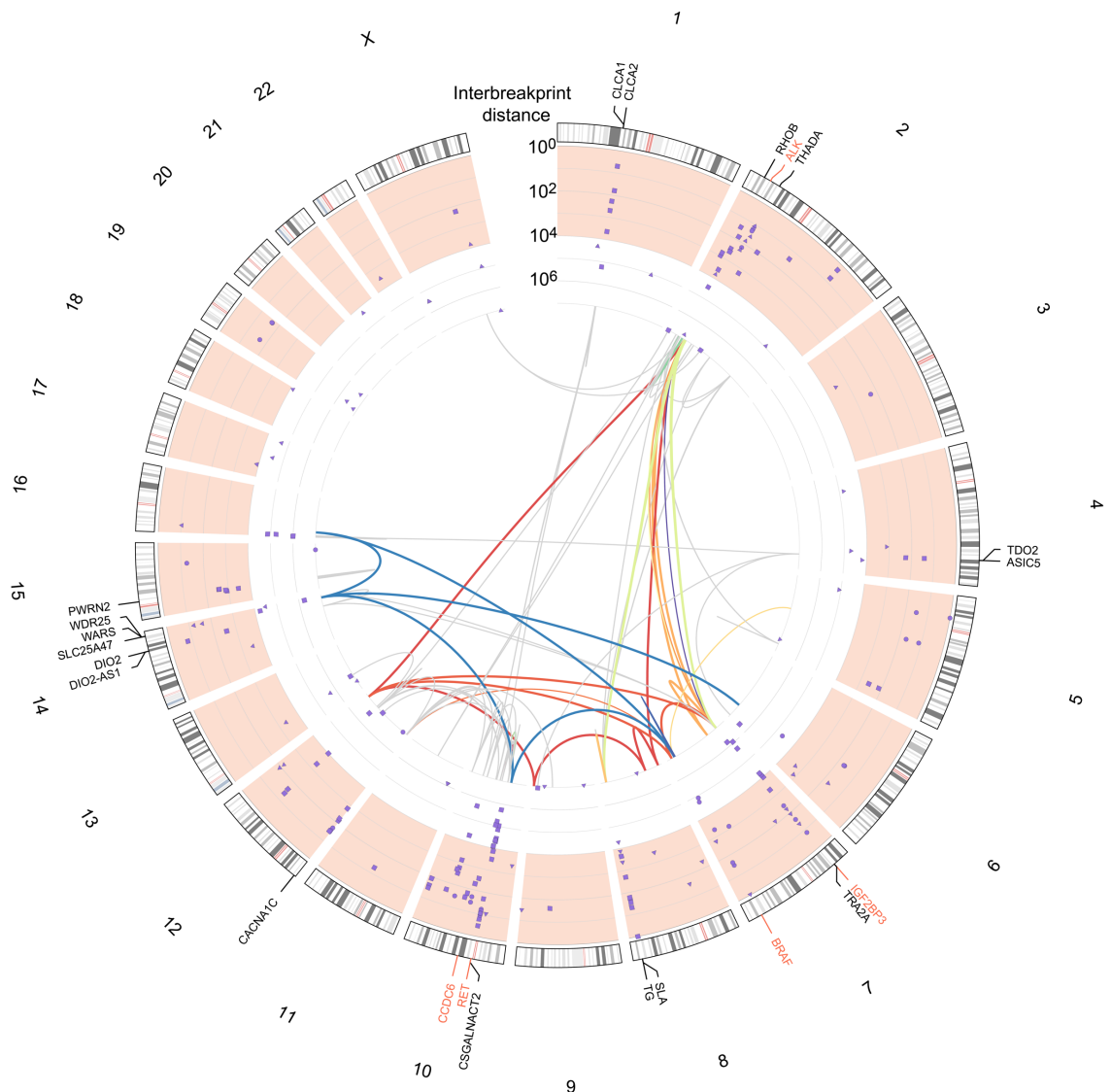


Figure 56 Circos plot of thyroid cancer breakpoints. The interbreakpoint distance for all thyroid cancers (48 samples). Genes within 30kb of clusters of breakpoints are annotated including known drivers (orange). Complex unclear (grey) and chromoplexy structural rearrangements are shown. Each chromoplexy event has a distinct colour.

5.3.9. Mutational timing of events suggests chromoplexy is an early and clonal event

In order to assess the evolutionary timing of chromoplexy, the clonality and timing of the underlying mutations was analysed. Mutational assignments were performed by the evolution and heterogeneity working groups of PCAWG (Gerstung *et al.*, 2018, D'Antonio *et al.*, 2018). The principles used to define the clonal nature of mutations has been discussed previously exploring the pattern of metastatic GCT (**Results 4.3.6**). Mutational timing requires the same principles used to designate the clonal order of mutations in malignant H3.3 mutated tumours (**Results 4.3.2**). The aggregation of this information for punctuated mutational events was performed in parallel for the clustered single nucleotide variant events, kataegis, and chromothripsis. Kataegis was detected through work by Jonas Demeulemeester and chromothripsis by Maxime Tarabichi. The timing of these events used a common methodology, devised in collaboration with Jonas Demeulemeester and Maxime Tarabichi.

Punctuated events were compared to simulated events generated by randomly sampling mutations of the same type from the same sample. Considering the small number of chromoplexy events, with relatively few constituent rearrangements, statistical power was limited for all but three cancer types: uterine adenocarcinoma, hepatocellular carcinoma of the liver and prostate cancer. For all three, chromoplexy events had an increased odds ratio of being clonal compared to a background of non-punctuated structural variants. A similar pattern was seen for chromothripsis, with the notable addition that the well-characterised chromothriptic events seen in liposarcoma also had a high odds of being clonal (Garsed *et al.*, 2014). In contrast kataegis was common and was, without exception, more subclonal than background substitutions.

A similar approach was taken to evaluate the relative timing of punctuated events in the clonal history of tumours. In lung squamous cell cancer and glioblastoma chromoplexy events appeared earlier than other non-punctuated structural variants. In all other cancer types, the confidence interval of the timing odds overlapped with 1. In contrast, most kataegis, if clonal, occurred late. The

exception to this was the kataegis associated with chromosome 12 chromothriptic amplifications in liposarcoma and leiomyosarcoma, which occurred early.

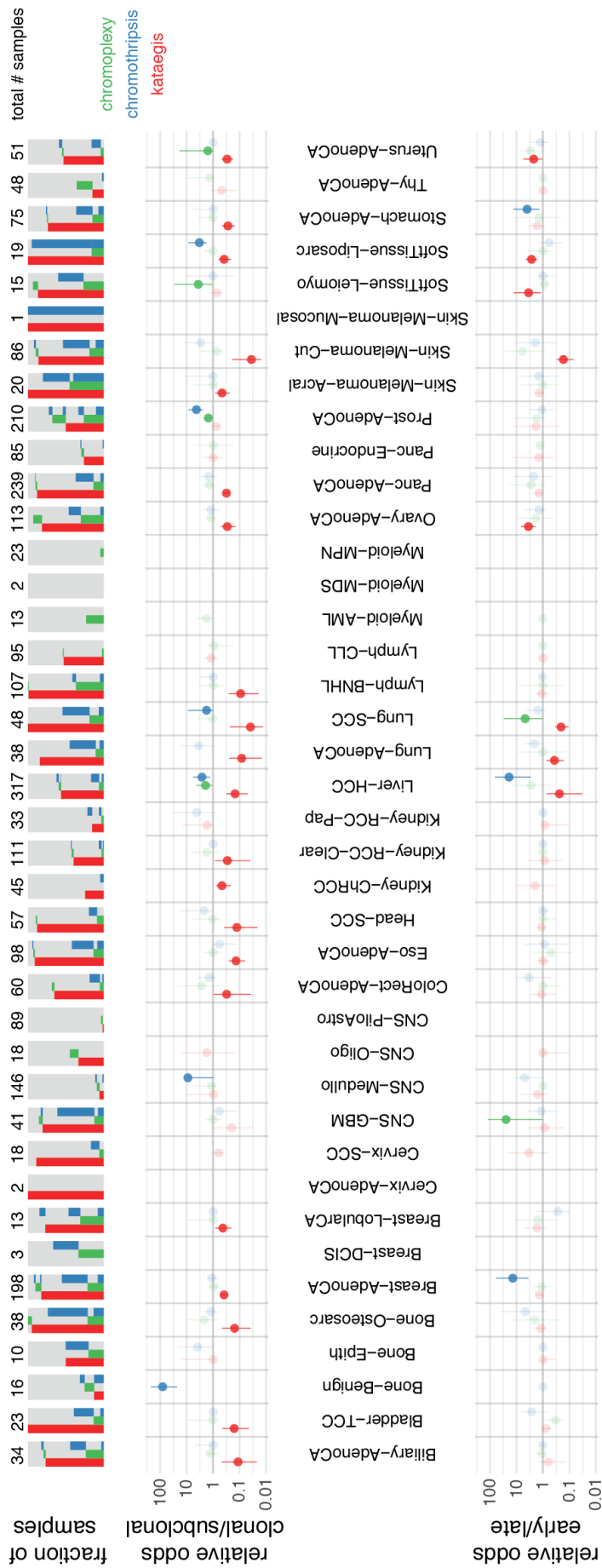


Figure 57 Timing of PCAWG punctuated events. (Top) The overlap of samples is shown in the tileplot. Each sample is shown as a row (size scaled to the total size of the group). (Middle) Odds of being clonal. For each event this is calculated relative to a simulated event comprised of the same number of non-punctuated mutations of the same type (kataegis-substitutions, chromothripsis and chromoplexy-structural variants). (Bottom) Relative odds of being early compared with simulated events.

5.4. Discussion

This analysis provides the first comprehensive review of chromoplexy in a pan-cancer setting. Chromoplexy remains a highly challenging configuration of genomic rearrangements to detect. ChainFinder, the only method relying on high-quality mutation calls, detects many events that poorly resemble the expected pattern for chromoplexy. ChainFinder events involve few chromosomes and structural variants and rarely involve deletions between adjacent breakpoints, even when expected to be detectable. Many of these events can be re-classified as simple rearrangements, retrotransposon or templated insertions, or chromothripsis.

Considerable work is still needed to define chromoplexy in order to more accurately appraise its frequency across cancer types. The overlap and distinction from templated insertion events is not well delineated. They are often difficult to distinguish using short read sequencing data, which provides limited phasing information to reconstruct the true configuration of chromosomes. Templated insertion events are thought to result in templated segments linked directly together in a single 'phased' derivative chromosome. Conversely chromoplexy events result in a series of distinct chromosomal rearrangements. Other sequencing technologies, such as linked read and long read sequencing would aid in distinguishing some of these outcomes. Simpler approaches, such as FISH and karyotyping can also delineate the gross structure of the chromosomes. The historical use of these simpler techniques in defining rearrangement events already provides a body of evidence supporting the occurrence of canonical fusion events through chromoplexy. These include the earliest reports of various well-recognised translocations, which involved multiple chromosomes, not only those containing the canonical fusion genes. These include the *ETV6-NTR3* fusion in infantile fibrosarcoma (Knezevich *et al.*, 1998), in which both initial reported cases involved a third chromosome that was not common to each case. In one early series reporting the *EWRS1-FLI1* fusion in Ewing's sarcoma, 2/4 fusions involved three or more chromosomes (Aurias *et al.*,

1984), consistent with the 42% of rearrangements generated by chromoplexy in the recently published sequencing study (Anderson *et al.*, 2018).

Using a strict definition of the appearance of chromoplexy breaks (footprints) can delineate a small but highly confident set of chromoplexy events. Even with this strict definition, chromoplexy appears common in prostate cancer but also occurs across most other cancer types. The frequency of events in thyroid cancer was unexpected, with a large number of driver events seemingly induced by chromoplexy. Thyroid cancer is unlikely to be exceptional in the frequency of chromoplexy but merely the ideal genomic landscape in which to detect it. The limited number of rearrangements in thyroid cancer mean that chromoplexy events remain unobscured. It is also possible that chromoplexy is more prevalent in thyroid cancer because of the high number of fusion events that involved highly expressed genes. In more complex genomes, rearrangements acquired independently but in genomic proximity to the chromoplexy event will be challenging to distinguish. As this conservative approach struggles to define events with a greater number of rearrangements, it is highly likely that the estimates presented here are also highly conservative. The true prevalence and complexity of chromoplexy is likely to be greater. Biological insights into the underlying mechanism of chromoplexy and genomic sites vulnerable to it will allow easier recognition and appreciation of its true prevalence.

The importance of chromoplexy is reinforced by the evidence presented here that, when assessable, chromoplexy occurred early in clonal evolution. This is consistent with the hypothesis that these events are responsible for critical early driver events in evolution and justifies further study.

Chapter 6. Discussion

This work presents key new findings relevant to the clinical management of rare bone tumours; osteoblastoma and giant cell tumours of bone. In each case the pattern of mutations identified was informative about shared cellular evolutionary origins. The exploration of the complex rearrangement pattern, chromoplexy, also highlighted the role in tumour evolution of mutational patterns that are more challenging to detect. Collectively these insights can suggest specific research directions both for these individual fields but also more generally for the study of tumour evolution.

6.1. Key Findings

The genomics of osteoblastoma were explored, for the first time, discovering a disease-defining mutation. Rearrangements in *FOS* or *FOSB* appeared to be ubiquitous in both osteoblastoma and its related counterpart osteoid osteoma. This provided the first biological evidence that osteoblastoma and osteoid osteoma are one disease. It also added them to the list of mesenchymal tumours defined by a simple recurrent mutation, and provided the basis for a clinical diagnostic biomarker (**Table 1**) (Mertens *et al.*, 2009, Amary *et al.*, 2019a).

As part of the AP-1 transcription factor complex, *FOS/FOSB* mutation might be expected to induce the tumour phenotype through change in gene expression. The specific impact on functional elements of *FOS/FOSB* suggests a quantitative change in their expression. This is either achieved through altered promoter activity for *FOSB* or loss of transcript or proteasomal degradation for *FOS*. These hypotheses still require definitive support. The cDNA from these specific fusions provides an ideal starting point for functional evaluation. The dynamics of mutant transcripts and proteins could be evaluated if transfected into suitable cell lines; either osteoblasts, or lineage guided stem cells. Once transfected, the binding of the mutated AP1 transcription complex across the genome could be analysed with chromatin immunoprecipitation sequencing, and the expression profile itself evaluated with cDNA sequencing. These experiments cannot be performed with

the data currently available because of the lack of directly available wildtype controls.

Chapter 4 explored the mutational and DNA methylation changes that distinguished malignant from benign H3.3 mutated bone tumours. Malignant tumours had a greater mutational burden, albeit with some variability, and acquired either replicative immortality or an additional epigenetic modulator. Timing analysis suggested that these malignant tumours had acquired histone mutation early in evolution, supporting the notion that they may have developed from conventional giant cell tumours of bone. Conversely, metastatic giant cell tumours lacked additional copy number changes. For the first time, whole genome sequencing was performed on a sample from the primary tumour and synchronous metastases from the same patient. Unlike malignant tumours, this case lacked additional driver mutations. These data suggest that benign metastatic cases closely resemble conventional giant cell tumours. The seeding pattern of metastasis in this one case also supported polyclonal seeding. This phenomenon has been reported in prostate cancer that was widely disseminated and thought to have involved sequential waves of metastasis but in general, cancer metastases are monoclonal (Gundem *et al.*, 2015, Yates *et al.*, 2017, Priestley *et al.*, 2018). This may support the prevailing theory of passive tumour emboli giving rise to metastases (Alberghini *et al.*, 2010, Fletcher *et al.*, 2013). DNA methylation profiles supported the close relationship of H3.3 mutated bone tumours but benign and malignant tumours could still be distinguished.

The epigenetic landscape of H3.3 mutated bone tumours requires further exploration. The epigenetic impact of the primary G34W *H3F3A* mutation is still not understood, particularly in comparison with the other oncohistone mutations (Harutyunyan *et al.*, 2019, Lu *et al.*, 2016). DNA methylation profiling could provide some insight into this, however appropriate controlled experiments, with and without the mutation, are required. Direct assay of histone methylation and acetylation, with the quantification of the activity of methyltransferases, such as SETD2, is warranted. The downstream effects of the shift in epigenetic landscape is largely unexplored even across the oncohistone mutated paediatric tumours. As an example, focal promoter changes, such as those seen in *CCND1* are likely

to have a functional impact, in a highly cell-type specific manner, and warrant further exploration.

Chapter 6 evaluated a mutational pattern at the pan-cancer scale. Chromoplexy proved challenging to define and detect, and the prevailing algorithm, ChainFinder, was shown to be highly non-specific. Using a classification method of structural variants with a tight definition of the configuration of chromoplexy, I demonstrated that chromoplexy-type rearrangements are pervasive in cancer. In particular, thyroid cancer fusions are commonly generated by chromoplexy. This approach was highly conservative and insensitive and therefore under-reports the true prevalence of chromoplexy. In contrast to kataegis, where assessable, chromoplexy, like chromothripsis was an early evolutionary event, providing evidence of its important impact early in tumorigenesis.

Further work is clearly required to identify these complex events more reliably. Some methodological work could improve their detection in currently available short read sequencing. As an example, when breakpoints are in close proximity, I demonstrated how the ambiguity between templated insertion events and chromoplexy can be clarified (**Results 5.3.6**). In addition, alternative technologies could be applied. Combining sequencing with well-established low-resolution technologies, such as FISH and karyotyping could better delineate the structure of rearranged derivative chromosomes. Linked read and long read sequencing also might aid in phasing rearrangements to one another over longer distances. These sequencing technologies may improve the sensitivity of structural rearrangement calling, which is inherently lower than for other mutation types (Cameron *et al.*, 2019, Ewing *et al.*, 2015).

6.2. Cross-cutting themes

Simple mutational patterns can uncover shared origins. That the highly pathologically similar diseases, osteoid osteoma and osteoblastoma, both possess *FOS* rearrangement makes their common evolutionary origin likely. Furthermore, that identical *FOS* rearrangements are shared with epithelioid haemangioma suggests they may also have a common origin. Variants of osteoblastoma can have epithelioid appearances and they are frequently highly

vascular (Deyrup and Montag, 2007). The histone mutations, in general, are highly disease specific (**Table 6**). The presence of the same histone mutation in benign and malignant tumour also suggests their common origin, further supported by the mutational analysis presented here. These are however likely to be relative exceptions, limited to disease defining rearrangements and the oncohistones. Most simple driver mutations are much more promiscuous reflecting evolutionary convergence on common mechanisms to generate the hallmarks of cancer. As well-known examples, *TP53*, *KRAS*, and *TERT* are seen across many cancer types (Zehir *et al.*, 2017, Bailey *et al.*, 2018). Mutations in *BRAF*, *KIT*, and *HER-2* might be best recognised and effectively targeted in melanoma, gastro-intestinal stromal tumours and breast cancer respectively but are also similarly mutated in many other cancers (Hyman *et al.*, 2015, Prins *et al.*, 2013). These highly recurrent driver mutations, seen even in normal tissues, might lead some to question the clonal origin theory of tumours. Fortunately, the wealth of other mutational data available from tumours reveal a much larger group of clonal mutations (**Figure 44**), making their clonal origin indisputable. The evolutionary impact of such simple mutations is also intriguing. That *FOS/FOSB* rearrangements have, to date, exclusively been seen in benign tumours suggests that they may preclude malignant progression. The direct malignant counterparts of *FOS* mutated tumours possess their own distinct defining mutations. Compared to the benign epithelioid vascular tumour epithelioid haemangioma which possess *FOS* rearrangements, the intermediate malignancy Epithelioid Haemangi endothelioma (EHE) is defined by *WWTR1* rearrangements (Errani *et al.*, 2011) and the malignant tumour, angiosarcoma, often possess angiogenesis signalling mutations (Behjati *et al.*, 2014). Osteosarcoma, the malignant counterpart to osteoblastoma, have a complex and largely unexplained genomic landscape but often possess *TP53* mutations (Behjati *et al.*, 2017). No other cancer genome, in the PCAWG collection possessed similar *FOS* rearrangements. Similarly, *FOSB* rearrangements are exclusive to the benign or intermediate grade tumours, osteoblastoma or pseudomyogenic haemangi endothelioma and absent from the malignant tumour epithelioid sarcoma (Agaram *et al.*, 2018, Sullivan *et al.*, 2013, Thway *et*

al., 2016). The reason for this evolutionary dead-end may become more apparent when the functional impact of the mutation is explored, as discussed above. Complex mutational patterns, like chromoplexy are also an important and defining feature of tumour evolutionary trajectory. Darwinian theory proposed species evolution as a gradual process, with changes acquired gradually. In tumours, computational approaches, such as those employed here, rely on a relatively even rate of some 'clock-like' mutations. This mutation rate does accelerate during a tumour's lifetime (Yates *et al.*, 2017, Gerstung *et al.*, 2018) but saltatory jumps in the mutational landscape are also increasingly recognised (Markowitz, 2016). Chromoplexy, perhaps, also stands apart from other bursts of chromosomal instability which can be global, like genome duplication (Bielski *et al.*, 2018) and punctuated chromosomal instability (Gao *et al.*, 2016), or focal like chromothripsis. These other events are likely to generate a near random array of mutations, with driver events generated under the influence of positive selection. Conversely, chromoplexy would seem to inherently involve gene targets and almost every component has a functional impact. It seems likely that a transcription dependent mechanism is responsible for this coding-region specificity, though selection almost certainly also has a role. Finally, the disease-type specificity of complex mutational patterns, hints at either a disease specific mechanism underlying their generation or disease specific selection.

6.3. Future work

Beyond the project specific future directions already discussed, some general directions of cancer genomic research present themselves. The study of osteoblastoma likely represents one of the last studies able to reveal a simple mutation in the coding genome that defines the fate of that tumour.

Heritable cellular traits beyond the non-coding genome are relatively poorly understood. The non-coding genome and the epigenome have vital roles in regulating gene expression but require complex integration of differing layers of analyses to disentangle (Ghavi-Helm *et al.*, 2019, Weischenfeldt *et al.*, 2016). Enhancer highjacking events (**Results 4.3.1**) as a simple example, are probably common and underappreciated events, as recently demonstrated for Androgen

Receptor upregulation in castrate resistant prostate cancer (Takeda *et al.*, 2018, Viswanathan *et al.*, 2018).

Some of these non-coding effects may be induced by complex mutational patterns. As explored in Chapter 5 these can be difficult to identify and the aetiology are not well understood. Appreciating their frequency and the processes creating them could unpick some of the early critical stages in tumorigenesis.

Finally, considering tumours in an evolutionary framework has both biological and clinical implications. These give us a window back to the earliest stages in tumour development, allowing us to appreciate the acquisition of the earliest hallmarks of cancer. This has implications for cancer prediction and prevention. More directly we can better elucidate the patterns of disease progression (**Chapter 4**). These are critical for shorter-term improvements in patient outcome, which will come from predicting the prognosis of patients and understanding why current treatments too frequently fail.

Chapter 7. Appendix

7.1. Chapter 3 Appendix

7.1.1. DNA discordant reads supporting *FOS/FOSB* rearrangement

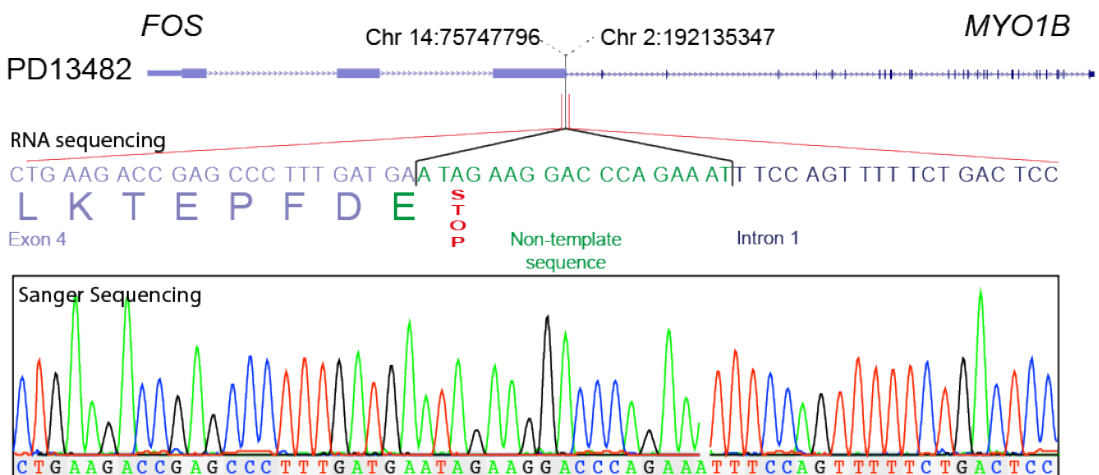
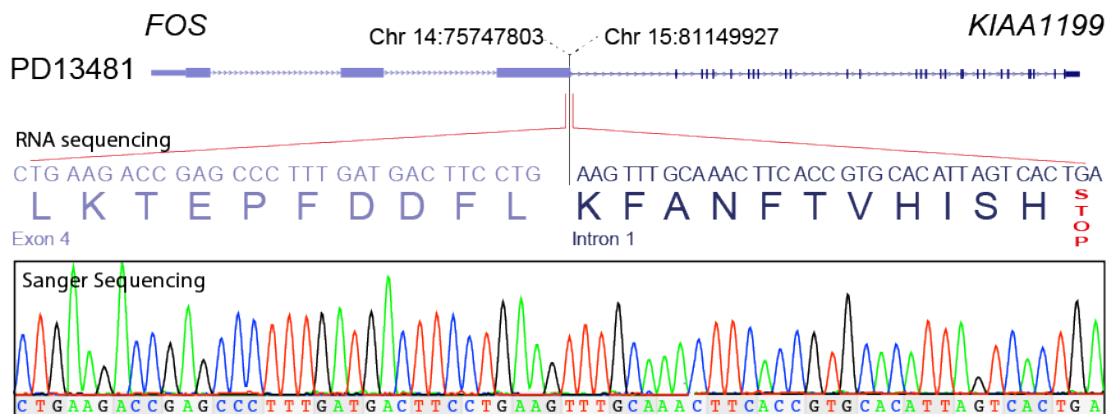
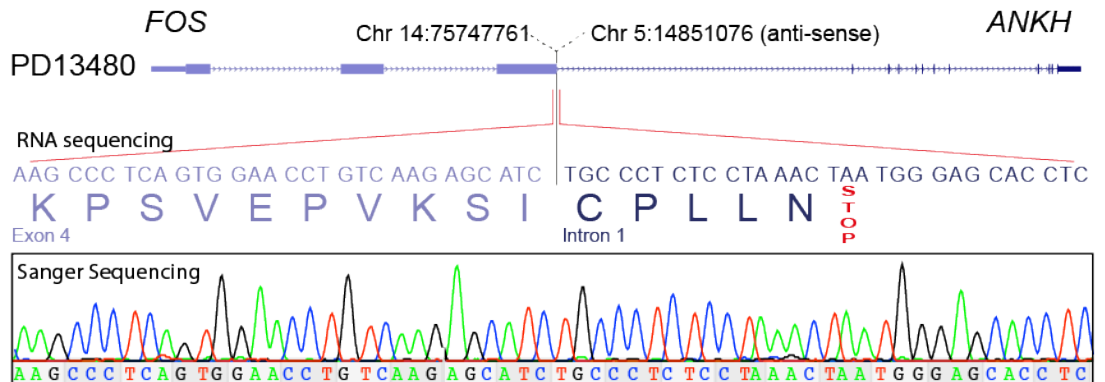
Sample	FLAG	RNAME	POS	MAPQ	CIGAR	RNEXT	PNEXT	TLEN	Chimaeric Mapping
PD13480a	2163	14	75747531	60	70S30M	5	14851416	0	SA:Z:5,14851058,+,28S72M,60,1;
PD13480a	2209	14	75747531	60	64S36M	=	75747919	488	SA:Z:5,14851058,-,34S66M,60,1;
PD13480a	99	14	75747531	60	23S77M	=	75747916	485	
PD13480a	163	14	75747531	60	24S76M	=	75747844	413	
PD13480a	2163	14	75747531	60	70S30M	5	14851416	0	SA:Z:5,14851058,+,28S72M,60,1;
PD13480a	99	14	75747531	60	45S55M	=	75747788	357	SA:Z:5,14851058,-,53S47M,60,1;
PD13480a	163	5	14851058	60	27S73M	=	14851434	476	
PD13480a	99	5	14851058	60	28S72M	=	14851416	458	SA:Z:14,75747531,-,70S30M,60,0;
PD13480a	177	5	14851058	60	34S66M	14	75747919	0	SA:Z:14,75747531,+,64S36M,60,0;
PD13480a	2163	5	14851058	60	53S47M	14	75747788	0	SA:Z:14,75747531,+,45S55M,60,0;
PD13481a	2129	14	75747758	60	46M54S	=	75747323	-481	SA:Z:15,81149925,-,44S56M,60,0;
PD13481a	2209	14	75747828	60	53S47M	=	75748132	404	SA:Z:15,81099370,+,48M52S,60,0;
PD13481a	81	15	81149925	60	44S56M	14	75747323	0	SA:Z:14,75747758,-,46M54S,60,0;
PD13481a	163	15	81149927	60	24S76M	=	81150277	450	
PD13482a	2129	14	75747757	60	40M60S	=	75747412	-385	SA:Z:2,192135346,-,58S42M,60,0;
PD13482a	81	2	192135346	60	58S42M	14	75747412	0	SA:Z:14,75747757,-,40M60S,60,0;
PD7519a	2179	14	75747716	60	36M64S	=	53309368	-22438349	SA:Z:14,53309647,-,64M36S,60,0;
PD7525a	2209	19	45971907	60	62S38M	=	45972098	291	SA:Z:6,30571217,-,41S59M,60,0;
PD7525a	177	6	30571217	60	41S59M	19	45972098	0	SA:Z:19,45971907,+,62S38M,60,0;

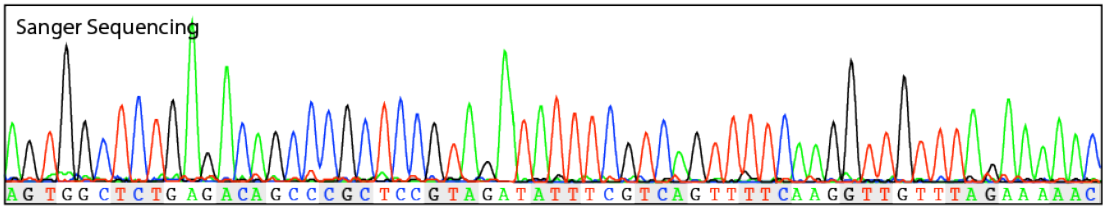
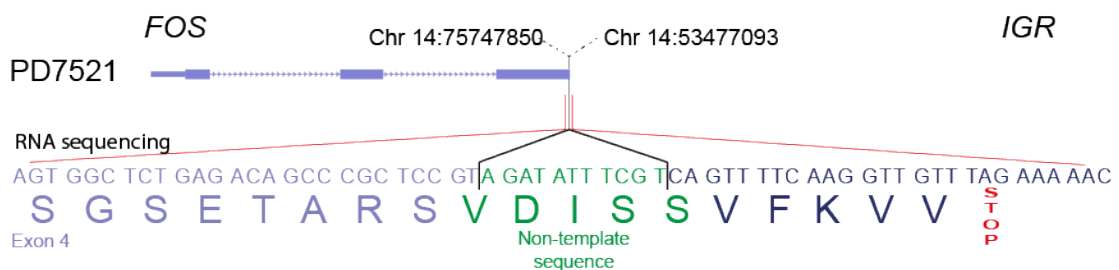
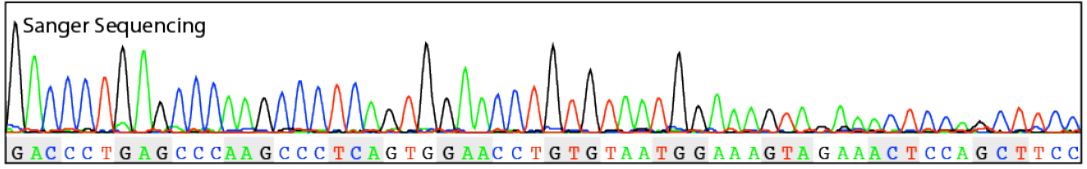
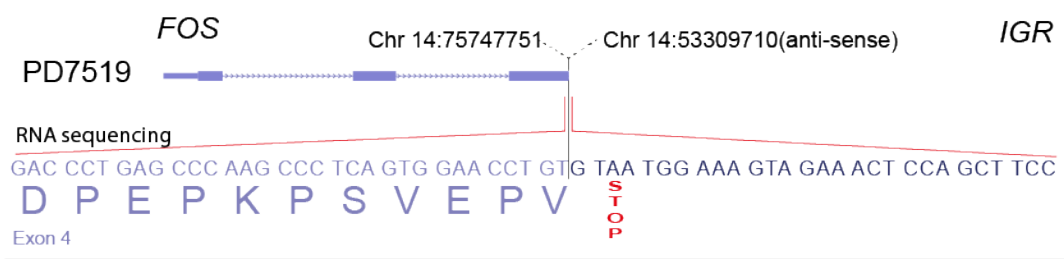
7.1.2. DNA discordant reads supporting FOS/FOSB rearrangement

Sample	FLAG	RNAME	POS	MAPQ	CIGAR	RNEXT	PNEXT	TLEN
PD13480a	129	14	75747435	60	100M	5	14850873	0
PD13480a	129	14	75747531	60	100M	5	14850838	0
PD13480a	129	14	75747544	60	100M	5	14850859	0
PD13480a	65	14	75747544	60	100M	5	14850859	0
PD13480a	177	14	75747589	60	100M	5	14851273	0
PD13480a	177	14	75747589	60	100M	5	14851335	0
PD13480a	65	14	75747669	60	93M7S	5	14850635	0
PD13480a	113	14	75747675	60	100M	5	14851295	0
PD13480a	113	14	75747919	60	100M	5	14851058	0
PD13481a	161	14	75747323	60	85M1D15M	15	81149925	0
PD13481a	161	14	75747433	60	100M	15	81149925	0
PD13481a	161	14	75747511	60	100M	15	81149990	0
PD13481a	1185	14	75747511	60	100M	15	81149990	0
PD13481a	97	14	75747529	60	100M	15	81150021	0
PD13481a	97	14	75747705	60	100M	15	81150033	0
PD13481a	145	14	75747828	60	15S85M	15	81099048	0
PD13481a	81	14	75747837	60	100M	15	81099007	0
PD13481a	145	14	75747846	60	100M	15	81099106	0
PD13481a	145	14	75747875	60	100M	15	81099069	0
PD13481a	145	14	75747937	60	100M	15	81099115	0
PD13481a	145	14	75747943	60	100M	15	81099187	0
PD13481a	81	14	75748111	60	100M	15	81099356	0
PD13481a	81	14	75748132	60	100M	15	81099370	0
PD13481a	81	14	75748153	60	100M	15	81099334	0
PD13482a	161	14	75747412	60	100M	2	192135346	0
PD13482a	97	14	75747421	60	100M	2	192135390	0
PD13482a	81	14	75748017	60	100M	2	192135104	0
PD13482a	145	14	75748181	60	100M	2	192135170	0
PD7519a	129	14	75747374	60	34M1D66M	=	53309571	-22437804
PD7519a	65	14	75747507	60	100M	=	53309572	-22437936
PD7519a	129	14	75747550	60	100M	=	53309438	-22438113
PD7519a	65	14	75747609	60	100M	=	53309571	-22438039
PD7521a	161	14	75747560	60	100M	=	53477208	-22270254
PD7521a	97	14	75747587	60	100M	=	53477175	-22270314
PD7521a	161	14	75747606	60	100M	=	53477156	-22270352
PD7521a	161	14	75747709	60	100M	=	53477164	-22270447
PD7521a	145	14	75747870	60	100M	=	53476670	-22271300
PD7521a	81	14	75747960	60	100M	=	53476886	-22271174
PD7525a	177	19	45971903	60	3M1D97M	6	30571299	0
PD7525a	177	19	45971907	60	9S91M	6	30571328	0
PD7525a	177	19	45971913	60	100M	6	30571388	0
PD7525a	113	19	45971919	60	100M	6	30571551	0
PD7525a	113	19	45971920	60	100M	6	30571261	0
PD7525a	177	19	45971920	60	100M	6	30571261	0
PD7525a	113	19	45971956	60	100M	6	30571336	0
PD7525a	177	19	45971961	60	100M	6	30571286	0
PD7525a	177	19	45971961	60	100M	6	30571306	0
PD7525a	177	19	45971994	60	100M	6	30571246	0
PD7525a	113	19	45972001	60	100M	6	30571224	0
PD7525a	177	19	45972001	60	100M	6	30571217	0
PD7525a	113	19	45972047	60	100M	6	30571367	0
PD7525a	113	19	45972098	60	100M	6	30571217	0
PD7525a	177	19	45972110	60	100M	6	30571294	0

7.1.3. Validation of *FOS* rearrangements

Upper schematics show the cDNA sequence revealed from bulk sequencing reads with the predicted



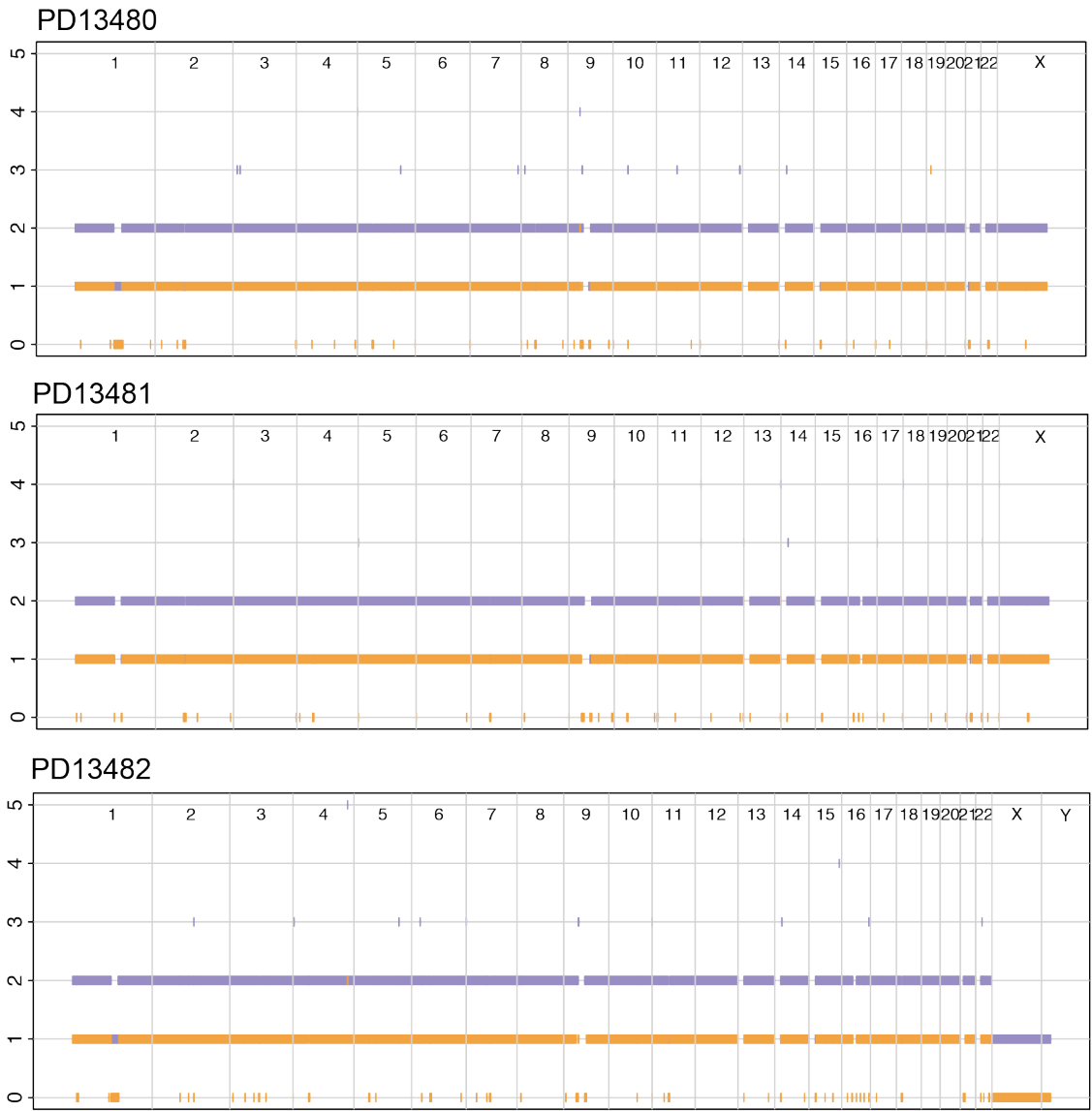


7.1.4. Discovery cohort sequencing results and FOS/FOSB rearrangements

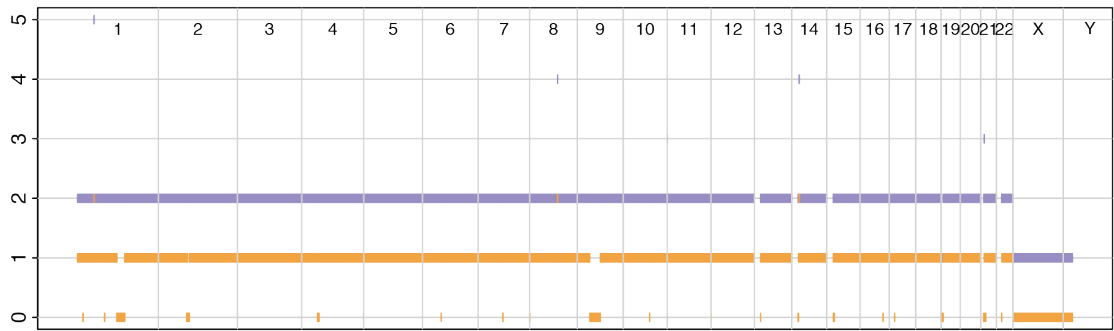
Clinical			Sequencing					Validation					FOS/FOSB Fusion details														
Sample	Diagnosis	Age	Sex	Anatomical Site	Bone Site	Site (mm)	Seq	Subs	Indels	CNAs	SVs	FOS FISH	FOSB FISH	FOSIHC	Chr 1	Post1	Str.1	Gene1	Chr2	Post2	Str.2	Gene2	micro-homology	non template insertion	clean break		
Discovery Cohort																											
PD7519	Osteoblastoma	33	M	Femur, right	Cortical	15	WGS+ RNAseq	700	22	29	12	Positive	Negative	Positive	14	53309710	-	IGR	14	75747751	+	FOS	.	.	.	clean break	
PD7521	Osteoblastoma	25	M	Femur, right	Cortical	40	WGS+ RNAseq	335	14	51	51	Positive	Negative	Positive	14	53477093	+	IGR	14	75747850	+	FOS	.	AGATATTCGT	.		
PD7525	Osteoblastoma	3	M	Humerus, left	Medullary	30	WGS+ RNAseq	555	37	177	41	Negative	Positive	Negative	6	30571219	+	PPP1R10	19	45971907	-	FOSB	.	CTG	.		
PD13480	Osteoblastoma	13	F	Lumbar vertebra, L3	Medullary	35	WGS+ RNAseq	303	25	64	36	Positive	Negative	Positive	5	14851076	-	ANKH	14	75747761	+	FOS	.	.	clean break		
PD13481	Osteoblastoma	11	F	Cervical vertebra, C5	Medullary	30	WGS+ RNAseq	169	50	75	34	Positive	Negative	Positive	14	75747803	+	FOS	15	81149927	+	KIAA1199	TG	.	clean break		
PD13482	Osteoid Osteoma	18	M	Lumbar vertebra, L1	NA	12	WGS+ RNAseq	123	31	70	7	Positive	Negative	Positive	2	192135347	+	MYO1B	14	75747796	+	FOS	.	ATAGAAGGAC CCAGAAAT	.		

7.1.5. Discovery cohort copy number profiles

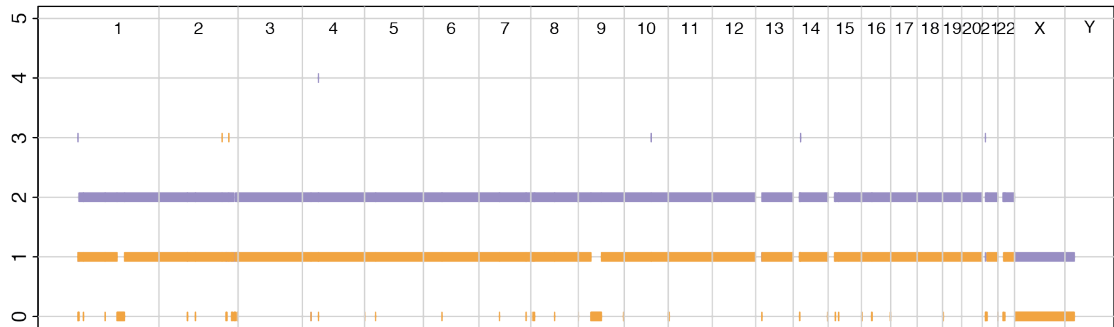
Genome wide views copy number views. Copy number is shown on the y-axis and genomic position on the x-axis. Total copy number is shown as purple segments and the minor copy number state is shown in orange.



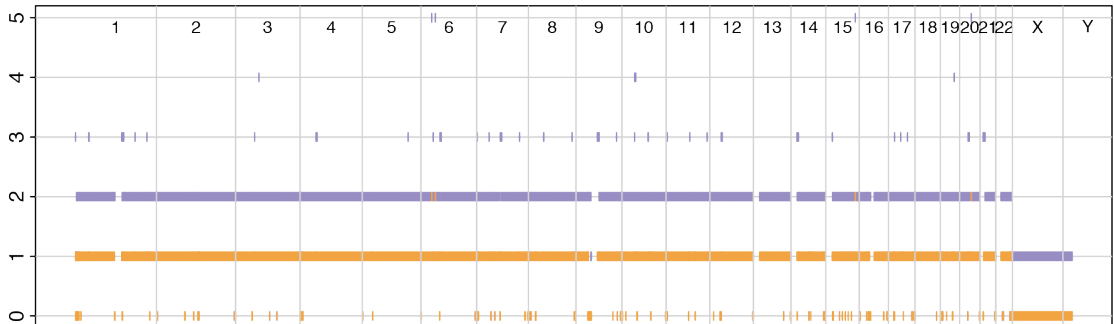
PD7519



PD7521

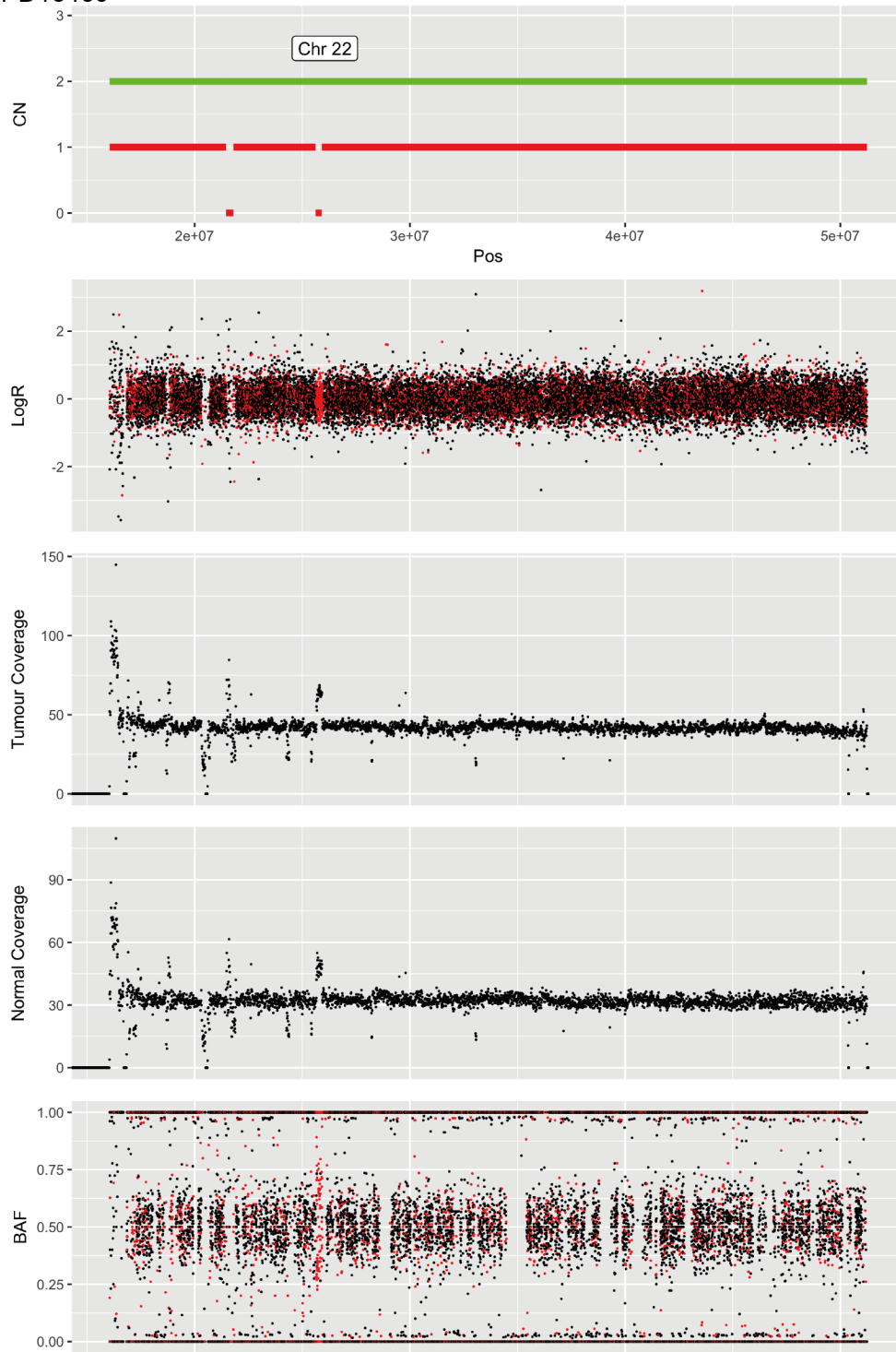


PD7525

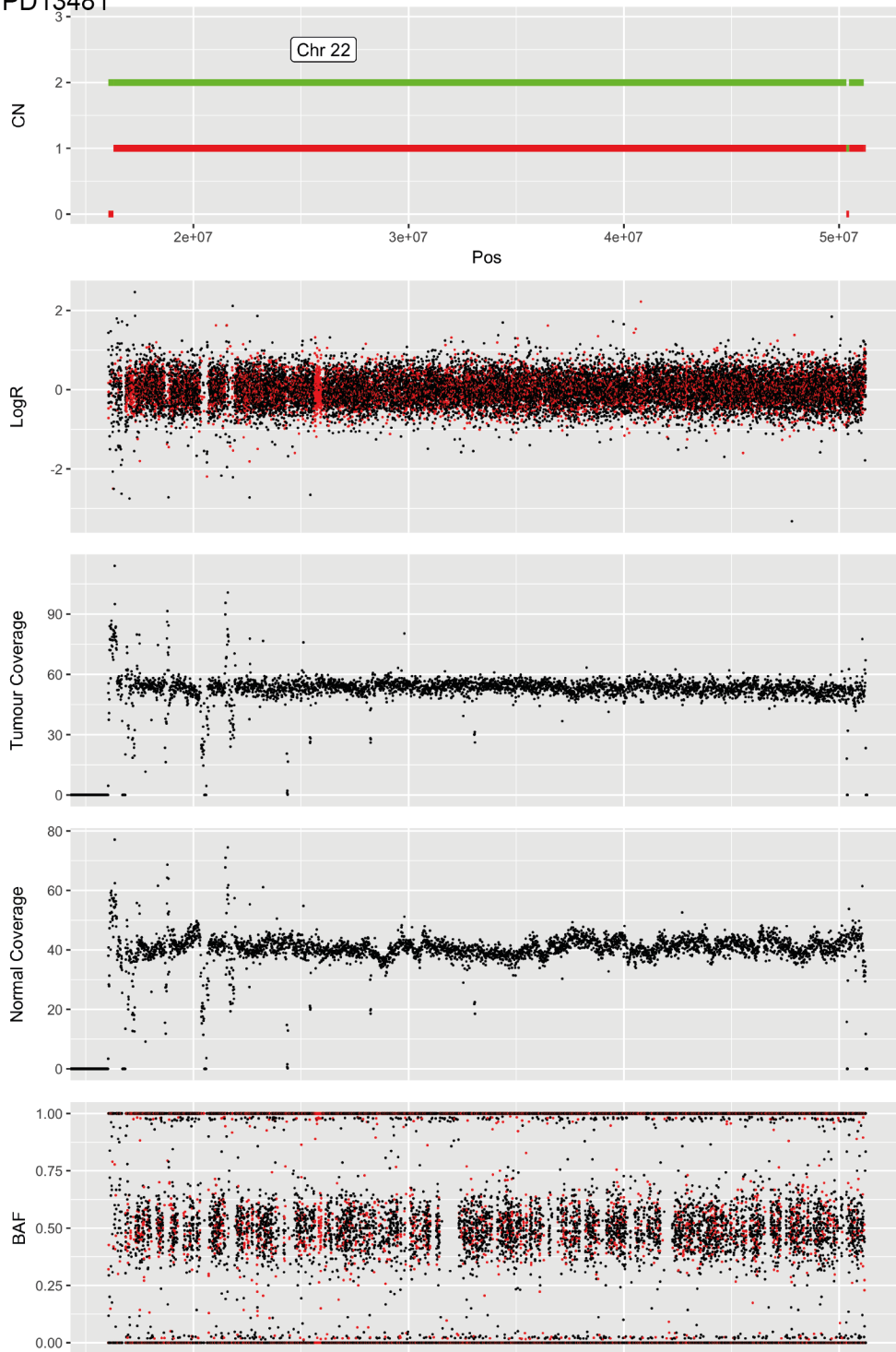


The subsequent pages show the detailed analysis of chromosome 22. Total copy number is shown in green segments, minor allele copy number is in green in the top panel. Sequentially below plots reveal LogR (ratio of tumour to normal coverage as computed by ASCAT), tumour coverage, normal sample coverage, B-Allele Frequency (BAF). In LogR and BAF plots, poorly performing snps are highlighted in red. No reliable copy number aberrations are seen across chromosome 22 in any sample.

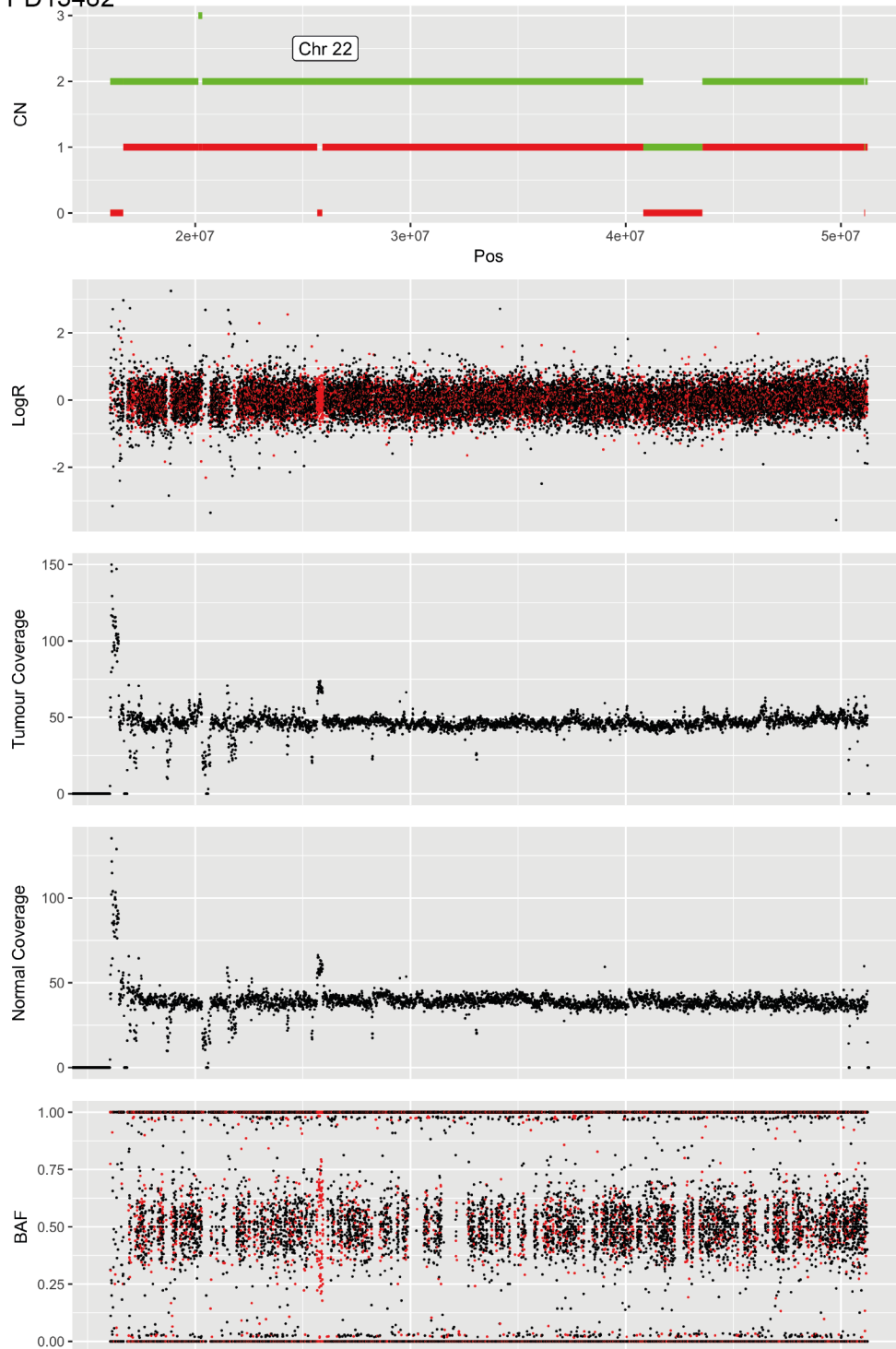
PD13480



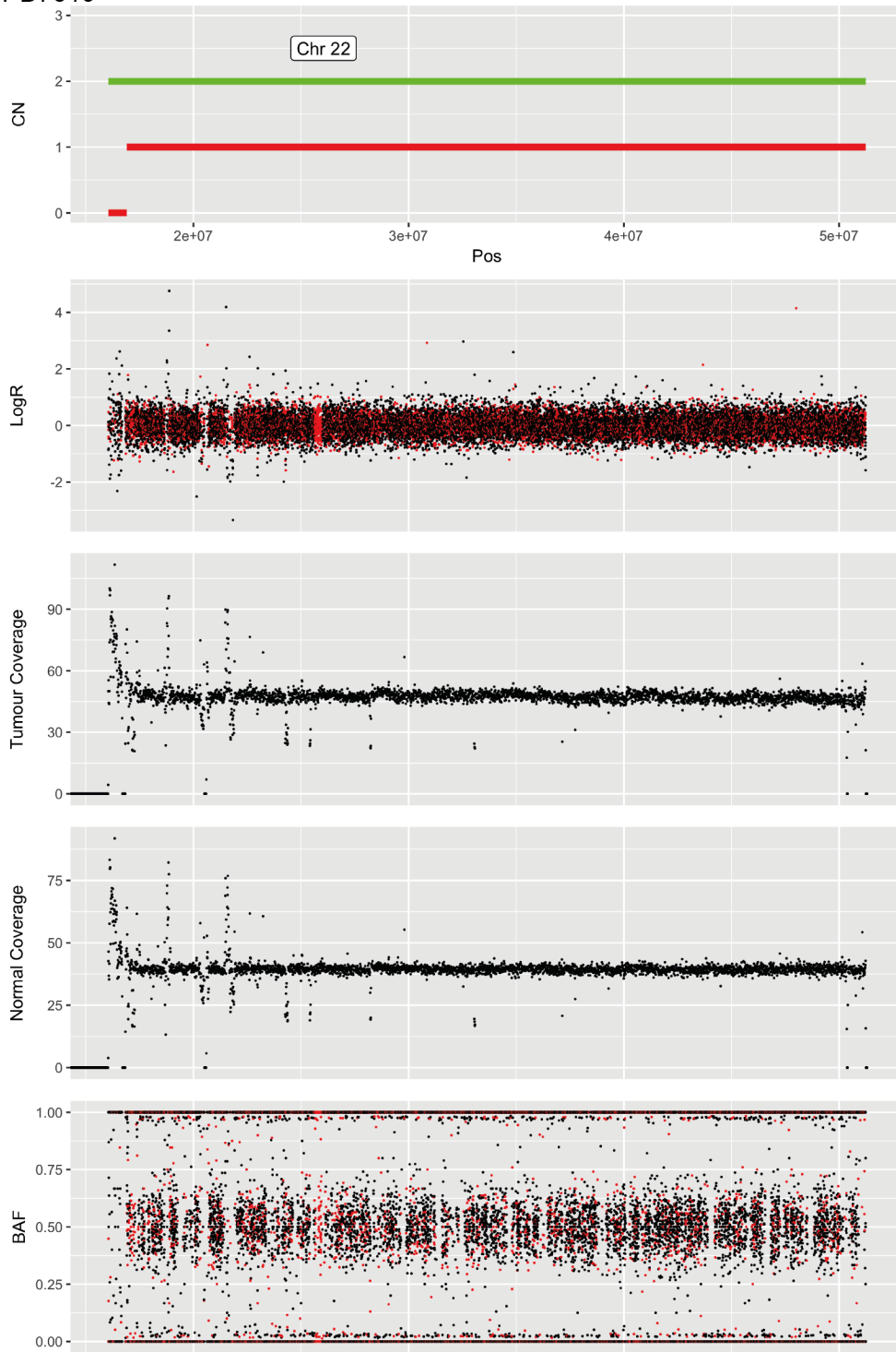
PD13481



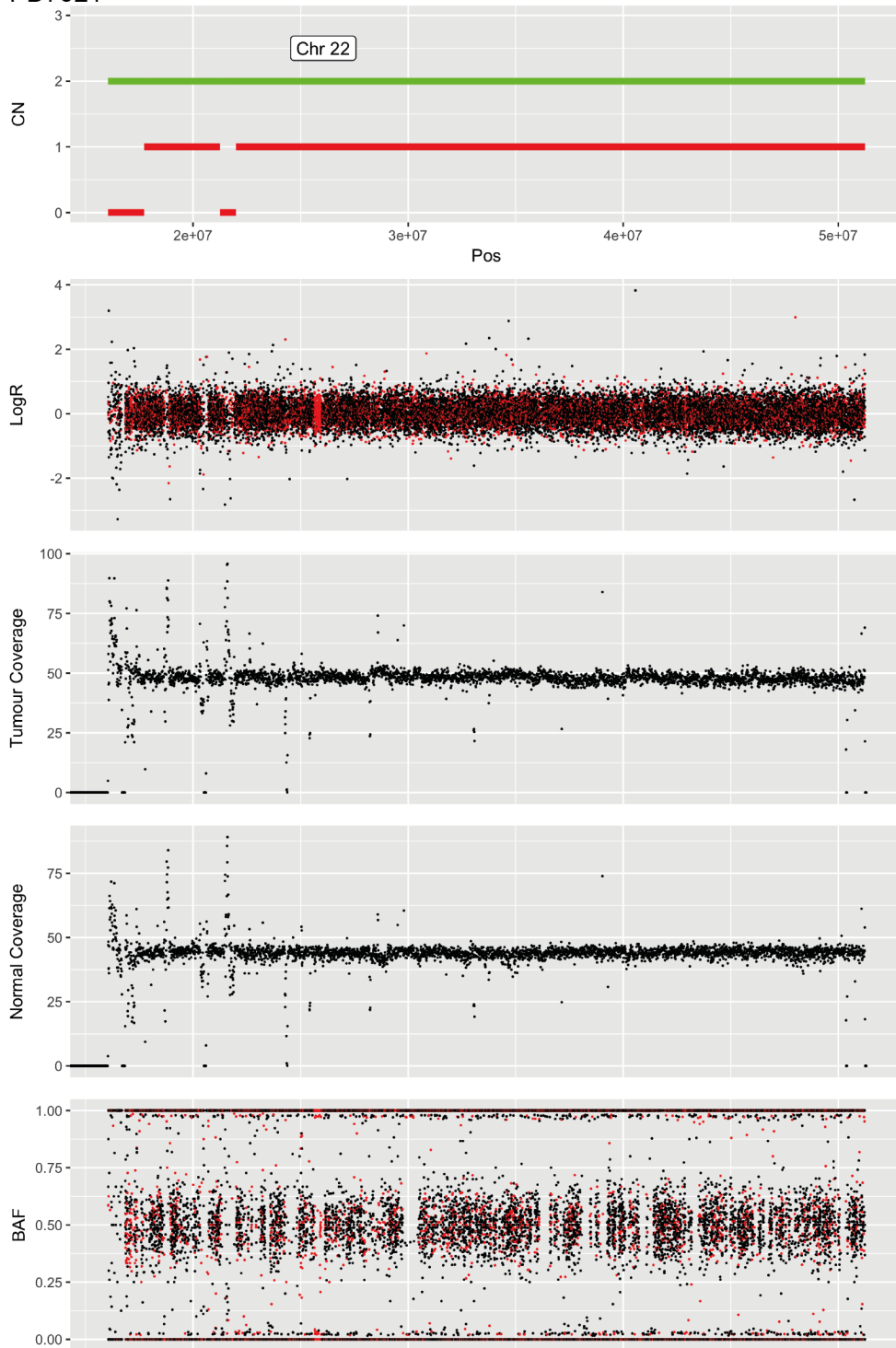
PD13482



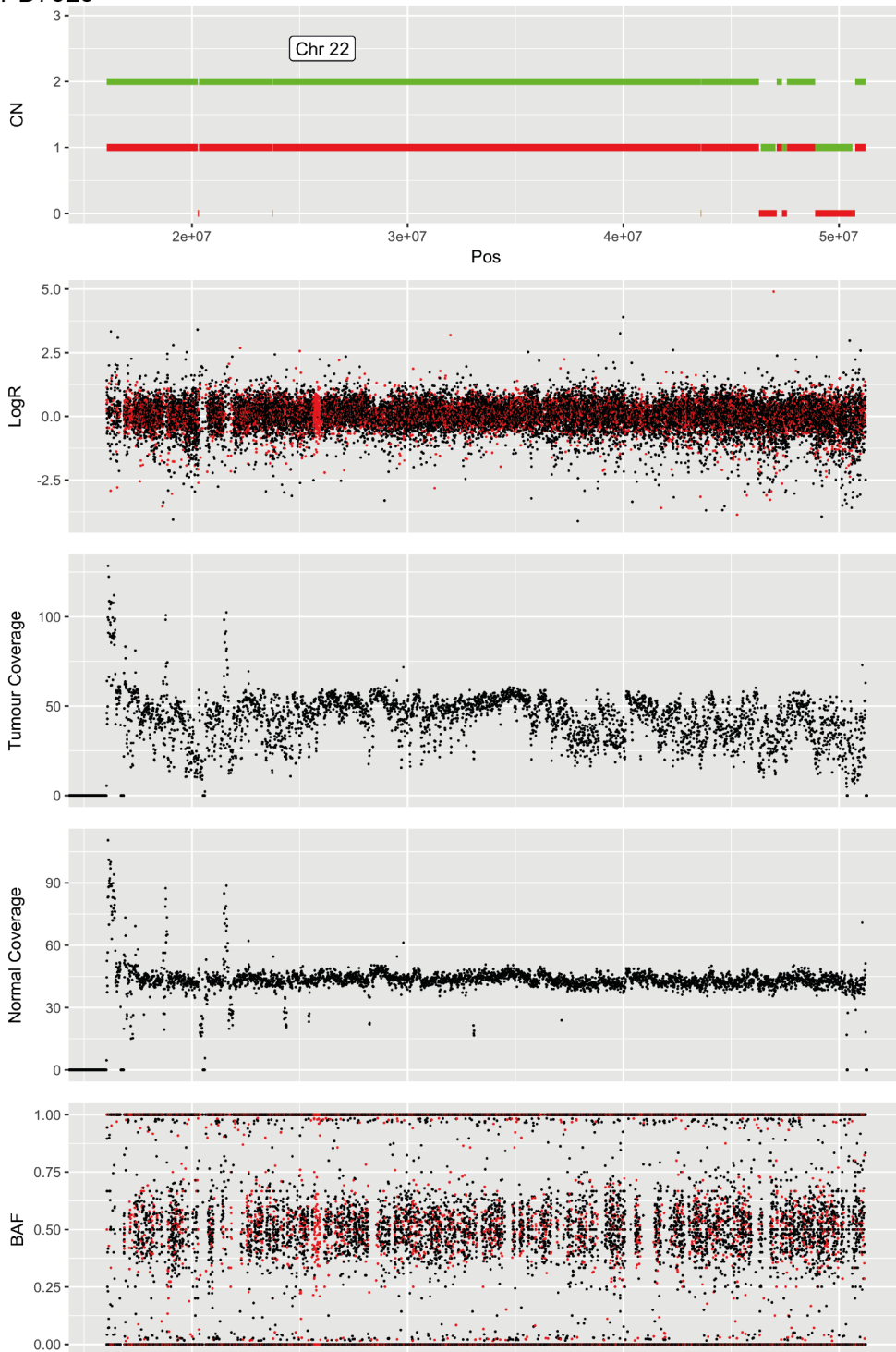
PD7519



PD7521



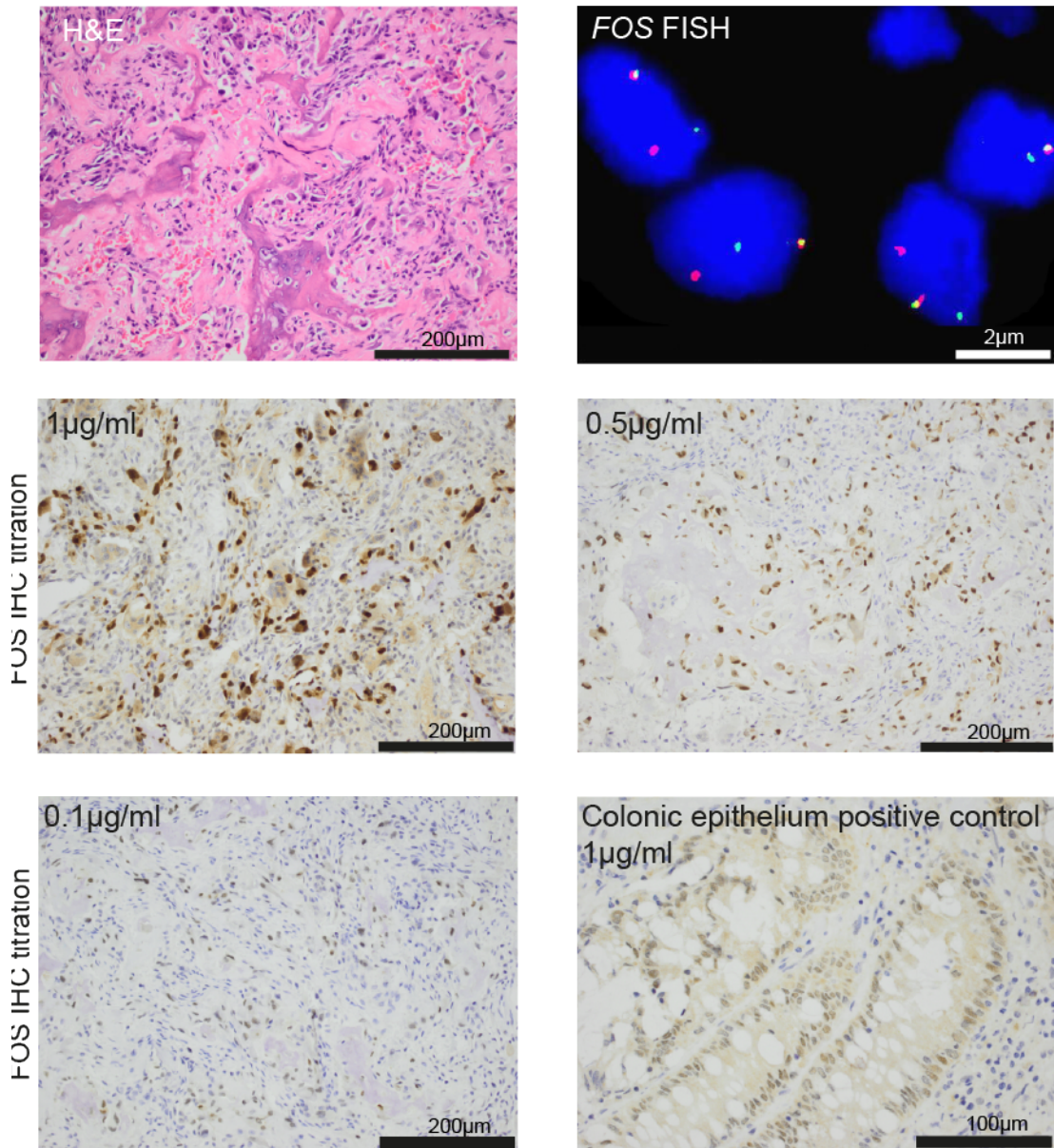
PD7525



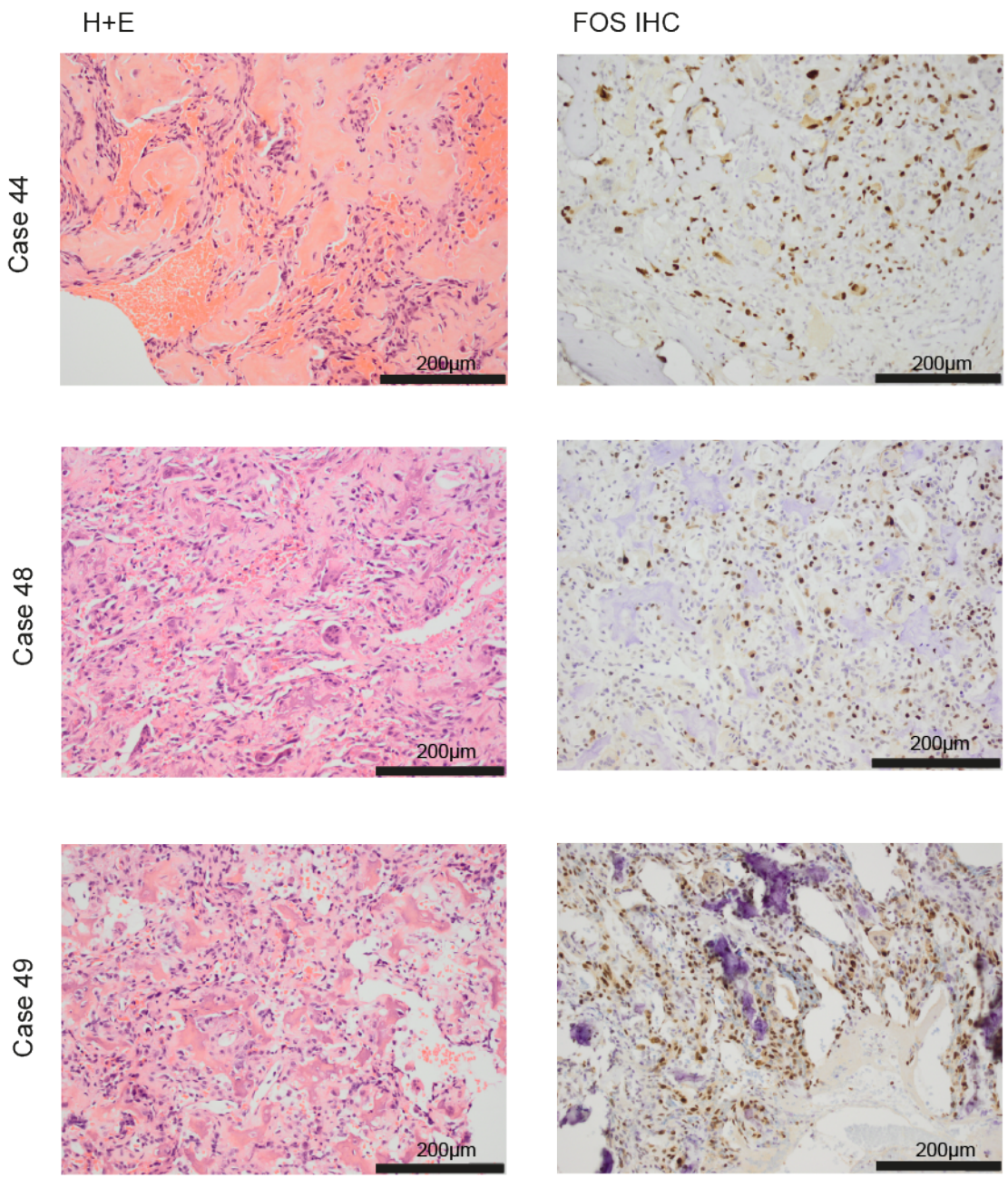
7.1.6. Immunohistochemistry and histology images

Demonstration of stronger nuclear reactivity of the N-terminal FOS antibody than the colonic epithelium positive control, even at the lowest concentration. H&E appearances for this sample are typical for osteoblastoma whilst FISH demonstrates a clear *FOS* break apart.

FOS fusion +ve osteoblastoma

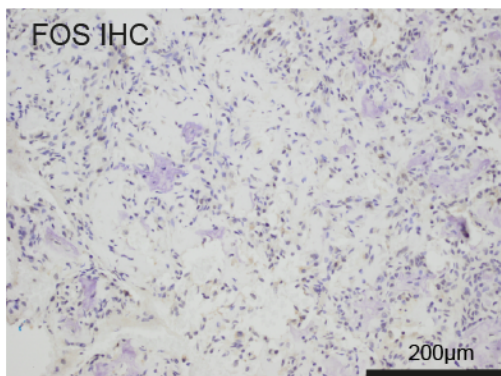
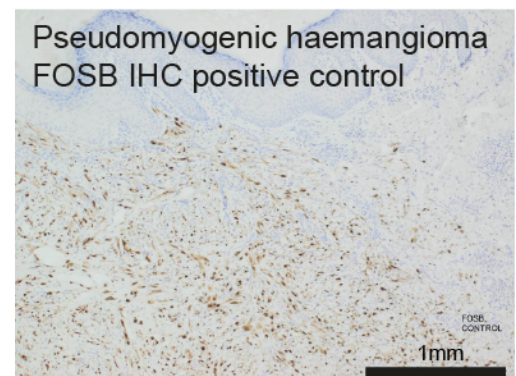
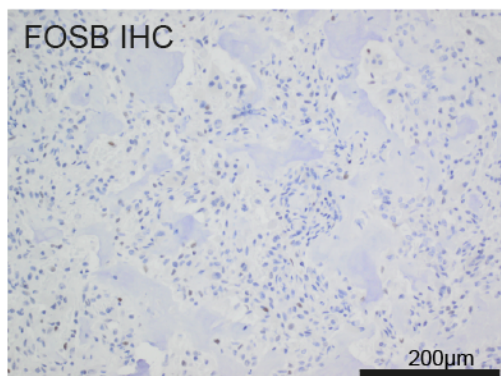
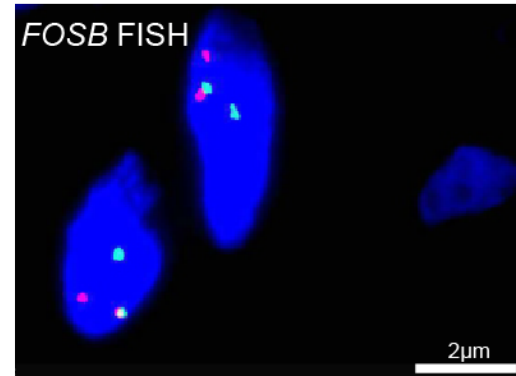
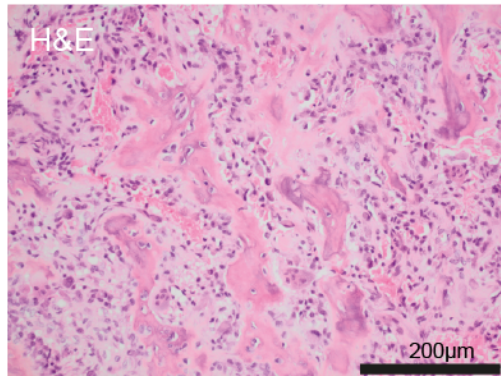


Strong N-terminal FOS immunoreactivity is seen in osteoblastoma cases even for which breakpoint signal could not be demonstrated by FISH.



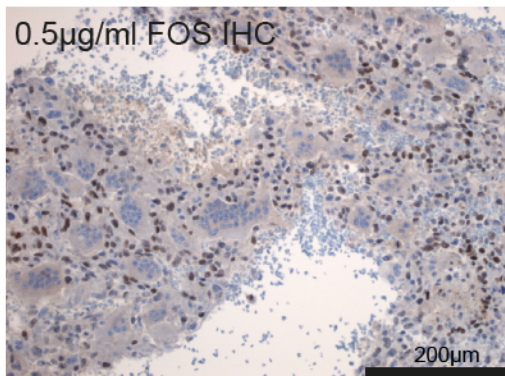
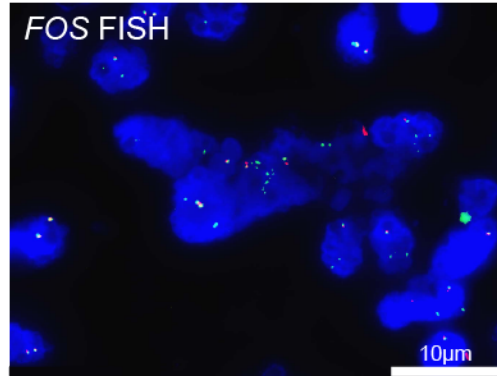
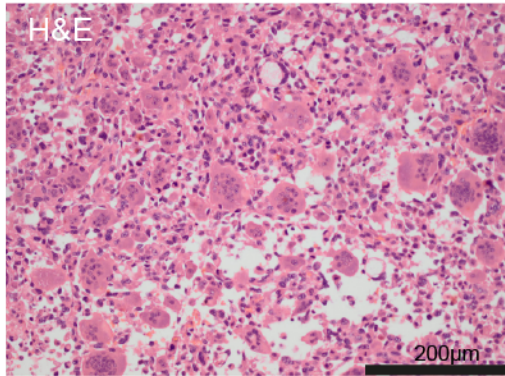
FOS or FOSB immunostaining was not seen in PD7525 likely owing to decalcification. PD7525 has a proven *FOSB* breakapart by DNA and RNA sequencing and FISH. FOSB antibody stains a pseudomyogenic haemangioma positive control well.

FOSB fusion +ve osteoblastoma



The single osteosarcoma sample that demonstrated strong FOS immunoreactivity, demonstrated a distinct histological pattern and evidence of no *FOS* breakapart, though there is evidence of an amplification near the locus of *FOS*.

Osteosarcoma



7.1.7. Validation cohort immunohistochemistry and FISH results

FISH results denote the presence of a breakapart signal in the respective gene.

*focally positive

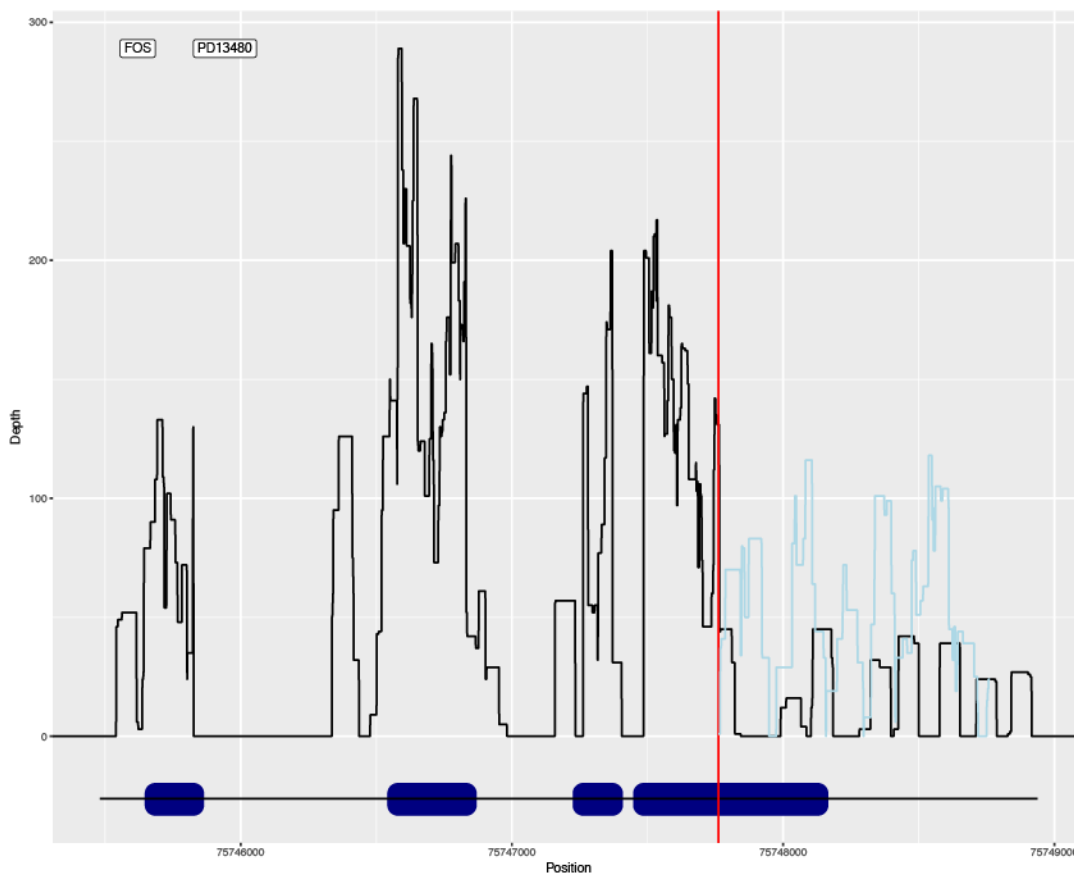
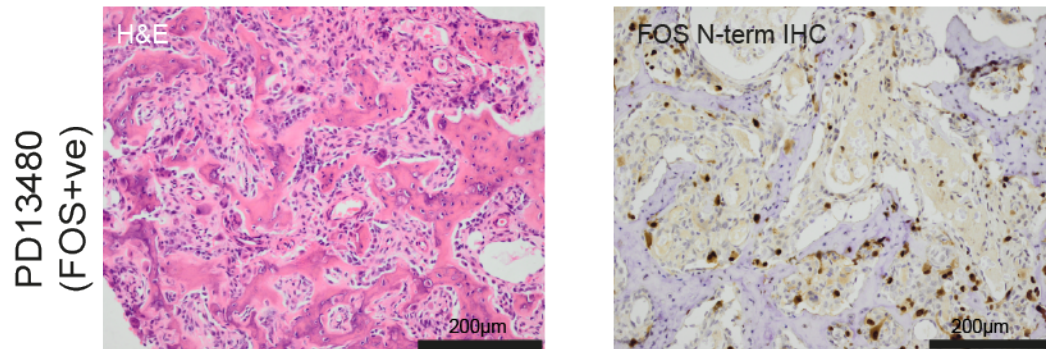
Sample	Diagnosis	Age	Sex	Anatomical Site	Size (mm)	FOS FISH	FOSB FISH	FOS IHC
1	Osteblastoma	13	M	Talus	NA	Positive	Negative	Positive
2	Osteblastoma	10	M	Rib	NA	Positive	Not done	Positive
3	Osteblastoma	18	NA	Lumbar vertebra, L5	15	Positive	Negative	Positive
4	Osteblastoma	15	M	Humerus	NA	Positive	Not done	Not done
5	Osteoid osteoma / osteblastoma	3	M	Humerus, left	15	Positive	Negative	Positive
6	Osteblastoma	12	F	Humerus, left	35	Positive	Negative	Positive
7	Osteblastoma	17	M	Tibia, left	30	Positive	Negative	Positive
8	Osteblastoma	39	M	Pelvis	NA	Positive	Not done	Not done
9	Osteblastoma	15	M	Tibia, right	3	Positive	Negative	Positive
10	Osteblastoma	20	M	Lumbar vertebra, L1	40	Positive	Negative	Positive
11	Osteoid osteoma	4	F	Femur	NA	Positive	Not done	Non-informative
12	Osteblastoma	28	M	Lumbar vertebra, L2	30	Positive	Negative	Not done
13	Osteoid osteoma	52	M	Humerus, left	40	Positive	Negative	Positive
14	Osteoid osteoma	34	M	Vertebra, T7	10	Positive	Negative	Positive
15	Osteblastoma	19	M	Talus, right	14	Positive	Negative	Not done
16	Osteblastoma	18	M	Talus, right	30	Positive	Negative	Negative
17	Osteoid osteoma	19	F	Humerus, right	18	Positive	Negative	Positive
18	Osteblastoma	39	M	Vertebra, C1	9	Positive	Negative	Positive
19	Osteblastoma	18	M	Metatarsal, right, fifth	30	Positive	Negative	Positive
20	Osteblastoma	19	M	Vertebra, T2	40	Positive	Negative	Positive
21	Osteblastoma	20	F	Ankle, right	10	Positive	Negative	Positive
22	Osteblastoma	21	M	Acetabulum, left	30	Positive	Negative	Positive
23	Osteblastoma	28	M	Humerus, left	15	Positive	Negative	Positive
24	Osteoid osteoma	26	M	Phalanx, right middle finger	5	Positive	Negative	Not done
25	Osteblastoma	26	M	Acetabulum, left	50	Positive	Not done	Positive
26	Osteblastoma	10	F	Vertebra, C5	60	Positive	Not done	Positive
27	Osteblastoma	22	M	Femur, left distal	15	Positive	Not done	Positive

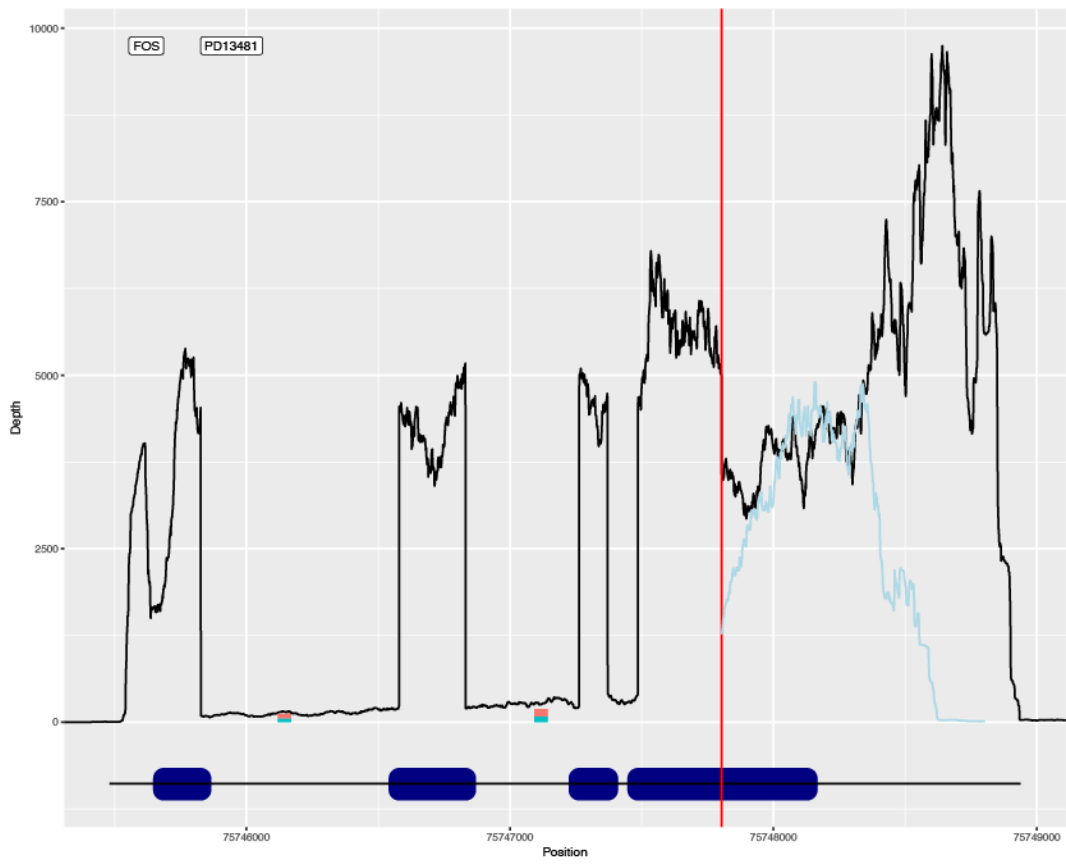
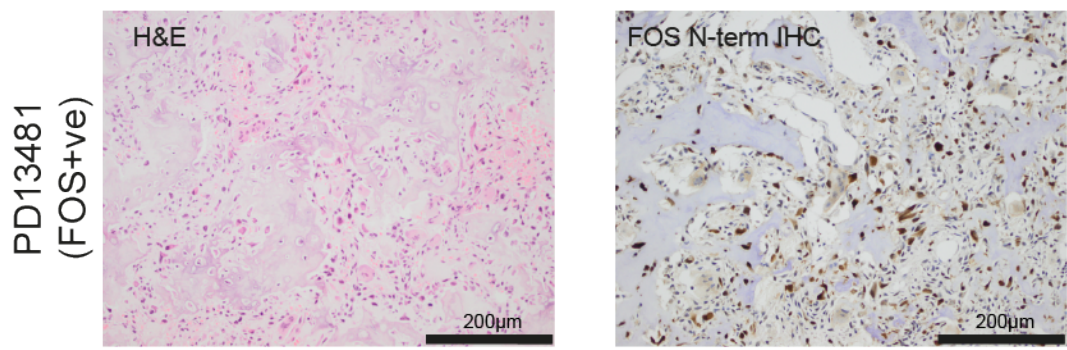
Sample	Diagnosis	Age	Sex	Anatomical Site	Size (mm)	FOS FISH	FOSB FISH	FOS IHC
28	Osteoid osteoma	32	M	Vertebra, C2	15	Positive	Not done	Positive
29	Osteoid osteoma	19	M	Vertebra, L3	15	Positive	Not done	Positive
30	Osteoid osteoma	32	F	Phalanx, left third tpe, distal	10	Positive	Not done	Positive
31	Osteoid osteoma	17	M	Vertebra, L1	5	Positive	Not done	Positive
32	Osteoid osteoma	16	F	Tibia, right	5	Positive	Not done	Positive
33	Osteoid osteoma	18	M	Vertebra, T2	10	Positive	Not done	Positive
34	Osteoid osteoma	15	M	Talus, left	4	Positive	Not done	Negative
35	Osteoid osteoma	12	F	Phalanx, left index finger	5	Positive	Not done	Positive
36	Osteoid osteoma	15	F	Femur, distal	18	Positive	Not done	Positive
37	Osteoid osteoma	20	F	Phalanx, left distal	8	Positive	Not done	Positive
38	Osteoid osteoma	19	M	Tibia, left distal	3	Positive	Not done	Positive
39	Osteoid osteoma	14	M	Phalanx, second right toe	25	Positive	Not done	Positive
40	Osteoblastoma	13	F	Metatarsal bone, left, fourth	15	Positive	Not done	Positive
41	Osteoid osteoma	14	M	Lumbar vertebra, L3	10	Positive *	Negative	Positive
42	Osteoblastoma	21	M	Vertebra, L1	NA	Positive *	Negative	Positive
43	Osteoblastoma	19	M	Vertebra, L5	NA	Positive *	Negative	Positive
44	Osteoid osteoma	55	M	Vertebra, T9, Left	4	focal amplification	Negative	Positive
45	Osteoid osteoma	9	F	Tibia, right	35	Positive *	Negative	Positive
46	Osteoblastoma	18	M	Metacarpal	8	Negative	Not done	Not done
47	Osteoblastoma	15	F	Sacrum	25	Negative	Not done	Not done
48	Osteoblastoma	9	F	Thoracic vertebra, T12	10	Negative	Negative	Positive
49	Osteoid osteoma	39	M	Phalanx, right toe, distal	20	Negative	Negative	Positive
50	Osteoblastoma	8	M	Lumbar vertebra	30	Negative	Positive	Negative
51	Osteoblastoma	24	F	Lumbar vertebra, L4	NA	Positive	Negative	Not done
52	Osteoblastoma	23	M	Sacrum	NA	Positive	Negative	Not done
53	Osteoid osteoma	36	M	Finger, middle, left	NA	Positive	Negative	Not done
54	Osteoid osteoma	30	M	Finger, middle, left	NA	Negative	Negative	Not done
55	Osteoblastoma	18	M	Ulna	NA	Positive	Negative	Not done

7.1.8. Allele specific expression results

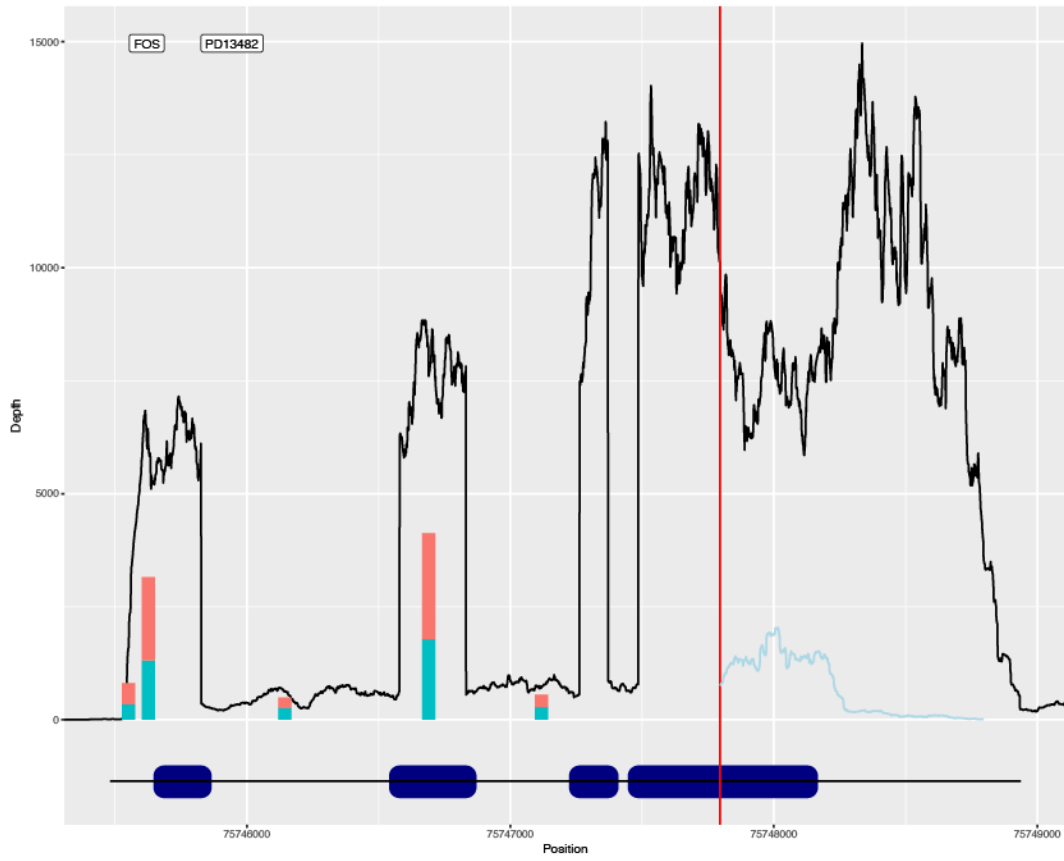
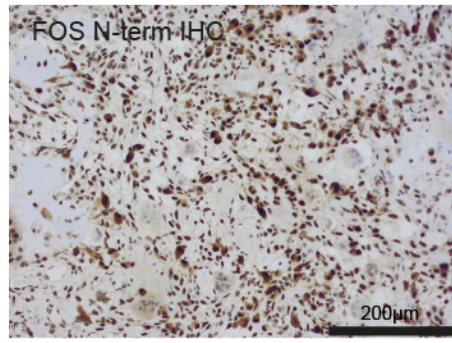
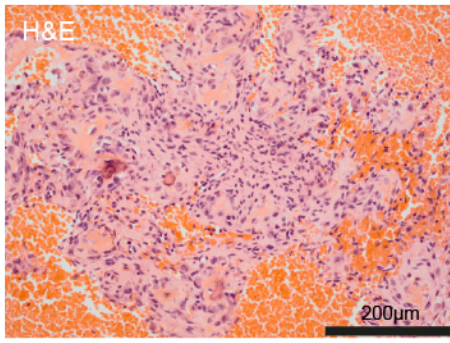
The subsequent pages show plots demonstrating RNAseq coverage with exons shown as blue rectangles. Breakpoints are shown with red lines. Fusion partner coverage is shown for 1kb after the breakpoint (light blue). Heterozygous SNPs, as identified in DNA, relative counts are shown as stacked bars.

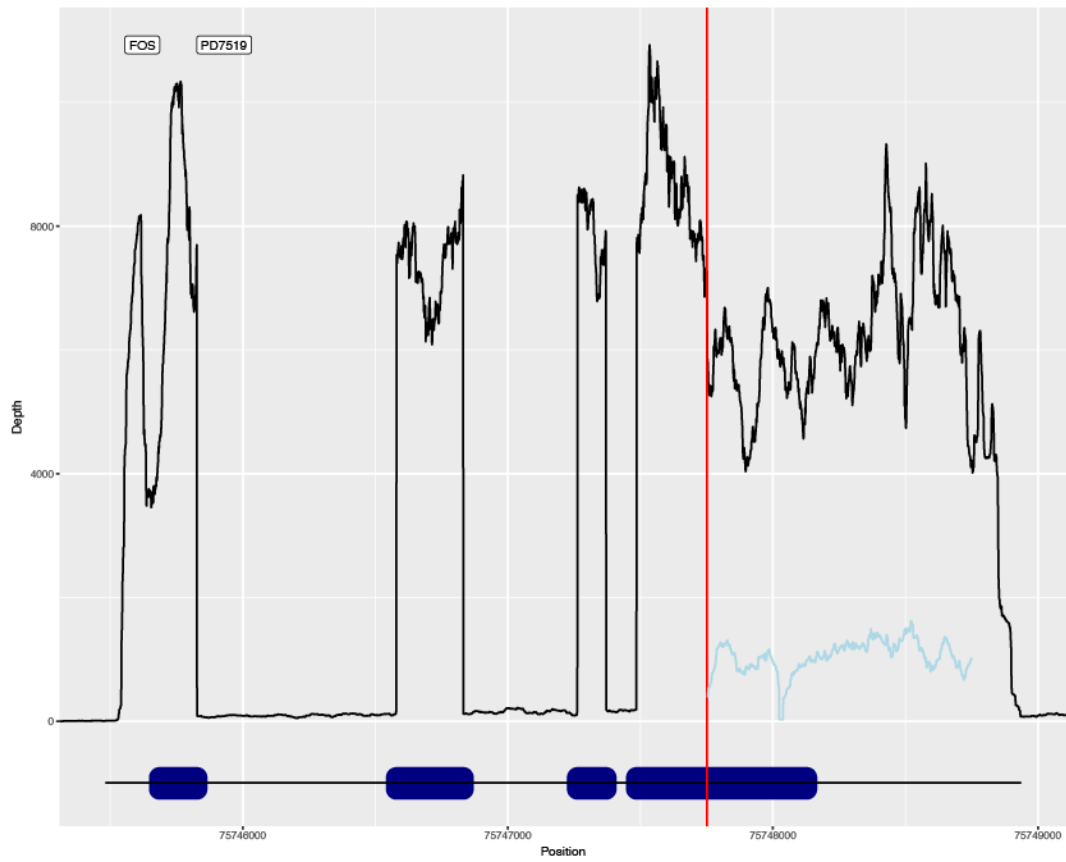
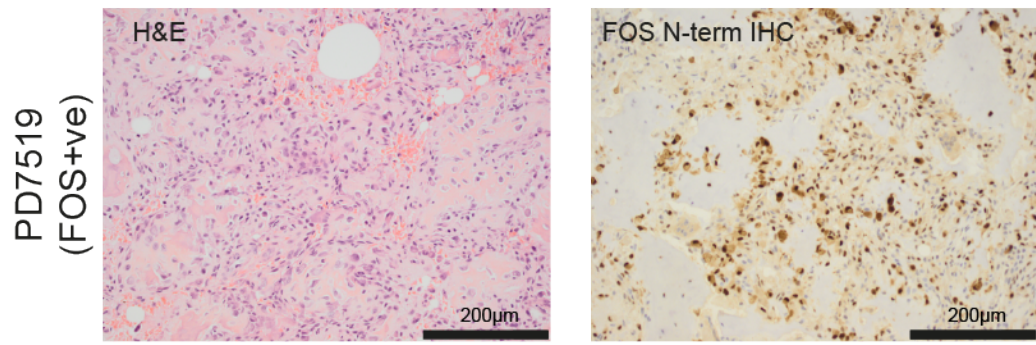
FOS fusion cases

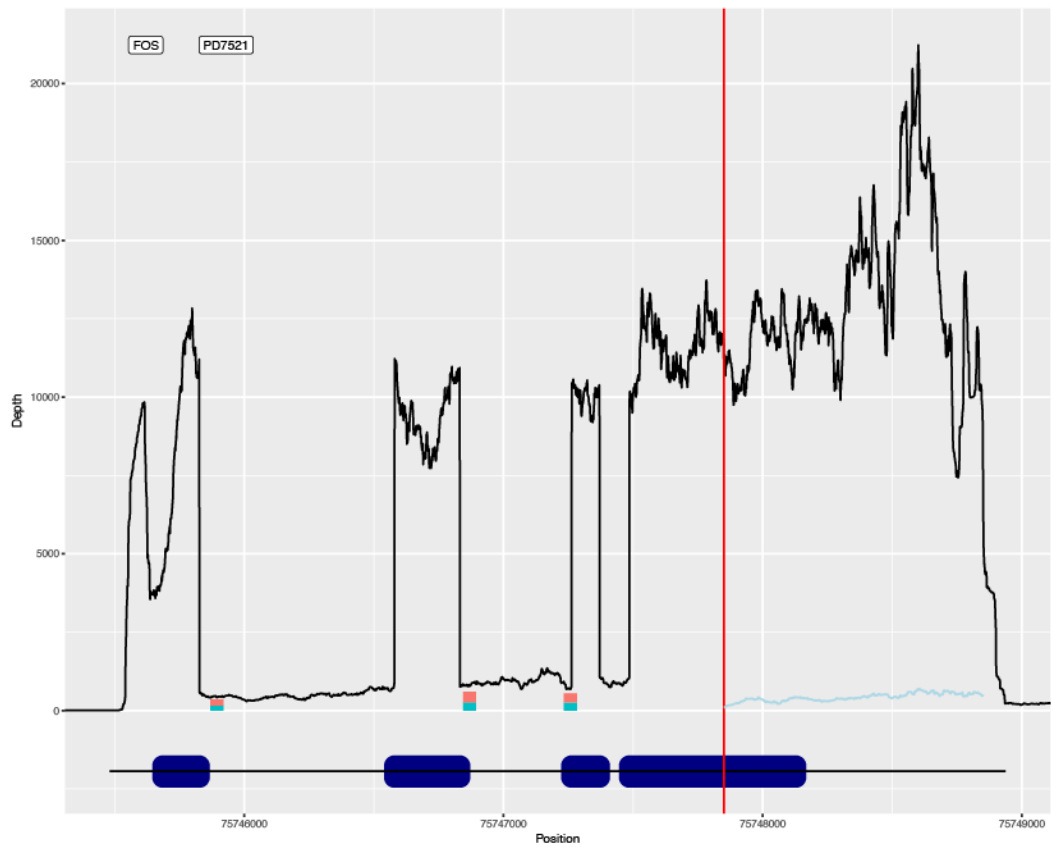
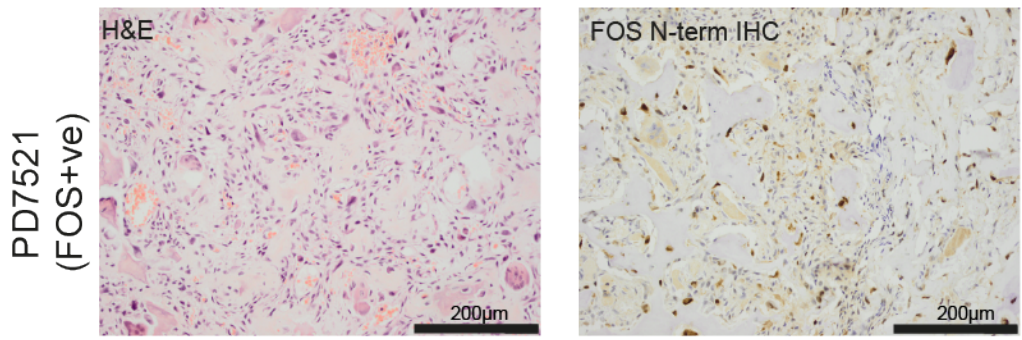


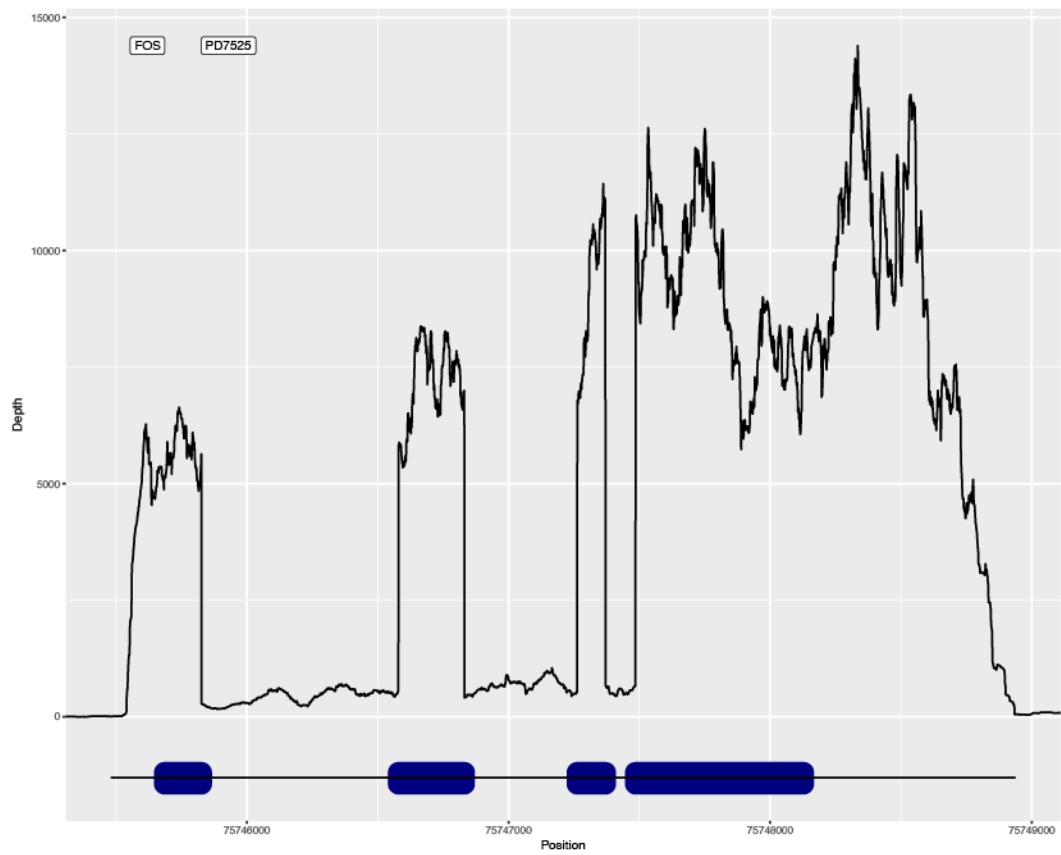
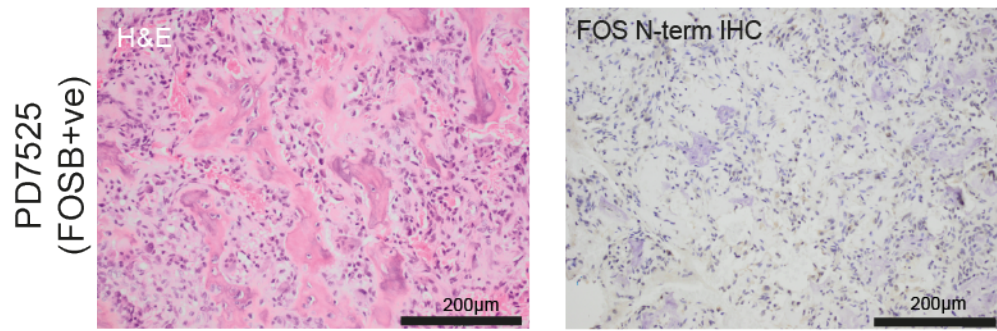


PD13482
(FOS+ve)

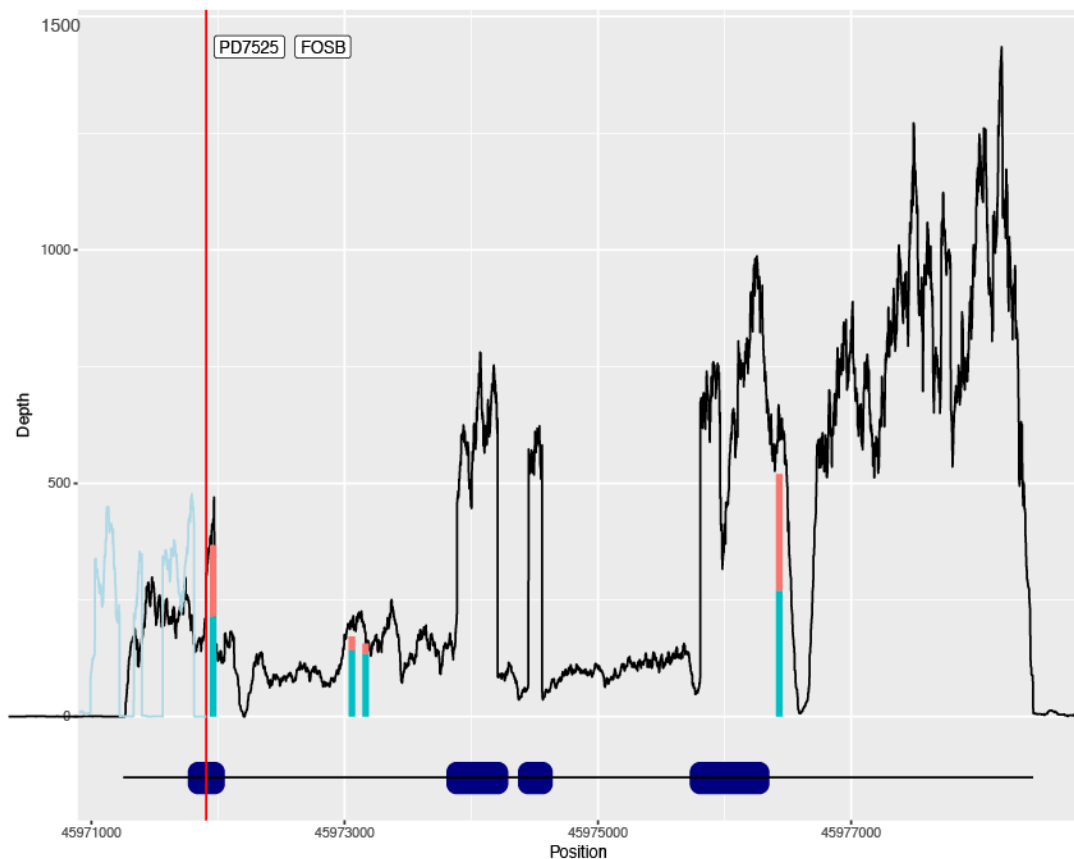








FOSB fusion. There is no clear evidence of allelic imbalance or fusion transcripts dominating wild-type transcripts. As tumour purity is low, it is likely that a significant proportion of RNAseq reads are contributed by normal contaminating cells. As these are not immunoreactive on FOS immunostaining (see panels above or PD7525 for *FOS* fusion negative sample), this strongly hints at predominantly post-transcriptionally effects of truncation. g) For this *FOSB* fusion case there is imbalance of an intron 1 heterozygous SNP, suggesting increased transcriptional activity of one allele.



7.2. Chapter 4 Appendix

7.2.1. SNP Samples

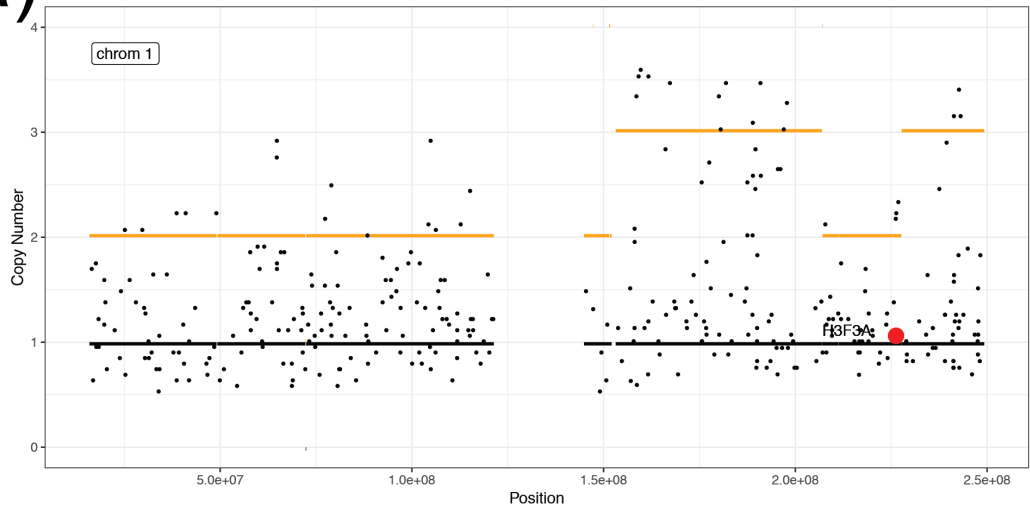
SNP_ID	Tumour_category	Diagnosis	Age	Sex	Sample_site	CN_score
S00067882	BMGCT	GCT	19	F	Ulna, left	0.087
S00067494	BMGCT	GCT	29	M	Femur, distal, left	0.004
S00067499	BMGCT	GCT	26	M	Radius, distal, right	0.006
S00067497	BMGCT	GCT	21	F	Femur, distal, right	0.000
S00069147	Mal	Malignant GCT	27	F	Sacrum	0.256
S00064048	Benign	GCT	37	M	Ulna, distal, right	0.051
S00064049	Benign	GCT	65	F	Tibia, proximal, left	0.392
S00064050	Benign	GCT	26	M	Femur, proximal, left	0.106
S00064051	Benign	GCT	34	M	Femur, distal, right	0.000
S00064063	Benign	GCT	41	M	Tibia, proximal, right	0.019
S00064065	Benign	GCT	18	F	Tibia, proximal, right	0.003
S00064067	Benign	GCT	48	M	Femur, distal	0.049
S00064068	Benign	GCT	28	M	Radius, distal, left	0.002
S00064069	Benign	GCT	64	M	Femur, distal, left	0.000
S00064070	Benign	GCT	40	M	Ulna, distal, right	0.105
S00064071	Benign	GCT	40	F	Radius, distal, right	0.081
S00064073	Benign	GCT	36	M	Radius, distal, right	0.055
S00064074	Benign	GCT	20	M	Radius, distal, left	1.464
S00064055	Benign	GCT	40	M	Tibia, proximal, left	0.020
S00067493	BMGCT	GCT	30	F	Tibia, left	0.000
S00067496	BMGCT	GCT	39	M	Ulna, right	0.066
S00068932	BMGCT	GCT	31	F	Awaiting information from RJAH	0.000
S00068941	Benign	GCT (difficult case)	50	M	Tibia, proximal, right	5.823
S00068945	Mal	GCT atypical (difficult case)	52	M	Tibia, proximal, right	5.269
S00069270	BMGCT	GCT	16	F	Capitate, left	0.898

7.2.2. Methylation array samples

Methylation_ID	Tumour_category	Initial_Diagnosis	Clustering clade	Centre	Age	Sex	Sample_site	Array type	CN score
203259060074_R08C01	BMGCT-Met	Giant Cell Tumour of Bone	G	UCL	15	Female	Lung	EPIC	NA
203259060079_R08C01	BMGCT-Met	Giant Cell Tumour of Bone	G	UCL	15	Female	Lung	EPIC	NA
202262730113_R02C01	BMGCT-Prim	Giant_Cell_Tumour_Bone	G	UCL	15	Female	Radius	EPIC	0.003
3999078064_R06C01	Chondroblastoma	Chondroblastoma	C	UCL	13	Female	Femur	450k	0.006
101130760087_R03C01	Chondroblastoma	Chondroblastoma	C	UCL	14	Male	Humerus	450k	0.003
101130760092_R01C01	Chondroblastoma	Chondroblastoma	C	UCL	15	Male	Tibia	450k	0.004
101130760092_R04C02	Chondroblastoma	Chondroblastoma	C	UCL	16	Male	Tibia	450k	0.008
101130760092_R05C02	Chondroblastoma	Chondroblastoma	C	UCL	30	Female	Femur	450k	0.031
101130760059_R06C01	Chondroblastoma	Chondroblastoma	C	UCL	39	Female	Pelvis	450k	0.012
101103430066_R01C01	Chondroblastoma	Giant_Cell_Tumour_Bone	C	UCL	22	Male	Foot	450k	0.013
101103430066_R06C01	Chondroblastoma	Chondroblastoma	C	UCL	17	Male	Tibia	450k	0.083
101103430066_R01C02	Chondroblastoma	Chondroblastoma	C	UCL	16	Male	Femur	450k	0.085
101103430087_R02C01	Chondroblastoma	Chondroblastoma	C	UCL	16	Female	Humerus	450k	0.008
101103430087_R03C02	Chondroblastoma	Chondroblastoma	C	UCL	15	Male	Femur	450k	0.007
101103430087_R05C02	Chondroblastoma	Chondroblastoma	C	UCL	18	Male	Femur	450k	0.005
101103430097_R05C01	Chondroblastoma	Chondroblastoma	C	UCL	19	Male	Patella	450k	0.021
101103430097_R04C02	Chondroblastoma	Chondroblastoma	C	UCL	5	Male	Femur	450k	0.007
101103430106_R03C02	Chondroblastoma	Chondroblastoma	C	UCL	16	Female	Tibia	450k	0.004
101103430106_R05C02	Chondroblastoma	Chondroblastoma	C	UCL	13	Female	Tibia	450k	0.004
101103430147_R06C01	Chondroblastoma	Pleomorphic Sarcoma	C	UCL	30	Female	Leg	450k	0.028
101103430147_R06C02	Chondroblastoma	Giant_Cell_Tumour_Bone	C	UCL	17	Female	Radius	450k	0.011
100994770005_R02C01	Chondroblastoma	Chondroblastoma	C	UCL	15	Female	Humerus	450k	0.003
3999078002_R03C02	GCT	Giant_Cell_Tumour_Bone	G	UCL	22	Female	Tibia	450k	0.057
3999078064_R01C01	GCT	Giant_Cell_Tumour_Bone	G	UCL	22	Female	Tibia	450k	0.003
3999078064_R03C02	GCT	Giant_Cell_Tumour_Bone	G	UCL	25	Male	Tibia	450k	0.012
101130760087_R02C02	GCT	Giant_Cell_Tumour_Bone	G	UCL	26	Female	Radius	450k	0.006
101130760087_R03C02	GCT	Giant_Cell_Tumour_Bone	G	UCL	50	Male	Femur	450k	0.021
101130760087_R04C02	GCT	Giant_Cell_Tumour_Bone	G	UCL	20	Male	Fibula	450k	0.016
101130760092_R01C02	GCT	Giant_Cell_Tumour_Bone	G	UCL	37	Female	Tibia	450k	0.009
101130760092_R02C02	GCT	Giant_Cell_Tumour_Bone	G	UCL	77	Male	Tibia	450k	0.015
101130760092_R03C02	GCT	Giant_Cell_Tumour_Bone	G	UCL	50	Male	Femur	450k	0.007
101130760059_R02C01	GCT	Giant_Cell_Tumour_Bone	G	UCL	30	Female	Fibula	450k	0.013
101130760059_R03C01	GCT	Giant_Cell_Tumour_Bone	G	UCL	36	Female	Tibia	450k	0.041
101130760059_R04C01	GCT	Giant_Cell_Tumour_Bone	G	UCL	21	Female	Tibia	450k	0.005
100994770004_R06C01	GCT	Osteoblastoma	M	UCL	18	Female	Rib	450k	0.014
101103430084_R02C02	GCT	Giant_Cell_Tumour_of_Bone	G	UCL	26	NA	NA	450k	0.008
101103430087_R01C01	GCT	Giant_Cell_Tumour_Bone	G	UCL	52	Male	Foot	450k	0.018
101103430087_R06C01	GCT	Giant_Cell_Tumour_Bone	G	UCL	35	Female	Femur	450k	0.016
101103430097_R01C02	GCT	Giant_Cell_Tumour_Bone	G	UCL	43	Male	Femur	450k	0.009
101103430106_R01C01	GCT	Giant_Cell_Tumour_Bone	G	UCL	22	Female	Humerus	450k	0.080
101103430147_R01C01	GCT	Giant_Cell_Tumour_Bone	G	UCL	19	Male	Foot	450k	0.026
101103430147_R03C01	GCT	Giant_Cell_Tumour_Bone	G	UCL	38	Male	Femur	450k	0.024
101103430147_R04C01	GCT	Giant_Cell_Tumour_of_Bone	G	UCL	30	NA	NA	450k	0.008
101103430147_R01C02	GCT	Giant_Cell_Tumour_Bone	G	UCL	39	Female	Femur	450k	0.007
101103430147_R02C02	GCT	Giant_Cell_Tumour_Bone	G	UCL	44	Male	Tibia	450k	0.029
101231000137_R01C02	GCT	Giant_Cell_Tumour_Bone	G	DKFZ	27	Female	Extremity (upper)	450k	0.014
101231000003_R05C01	GCT	Giant_Cell_Tumour_Bone	G	DKFZ	33	Male	Extremity (lower)	450k	0.037
101231000003_R06C01	GCT	Giant_Cell_Tumour_Bone	G	DKFZ	29	Male	Extremity (upper)	450k	0.004
3998568071_R06C02	GCT	Giant_Cell_Tumour_Bone	G	DKFZ	75	Female	Extremity (lower)	450k	0.018
3998568072_R01C01	GCT	Giant_Cell_Tumour_Bone	G	DKFZ	32	Male	Extremity (upper)	450k	0.021
3998568072_R03C01	GCT	Giant_Cell_Tumour_Bone	G	DKFZ	20	Female	Pelvis	450k	0.009
3998568072_R04C01	GCT	Giant_Cell_Tumour_Bone	G	DKFZ	27	Female	Extremity (upper)	450k	0.002
200091640036_R04C02	GCT	Giant_Cell_Tumour_Bone	G	DKFZ	34	Female	Extremity (lower)	450k	0.002
200091640036_R05C02	GCT	Giant_Cell_Tumour_Bone	G	DKFZ	58	Male	Extremity (lower)	450k	0.051
200091640036_R06C02	GCT	Giant_Cell_Tumour_Bone	G	DKFZ	59	Male	Extremity (upper)	450k	0.019
200091640026_R02C01	GCT	Giant_Cell_Tumour_Bone	G	DKFZ	26	Female	Extremity (upper)	450k	0.001
200109360096_R04C01	GCT	Giant_Cell_Tumour_Bone	G	DKFZ	22	Female	Extremity (lower)	450k	0.001
200362700204_R05C01	GCT	Giant_Cell_Tumour_Bone	G	DKFZ	29	Female	Extremity (lower)	450k	0.000
200788220019_R07C01	GCT	Giant_Cell_Tumour_Bone	G	UCL	15	Female	Tibia	EPIC	0.001
200788220001_R08C01	GCT	Giant_Cell_Tumour_Bone	G	UCL	21	Male	Metatarsal	EPIC	0.001
200788220051_R02C01	GCT	Osteosarcoma	M	UCL	20	Female	Fibula	EPIC	0.001
202273260117_R05C01	GCT	Giant_Cell_Tumour_Bone	G	UCL	65	Male	Tibia	EPIC	0.002
202273260117_R06C01	GCT	Giant_Cell_Tumour_Bone	G	UCL	34	Male	Femur	EPIC	0.002
101103430106_R04C01	GCT	Giant_Cell_Tumour_Bone	G	UCL	29	Female	Femur	450k	0.035
100994770004_R04C02	Mal_G34	Giant_Cell_Tumour_Bone	G	UCL	35	Female	Vertebra	450k	0.101
3999112146_R01C02	Mal_G34	Osteosarcoma	M	DKFZ	59	Male	Extremity (upper)	450k	0.392
3998909204_R06C01	Mal_G34	Osteosarcoma	M	DKFZ	75	Male	Extremity (lower)	450k	0.089
3998568072_R02C01	Mal_G34	Giant cell tumor of bone (malignant)	G	DKFZ	18	Male	Extremity (upper)	450k	0.019
200325530180_R01C01	Mal_G34	Osteosarcoma	M	DKFZ	71	Male	Extremity (lower)	450k	0.067
200325530180_R02C01	Mal_G34	Osteosarcoma	M	DKFZ	34	Female	Extremity (lower)	450k	0.013
200925700120_R03C01	Mal_G34	Osteosarcoma	M	DKFZ	75	Male	Extremity (upper)	EPIC	0.459
200788220016_R05C01	Mal_G34	Osteosarcoma	M	UCL	45	Female	Femur	EPIC	0.064
200788220001_R01C01	Mal_G34	Osteosarcoma	M	UCL	47	Male	Femur	EPIC	0.013
200788220049_R06C01	Mal_G34	Pleomorphic Sarcoma	M	UCL	36	Female	Femur	EPIC	0.274

Methylation_ID	Tumour_category	Initial_Diagnosis	Clustering clade	Centre	Age	Sex	Sample_site	Array type	CN score
200788220003_R03C01	Mal_G34	Osteosarcoma	M	UCL	53	Male	Tibia	EPIC	0.473
200788220019_R02C01	Mal_G34	Osteosarcoma	M	UCL	25	Male	Tibia	EPIC	0.016
200788220019_R05C01	Mal_G34	Osteosarcoma	M	UCL	40	Female	Tibia	EPIC	0.012
200788220049_R01C01	Mal_G34	Osteosarcoma	M	UCL	29	Female	Pelvis	EPIC	0.391
200788220001_R05C01	Mal_G34	Giant_Cell_Tumour_Bone	G	UCL	17	Female	Femur	EPIC	0.001
3999112131_R03C01	Osteosarcoma	Osteosarcoma	Os	DKFZ	6	Female	Extremity (lower)	450k	0.574
3999112131_R04C01	Osteosarcoma	Osteosarcoma	Os	DKFZ	18	Male	Extremity (upper)	450k	0.308
3999112137_R05C01	Osteosarcoma	Osteosarcoma	Os	DKFZ	16	Male	Extremity (lower)	450k	0.214
3999112137_R06C01	Osteosarcoma	Osteosarcoma	Os	DKFZ	18	Male	Extremity (lower)	450k	0.334
3999112137_R01C02	Osteosarcoma	Osteosarcoma	Os	DKFZ	14	Male	Extremity (lower)	450k	0.269
3999112146_R06C01	Osteosarcoma	Osteosarcoma	Os	DKFZ	20	Male	Extremity (lower)	450k	0.484
3999112146_R03C02	Osteosarcoma	Osteosarcoma	Os	DKFZ	29	Male	Extremity (lower)	450k	0.352
3999112146_R05C02	Osteosarcoma	Osteosarcoma	Os	DKFZ	34	Female	Extremity (lower)	450k	0.001
3998909204_R03C01	Osteosarcoma	Osteosarcoma	Os	DKFZ	14	Female	Scapula	450k	0.211
3998909203_R05C01	Osteosarcoma	Osteosarcoma	Os	DKFZ	65	Male	Spine	450k	0.376
3998909204_R04C02	Osteosarcoma	Osteosarcoma	Os	DKFZ	9	Male	Extremity (lower)	450k	0.416
3998909203_R05C02	Osteosarcoma	Osteosarcoma	Os	DKFZ	35	Male	Extremity (upper)	450k	0.162
3998909203_R06C02	Osteosarcoma	Osteosarcoma	Os	DKFZ	12	Female	Extremity (lower)	450k	0.466
3998920094_R02C02	Osteosarcoma	Osteosarcoma	Os	DKFZ	24	Female	Extremity (lower)	450k	0.312
3998920096_R01C02	Osteosarcoma	Osteosarcoma	Os	DKFZ	12	Female	Extremity (lower)	450k	0.246
3998523055_R05C02	Osteosarcoma	Osteosarcoma	Os	DKFZ	27	Male	Head	450k	0.347
200397540005_R06C02	Osteosarcoma	Osteosarcoma	Os	DKFZ	14	Female	Extremity (lower)	450k	0.186
200397540010_R02C01	Osteosarcoma	Osteosarcoma	Os	DKFZ	21	Male	Extremity (lower)	450k	0.567
200360420062_R01C01	Osteosarcoma	Osteosarcoma	Os	DKFZ	20	Female	Jaw	450k	0.302
200362700194_R06C02	Osteosarcoma	Osteosarcoma	Os	DKFZ	30	Male	Spine	450k	0.009
200925700157_R01C01	Osteosarcoma	Osteosarcoma	Os	DKFZ	34	Female	Extremity (lower)	EPIC	0.477
202259490096_R02C01	Osteosarcoma	Osteosarcoma	Os	UCL	56	Male	Tibia and Fibula	EPIC	0.292
202262730037_R08C01	Osteosarcoma	Osteosarcoma	Os	UCL	11	Female	Tibia	EPIC	0.443
202262730098_R04C01	Osteosarcoma	Osteosarcoma	Os	UCL	56	Female	Femur	EPIC	0.324
202273260008_R03C01	Osteosarcoma	Osteosarcoma	Os	UCL	22	Male	Rib	EPIC	0.264
202273260008_R08C01	Osteosarcoma	Osteosarcoma	Os	UCL	18	Male	Fibula	EPIC	0.504
202273260019_R03C01	Osteosarcoma	Osteosarcoma	Os	UCL	14	Male	Femur	EPIC	0.259
202273260019_R04C01	Osteosarcoma	Osteosarcoma	Os	UCL	19	Male	Femur	EPIC	0.035
202273260020_R02C01	Osteosarcoma	Osteosarcoma	Os	UCL	15	Male	Femur	EPIC	0.331
202273260020_R04C01	Osteosarcoma	Osteosarcoma	Os	UCL	13	Male	Tibia	EPIC	0.489
202273260054_R01C01	Osteosarcoma	Osteosarcoma	Os	UCL	20	Male	Tibia	EPIC	0.452
202273260058_R02C01	Osteosarcoma	Osteosarcoma	Os	UCL	36	Female	Humerus	EPIC	0.294
202273260058_R03C01	Osteosarcoma	Osteosarcoma	Os	UCL	10	Male	Femur	EPIC	0.127
202273260064_R04C01	Osteosarcoma	Osteosarcoma	Os	UCL	36	Male	Tibia	EPIC	0.019
202273260065_R02C01	Osteosarcoma	Osteosarcoma	Os	UCL	17	Male	Tibia	EPIC	0.406
202273260066_R07C01	Osteosarcoma	Osteosarcoma	Os	UCL	12	Male	Femur	EPIC	0.547
202273260111_R03C01	Osteosarcoma	Osteosarcoma	Os	UCL	13	Female	Femur	EPIC	0.179
202273260111_R06C01	Osteosarcoma	Osteosarcoma	Os	UCL	15	Male	Humerus	EPIC	0.498
203259060009_R03C01	Osteosarcoma	Osteosarcoma	Os	UCL	16	Male	Pelvis	EPIC	NA
203259060009_R04C01	Osteosarcoma	Osteosarcoma	Os	UCL	78	Male	Fibula	EPIC	NA
203259060043_R01C01	Osteosarcoma	Osteosarcoma	Os	UCL	19	Male	Femur	EPIC	NA
203259060043_R03C01	Osteosarcoma	Osteosarcoma	Os	UCL	7	Female	Femur	EPIC	NA

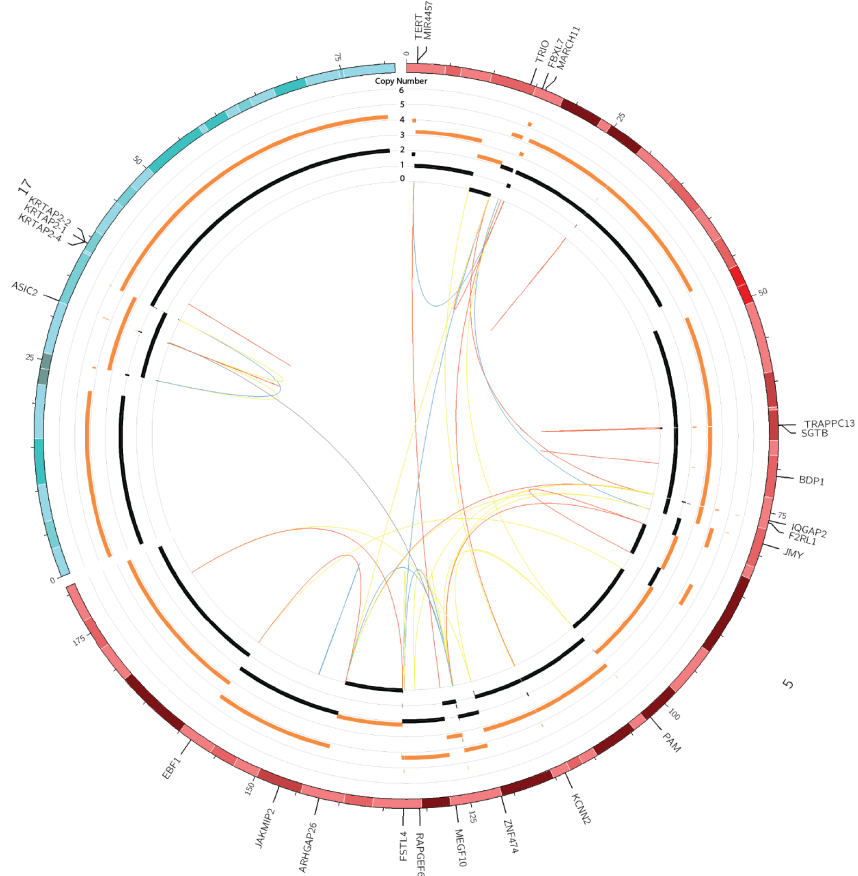
a)



b)



c)



7.3. Chapter 5 Appendix

7.3.1. PCAWG cohort

Organ	Abbreviation	Included Subtypes	Jurisdictions	Cases	Sex		Age	
					F	M	Med	10-90th
Neural Crest								
CNS	CNS-GBM	Glioblastoma	US	41	13	28	60	43-72
CNS	CNS-Medullo	Medulloblastoma; Desmoplastic medullo.; Large cell medulloblastoma	DE	146	67	79	9	3-28
CNS	CNS-Oligo	Oligodendroglioma	US	18	9	9	41	21-62
CNS	CNS-PiloAstro	Pilocytic astrocytoma	DE	89	47	42	8	2-17
Skin	Skin-Melanoma	Malignant melanoma	AU, US	107	38	69	57	37-78
Endoderm								
Biliary	Biliary-AdenoCA	Papillary cholangiocarcinoma	JP, SG	34	15	19	64	53-76
Bladder	Bladder-TCC	Transitional cell carcinoma; Papillary transitional cell carcinoma	US	23	8	15	65	52-80
Colon/Rectum	ColoRect-AdenoCA	Adenocarcinoma; Mucinous adeno.	US	60	30	30	67	46-81
Esophagus	Eso-AdenoCA	Adenocarcinoma	UK	98	14	84	70	56-79
Liver	Liver-HCC	Hepatocellular carcinoma; Combined HCC/cholangio; Fibrolamellar HCC	FR, JP, US	317	89	228	67	50-78
Lung	Lung-AdenoCA	Adenocarcinoma; Adenocarcinoma <i>in situ</i> ; Mucinous adenocarcinoma	US	38	20	18	66	47-77
Lung	Lung-SCC	Squamous cell carcinoma; Basaloid SCC	US	48	10	38	68	54-77
Pancreas	Panc-AdenoCA	Adenocarcinoma; Acinar cell Ca.; Mucinous adeno.; Adenosquaous Ca.	AU, CA	239	119	120	67	50-79
Pancreas	Panc-Endocrine	Neuroendocrine carcinoma	AU, IT	85	30	55	59	38-75
Prostate	Prost-AdenoCA	Adenocarcinoma	CA, DE, UK, US	210	0	210	59	47-71

Stomach	Stomach-AdenoCA	Adenocarcinoma; Mucinous adeno.; Papillary adeno.; Tubular adeno.	CN, US	75	18	57	65	47-79
Thyroid	Thy-AdenoCA	Adenocarcinoma; Adeno., columnar cell; Adeno., follicular type	US	48	37	11	51	26-75
Mesoderm								
Bone/Soft Tissue	Bone-Benign	Osteoblastoma; Osteofibrous dysplasia	UK	7	4	3	18	12-30
Bone/Soft Tissue	Bone-Benign	Chondroblastoma; Chondromyxoid fibroma	UK	9	2	7	16	14-38
Bone/Soft Tissue	Bone-Epith	Adamantinoma; Chordoma	UK	10	4	6	60	37-67
Bone/Soft Tissue	Bone-Osteosarc	Osteosarcoma	UK/NO	38	20	18	20	9-58
Bone/Soft Tissue	SoftTissue-Leiomyo	Leiomyosarcoma	US	15	10	5	61	51-78
Bone/Soft Tissue	SoftTissue-Liposarc	Liposarcoma	US	19	5	14	n/a	n/a
Cervix	Cervix-AdenoCA	Adenocarcinoma	US	2	2	0	39	33-46
Cervix	Cervix-SCC	Squamous cell carcinoma	US	18	18	0	39	25-58
Head/Neck	Head-SCC	Squamous cell carcinoma	IN, US	57	10	47	53	34-71
Kidney	Kidney-ChRCC	Adenocarcinoma, chromophobe type	US	45	19	26	47	34-69
Kidney	Kidney-RCC	Adenocarcinoma, clear cell type; Adenocarcinoma, papillary type	EU, US	144	54	90	60	48-75
Lymphoid	Lymph-BNHL	Burkitt; Diffuse large B-cell; Follicular; Marginal zone; Post-transplant	DE, US	107	51	56	57	10-74
Lymphoid	Lymph-CLL	Chronic lymphocytic leukaemia	ES	95	31	64	62	46-78
Myeloid	Myeloid-AML	Acute myeloid leukaemia	KR, UK	10	3	7	50	35-56
Myeloid	Myeloid-MDS	Chronic myelomonocytic leukaemia; MDS with ring sideroblasts	UK	2	1	1	76	74-77
Myeloid	Myeloid-MPN	Essential thrombocythemia; Polycythemia vera; Myelofibrosis	UK	26	14	12	56	38-75
Ovary	Ovary-AdenoCA	Adenocarcinoma; Serous cystadenocarcinoma	AU, US	113	113	0	60	48-74

Uterus	Uterus-AdenoCA	Adenocarcinoma, endometrioid; Serous cystadenocarcinoma	US	51	51	0	69	57-81
Ectoderm								
Breast	Breast-AdenoCA	Infiltrating duct carcinoma; Medullary carcinoma; Mucinous adeno.	EU, UK, US	198	197	1	56	39-76
Breast	Breast-DCIS	Duct micropapillary carcinoma	EU, UK	3	3	0	55	43-60
Breast	Breast-LobularCA	Lobular carcinoma	EU, UK, US	13	13	0	53	42-69
Total				2658	1189	1469	59	21-76

7.4. Related authored papers

Publications in press resulting from work in this thesis:

FITTALL, M. W. & VAN LOO, P. 2019. Translating insights into tumor evolution to clinical practice: promises and challenges. *Genome Medicine*, 11, 20. ***This is the basis of much of the introduction 1.2.2***

FITTALL, M. W., MIFSUD, W., PILLAY, N., YE, H., STROBL, A. C., VERFAILLIE, A., DEMEULEMEESTER, J., ZHANG, L., BERISHA, F., TARABICHI, M., YOUNG, M. D., MIRANDA, E., TARPEY, P. S., TIRABOSCO, R., AMARY, F., GRIGORIADIS, A. E., STRATTON, M. R., VAN LOO, P., ANTONESCU, C. R., CAMPBELL, P. J., FLANAGAN, A. M. & BEHJATI, S. 2018. Recurrent rearrangements of FOS and FOSB define osteoblastoma. *Nat Commun*, 9, 2150. ***This is the basis of chapter Chapter 3***

PCAWG 2020. Pan-cancer analysis of whole genomes. *Nature*, 578, 82-93. ***This flagship paper for the PCAWG consortium features nine scientific highlights with their own structured author lists. Chapter 5 forms the basis of 3/9 scientific highlights, for which Maxime Tarabichi, Jonas Demeulemeester and I are jointly first authors***

Publications in preparation/submission from work in this thesis:

FITTALL, M.W., LOMBARD, P., ELLERY, P., STROBL, A. C., TARABICHI, M., SILL, M., KOELSHE C, DEMEULEMEESTER, J., TIRABOSCO, R., AMARY, F., VAN LOO, P., CAMPBELL, P. J., JONES, D.T.W, BEHJATI, S. & FLANAGAN, A. M. Patterns of progression in H3.3 mutated bone tumours. Manuscript in preparation. ***This is the basis of chapter Chapter 4***

REVIEW

Open Access

Translating insights into tumor evolution to clinical practice: promises and challenges



Matthew W. Fittall^{1,2,3} and Peter Van Loo^{1,4*}

Abstract

Accelerating technological advances have allowed the widespread genomic profiling of tumors. As yet, however, the vast catalogues of mutations that have been identified have made only a modest impact on clinical medicine. Massively parallel sequencing has informed our understanding of the genetic evolution and heterogeneity of cancers, allowing us to place these mutational catalogues into a meaningful context. Here, we review the methods used to measure tumor evolution and heterogeneity, and the potential and challenges for translating the insights gained to achieve clinical impact for cancer therapy, monitoring, early detection, risk stratification, and prevention. We discuss how tumor evolution can guide cancer therapy by targeting clonal and subclonal mutations both individually and in combination. Circulating tumor DNA and circulating tumor cells can be leveraged for monitoring the efficacy of therapy and for tracking the emergence of resistant subclones. The evolutionary history of tumors can be deduced for late-stage cancers, either directly by sampling precursor lesions or by leveraging computational approaches to infer the timing of driver events. This approach can identify recurrent early driver mutations that represent promising avenues for future early detection strategies. Emerging evidence suggests that mutational processes and complex clonal dynamics are active even in normal development and aging. This will make discriminating developing malignant neoplasms from normal aging cell lineages a challenge. Furthermore, insight into signatures of mutational processes that are active early in tumor evolution may allow the development of cancer-prevention approaches. Research and clinical studies that incorporate an appreciation of the complex evolutionary patterns in tumors will not only produce more meaningful genomic data, but also better exploit the vulnerabilities of cancer, resulting in improved treatment outcomes.

Background

Over time, the therapeutic approach to cancer is evolving from targeting the clinical phenotype (tumor size, location, stage, histological type, and grade), to targeting a molecular phenotype (such as surface receptor status or the presence of activating or sensitizing mutations) [1, 2]. The clinical phenotype can be targeted spatially with surgery and radiotherapy or systemically using cytotoxic chemotherapies. The molecular phenotype has been targeted by both direct and indirect endocrine manipulation, by an array of small molecule inhibitors, and by monoclonal antibody therapies. Both approaches typically consider the target to be static (to be treated until clinical failure) and homogeneous (one sample represents all tumor cells).

The application of evolutionary concepts to cancer was proposed several decades ago by Peter Nowell [3]. Reliable exploration of the degree of variation within and between cancers has only become possible with the increasing availability of next generation sequencing and associated computational analysis [4–6].

All of the cells within a tumor are unique, comprising different somatic variants and epigenetic and transcriptomic states. Even normal cells are likely to accrue approximately three somatic mutations every cell cycle [7, 8]. Most of these changes will have no functional impact and are ‘passengers’ on the cells’ evolutionary journey (Box 1). Somatic mutations (or epigenetic changes) that have an advantageous functional impact are ‘drivers’ and will allow a cell to expand clonally and outcompete its neighbors. When a clonal expansion goes to completion, the entire population will be ‘clonally’ descended from that founder cell, or clone. The last complete clonal expansion will have arisen from the most recent common

* Correspondence: Peter.VanLoo@crick.ac.uk
¹The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK
⁴University of Leuven, Herestraat 49, B-3000 Leuven, Belgium
Full list of author information is available at the end of the article



© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Box 1	
Glossary	
Clone	A group of cells that are all descended from a single ancestor. Mutations that are shared between these cells are commonly described as 'clonal'.
Subclone	Cells originating from a more recent cell than the most recent common ancestor. These will possess both the clonal mutations and also subclonal mutations that are private to the subclone.
Driver mutation	A mutation with a beneficial functional impact on a cell (for example, affecting growth, invasion, or metastasis).
Passenger mutation	A mutation with no functional impact. Both driver and passenger mutations (the latter representing the large majority of mutations) can still be used to identify clonal or subclonal populations.
Most recent common ancestor (MRCA)	The theoretical founder cell of the tumor, from which all cancer cells in a cancer sample are derived. The most recent common ancestor possesses all mutations that are common to all of the tumor cells.
Branching evolution	Divergence in tumor evolution leading to separate subclonal populations.
Linear evolution	The absence of apparent divergence or branches in evolution. All evolution prior to the MRCA will always appear linear as all other pre-MRCA branches have become extinct.
Gradual evolution	An iterative pattern of mutation acquisition and selection over time.
Punctuated evolution	Discontinuous acquisition of mutations over time with periods of relative stasis. Mutations may be acquired in distinct patterns and be co-located, or can be distributed across the genome.

ancestor (MRCA), defined as the most recent individual cell from which all existing cancer cells in a cancer sample are descendants. If a clonal expansion or sweep is incomplete, the expanded population is subclonal, comprising only a fraction of the tumor cells. Diverging subclones with mutually exclusive mutations can co-exist within a tumor [9]. Intra-tumor heterogeneity, or the presence of subclones possessing private mutations within a tumor, has been observed across many cancer types and seems to be nearly ubiquitous [10, 11].

The dynamics of evolution in cancer are still not fully understood [12]. Traditionally, mutation and selection are thought to be slow iterative processes that occur throughout a cancer's lifetime, a process of gradual evolution. The patterns of mutations observed in some tumors, however, suggest that mutations can also be acquired in sudden bursts, leading to punctuated evolutionary steps [13–19].

An emerging wealth of cancer genome sequencing data is informing our understanding of tumor evolution, and will cause a fundamental paradigm shift in our approach to cancer. This will impact all aspects of cancer management, including cancer therapy, monitoring, early detection, and prevention (Table 1).

Measuring intra-tumor heterogeneity and tumor evolution

Implicit in the heterogeneity of tumor cells and essential for evolution is variation in either the genome or the epigenome [20–22]. Although epigenetic heterogeneity has been shown to have prognostic utility [23–26] and is the subject of intense study, genetic heterogeneity is better understood at present, and is the focus of this review.

Intra-tumor heterogeneity and evolution can be inferred from the pattern of mutations that is detected. Clonal mutations, which are common to all cells within a tumor, were present in the tumor cells' most recent common ancestor, whereas subclonal mutations were acquired later and are therefore only found in a proportion of tumor cells (Box 1). The frequency of a mutation in sequencing data (the variant allele frequency (VAF)) can be used to establish its

Table 1 Promises and challenges in translating insights into tumor evolution to clinical practice

	Therapy	Monitoring	Early diagnosis and stratification	Prevention
Promises	<ul style="list-style-type: none"> Clonal therapy targeting clonal mutations to eradicate all tumor cells (such as targeted therapy or immunotherapy) Preempt resistance Adaptive therapy to chronically control disease 	<ul style="list-style-type: none"> Bespoke monitoring based on tumor-specific mutations 	<ul style="list-style-type: none"> Identify genetic changes meriting intervention 	<ul style="list-style-type: none"> Mutational signatures can suggest etiological factors that drive early tumorigenesis
Challenges	<ul style="list-style-type: none"> Sampling strategy Inevitable clonal monotherapy resistance Bespoke combination therapies complicate toxicity and licensing 	<ul style="list-style-type: none"> High cost Novel mutations or subclones may be missed Early detection of relapse may not improve outcome 	<ul style="list-style-type: none"> Normal tissues contain canonical cancer mutations Early diagnosis may not improve outcome 	<ul style="list-style-type: none"> Exogenous factors may not be preventable Some tumors may not be preventable (such as those of children or young adults)

clonality. VAF is influenced by both the proportion of cells that possess the mutation and the number of both mutated and un-mutated copies of that DNA locus. Mutation frequencies can be estimated by sampling, which has intrinsic spatial, genomic, and statistical limitations (Fig. 1). Intra-tumor heterogeneity has been extensively explored using exome or genome sequencing of multiple regions of resected primary tumors [9, 12, 19, 27–29]. Paired primary–metastasis studies and post-mortem studies have allowed detailed insight into the evolution and patterns of spread of metastases [30–33]. Intra-tumor heterogeneity has been shown to be prognostic across cancer types [10, 34, 35], and is predominantly associated with the degree and heterogeneity of aneuploidy. It has also been shown to impact therapy: potentially targetable driver mutations can be subclonal, suggesting that treatment would only be partially effective [36].

Describing tumor evolution requires measurement over time. Models of tumorigenesis, such as the ‘Vogelgram’, were created by sampling different stages of cancer progression across a population [37, 38]. The initial Vogelgram in colorectal cancer was established by probing a limited number of putative oncogenes, identified from hereditary cases, across the histologically defined spectrum of disease [39]. Mutations that are found across different stages of disease are assumed to arise early in tumor evolution, whereas those found only in established invasive cancers can be assumed to occur later in tumor evolution. Rarely, in individuals with predisposing risk factors such as those who have colitis-associated colon cancer, the whole spectrum of tumor progression can be observed simultaneously [27]. Despite applying modern genomic techniques, models of progression can remain elusive if the genome is already markedly aberrated in pre-invasive lesions, as in the precursors of lung squamous cell carcinoma [40]. This modeling approach also relies on the assumption that cancers of the same histology have a highly stereotyped genetic progression that is common to different tumors.

Computational approaches have been developed to infer the history of an individual tumor that is already established from its own genome, as recently reviewed [41, 42]. Although these approaches typically allow only partial reconstruction of a tumor’s evolutionary history, from a single biopsy, aggregating results across multiple tumors can be a powerful approach [42]. Taking multiple samples from the same tumor over time or across space can also significantly increase the power of these reconstruction approaches [41, 42]. In metastatic solid organ tumors, repeated sampling over time is challenging, so hematological malignancies have been studied most extensively in this context [43–46]. Circulating tumor DNA (ctDNA) and cells shed from solid tumors offer the potential to track subclonal mutations, albeit with limited sensitivity and specificity.

Most DNA sequencing has been performed on pooled DNA from multiple cells and, consequently, ambiguity can remain as to whether mutations co-occur in the same cell. Single-cell sequencing can overcome this, albeit at higher cost and at the expense of substantial sequencing artifacts [47–50]. High-throughput techniques have been developed for analyzing large numbers of single cells, although these methods are most advanced for transcriptome sequencing [51]. Single-cell sequencing of other ‘omic layers is currently relatively costly and available for fewer cells [52], but exciting high-throughput approaches are now emerging [53]. Techniques to analyze multiple layers simultaneously have also been developed recently [54–56], but these are currently costly and lower throughput. These ‘multi-omic’ approaches are likely to significantly improve the interpretation of non-genetic cellular heterogeneity. Such interpretation is also confounded by heterogeneity among non-tumor cells that results from the variety of cell types and states within a tumor [57, 58].

Future approaches for measuring tumor heterogeneity that could be used clinically would need to satisfy the following criteria: (i) sampling should be minimally invasive or performed as part of tumor resection; (ii) sampling of the tumor should be as comprehensive as possible, ideally without any spatial biases; (iii) sample handling and preservation will need to be simple and readily available in the clinic; (iv) simple proxy biomarkers need to be available to assay heterogeneity reliably; and (v) assays need to be rapid and cost-effective.

Recently, a conceptual consideration of how evolution and heterogeneity could be summarized was explored in a consensus statement by Maley et al. [59]. They proposed binary divisions of the degree of heterogeneity (diversity, D) and evolution (rate of change, Δ) that could be combined in a single four-level Evo-Index. As yet, it is not clear how these scores would be generated or whether such a simple binary system is informative.

Can tumor evolution guide cancer therapy?

The rational design of cancer therapies based on genomic data has to date, with a few notable exceptions, been expensive and has delivered limited benefit to patients [60]. Even therapies specifically targeting prevalent tumor mutations, such as the *BRAF* V600E mutation in melanoma [61] and a variety of *EGFR* point mutations in lung cancer [62], only lead to relatively short-lived tumor responses. Understanding the heterogeneity that exists within tumors and their ability to evolve in response to therapy may allow more optimized treatment strategies (Table 1).

Individual clonal therapies

The simplest conceivable therapeutic approach is to target individual clonal mutations. By targeting mutations that are present in all tumor cells, the entire tumor

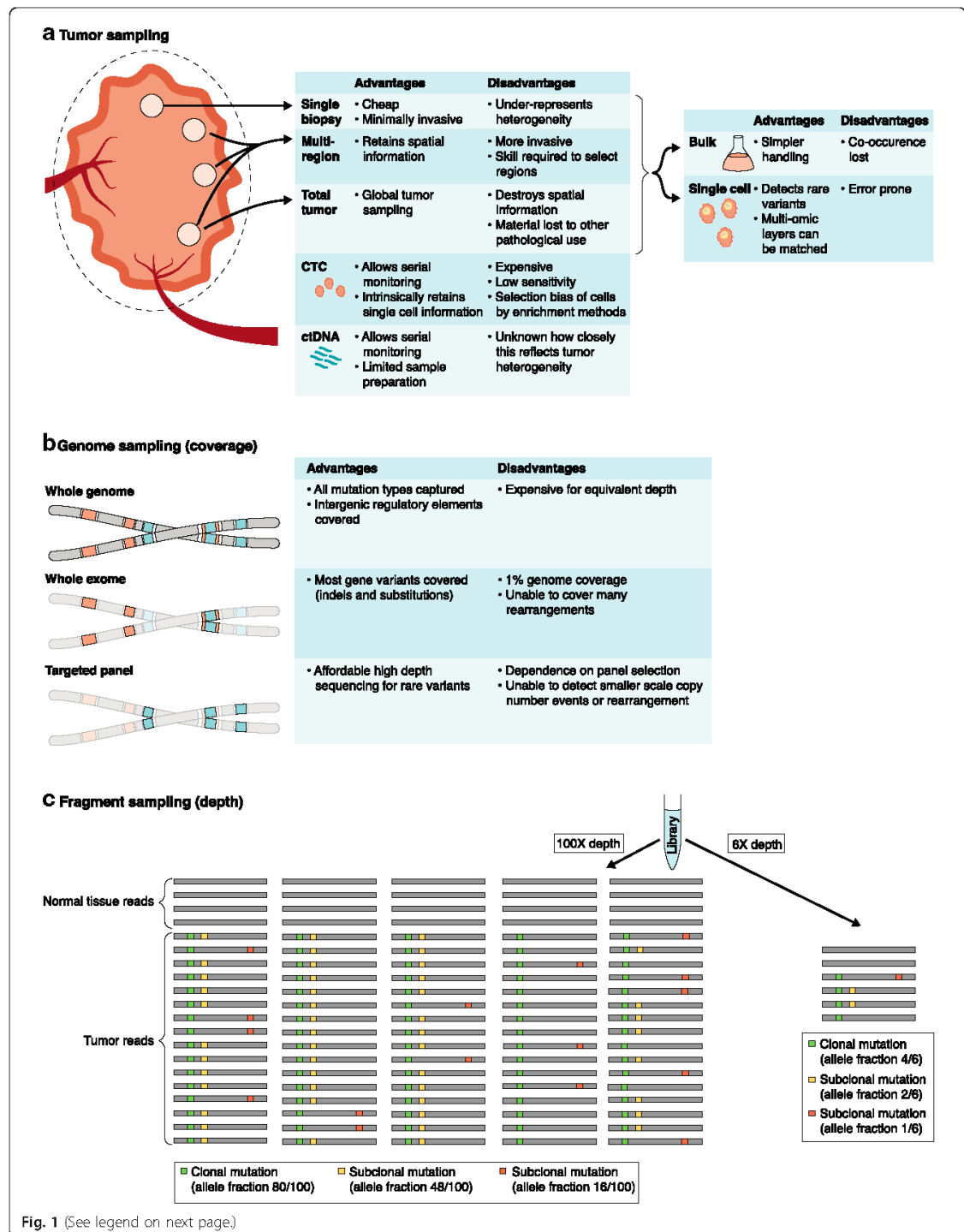


Fig. 1 (See legend on next page.)

(See figure on previous page.)

Fig. 1 Sampling decisions required for comprehensive and evolutionary description of tumors. Tumor genomic sampling can be considered to fall into three separate domains. **a** Sampling of tumor material, either directly from a tumor mass or shed into the circulation. Samples from the tumor mass can either be pooled as a bulk specimen or disaggregated into single cells. **b** Only portions of genomic material are sampled and assessed; either targeted panels of a few hundred genes can be used or the whole exome or whole genome can be profiled. **c** Bulk DNA extractions may contain millions of DNA molecules. These are contributed by different parental alleles from both tumor and normal cells. Samples frequently contain 10–80% normal cells. Library preparation and sequencing only samples a tiny fraction of the available DNA fragments. The schematic shows a representation of sampling at two different sequencing depths (100X and 6X) and illustrates how higher sequencing depths allow more accurate determinations of the frequencies of specific mutations and their clonal or subclonal status. *ctDNA* circulating tumor DNA

could in theory be eradicated. Previous targeted therapies have, to some degree, implicitly relied on the presumption that mutations that are highly prevalent in different tumors are probably early events in tumorigenesis and therefore likely to be clonal.

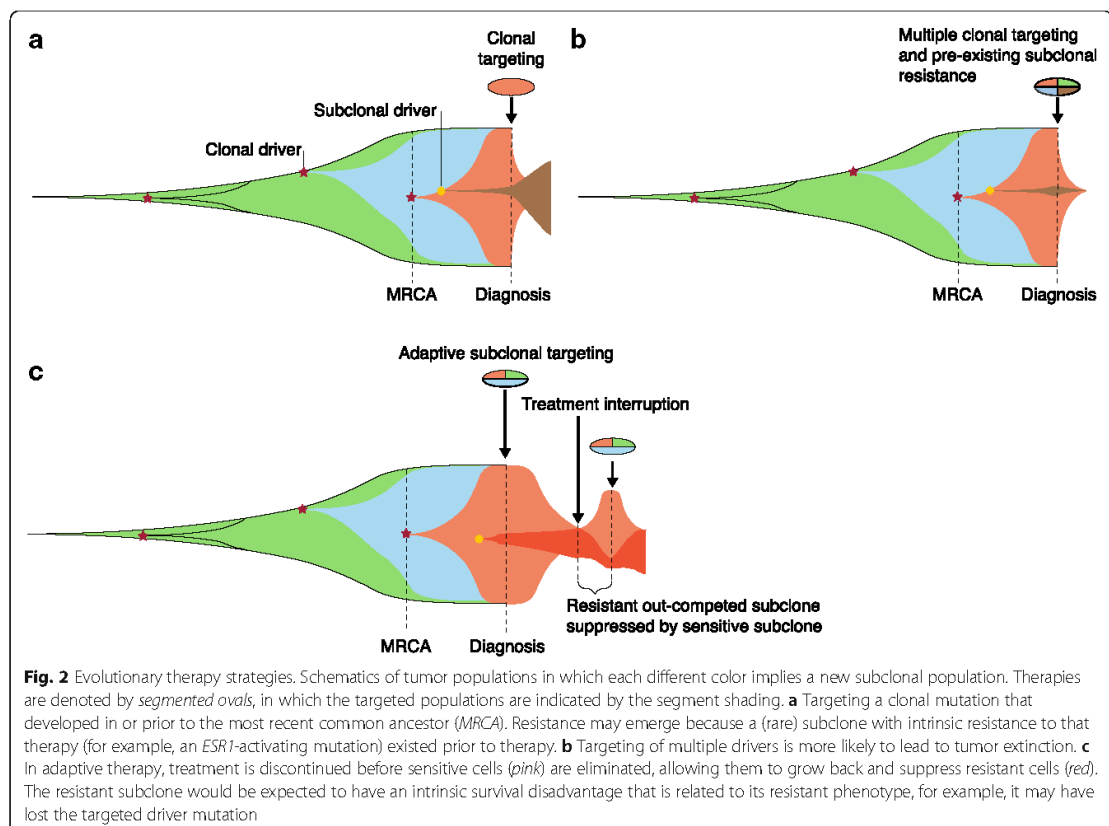
In most cases, single clonal mutations, which are thought to be functionally relevant driver mutations, have been targeted directly. In established cancers, this invariably results in the acquisition of treatment resistance. The simplest examples are the resistance to endocrine therapy in metastatic breast and prostate cancer. The mechanisms of these resistance phenomena are now relatively well understood. Many breast cancers depend on estrogen signaling and are initially sensitive to therapies that reduce the level of circulating estrogen or that target the cellular estrogen receptor, such as aromatase inhibitors or selective estrogen receptor modulators, respectively. Treatment resistance frequently arises when tumor cells develop constitutive activity in the estrogen receptor through mutation of its gene, *ESR1* [63]. Likewise, prostate cancers are almost ubiquitously driven by androgen signaling, sensitizing them to chemical or surgical castration. Prostate cancer cells compensate for medically depleted circulating androgen levels through a number of different mechanisms, including amplification of the androgen receptor [64]. Gundem et al. [31] demonstrated that multiple separate tumor cell populations, across distinct metastatic sites, can develop unique androgen receptor amplifications—a demonstration of convergent evolution. The widespread evolution of resistance suggests that clonal monotherapies are unlikely to achieve permanent tumor control or cure. For those with slow-paced advanced disease, or those who would not tolerate more intensive therapy, individual therapies will continue to play an important role. Most responses to targeted therapies, however, are both incomplete and short-lived and require improvement (Fig. 2a).

Even when a mutation is not treated directly, tumors can develop resistance. Synthetic lethality is a treatment approach that exploits a cellular vulnerability exposed by a clonal driver mutation. *BRCA* mutations in breast and ovarian cancer, both inherited or acquired, increase genomic instability due to disruption of the repair of double-strand DNA breaks, which not only produces variation during tumorigenesis but also increases the reliance

of these tumors on other DNA-repair mechanisms. This is exploited for therapy by inhibiting the single-stranded DNA repair PARP enzymes [65, 66]. PARP inhibition causes the accumulation of lethal DNA damage specifically in tumor cells. *BRCA* mutations can, however, undergo somatic reversal in multiple tumor subclones, leading to resistance to PARP inhibition [67–69].

Resistance to therapy typically results from mutations, which may pre-exist or can appear subsequent to the therapy, or from non-genetic factors. Mutations that exist prior to treatment exposure might be rare, and therefore undetectable by present assays. Once treatment creates selective pressure, resistant cells carrying these mutations will persist and become apparent. It is possible, and perhaps likely in larger tumors, that most resistance mutations exist prior to therapy exposure, even for conventional cytotoxic therapies [52]. Resistance mutations may also occur *de novo* after treatment exposure, perhaps having been induced by iatrogenic mutagenesis [70, 71]. Unless these mutations are of a distinct type, known to be induced by therapy, it is difficult to exclude the possibility that they did not exist prior to treatment at a very low and undetectable frequency. Resistance may also be non-genetic and either related to cell state plasticity or to a specific molecular resistance pathway, such as aurora kinase activation in anti-EGFR-treated lung cancer [72, 73], but further understanding of these non-heritable resistance mechanisms is needed. There are broad principles of treatment resistance that are common between cancer and infectious diseases [74]: like tumor cell populations, pathogens can be also be genetically heterogeneous [75], and as observed in the examples of HIV and *Mycobacterium tuberculosis*, they rarely have prolonged responses to monotherapy.

In principle, individual clonal therapies may still be used curatively if employed very early in tumor evolution, as proposed by Mitchell et al. [76]. Clear cell kidney cancers were modeled to have deleted *VHL* (on chromosome 3p) several decades prior to a second mutational hit to the remaining *VHL* allele. Proliferation and tumorigenesis only accelerate after both alleles of *VHL* are mutated. Therefore, the pool of mutated cells is probably only a few hundred cells for a prolonged period. Depleting this small cell population—even



marginally—with a therapy, sensitized by 3p loss, would reduce the probability of a cell with a second hit mutation ever arising. This would have to be achieved decades before these cells become detectable, and therefore would most likely involve the preemptive treatment of healthy individuals. However, considering that most tissues may harbor equivalent cell populations [8, 77, 78], such prophylactic management may not be clinically, economically, or ethically feasible.

Combined clonal therapies

Predictions of a tumor's evolutionary response to a therapy can allow pre-emptive measures to prevent resistance. For example, *ABL1* inhibition in chronic myeloid leukemia (CML), characterized by clonal *BCR-ABL1* fusions, has revolutionized therapy for this disease, yet the development of resistance remains a challenge in a proportion of patients. Combining different classes of *ABL1* inhibitors with mutually exclusive profiles of resistance mutations can preempt the emergence of resistant subclones (Fig. 2b). Preclinical application of this approach has resulted in durable responses [79].

Combining different clonal therapies might also reduce the emergence of resistance. Many breast cancers are thought to have cell-cycle dysregulation related to the cyclin-CDK-Rb pathway, in addition to estrogen sensitivity [80]. The addition of CDK4/6 inhibition to aromatase inhibition does indeed prolong the response in patients with metastatic disease. This delays the need for conventional cytotoxic therapy, but at the price of increased toxicity compared to endocrine therapy alone [81].

The development of effective combination therapies requires a comprehensive understanding of mutation clonality and resistance mechanisms. Metastatic melanomas frequently have activating mutations in the MAPK pathway, and resistance to BRAF inhibitors was thought to result from downstream MEK activation [82, 83]. Trials combining MEK and BRAF inhibition in melanoma have demonstrated modest clinical benefit [84, 85]; however, resistant tumors often have multiple different detectable MAPK mutations, suggesting convergent evolution [86].

Ideally, larger numbers of drivers could be targeted simultaneously or sequentially, depending on the pace and nature of the evolutionary response of the tumor.

Such combination therapies will impact toxicity management, although not always detrimentally. In fact, the addition of MEK inhibition to BRAF inhibitors reduces the cutaneous side-effects that are associated with BRAF inhibitors. The toxicities resulting from combination treatments may, however, require complex pharmacological adjustments that have implications for trial design, drug licensing and healthcare economic assessments.

Many tumors have only few clonal driver mutations and will require alternative strategies [87–89]. Effective pharmacological options for targeting driver mutations are also relatively limited. Some driver mutations may be treatable indirectly, either by collateral lethality, whereby susceptibilities created by the loss of genes adjacent to deleted tumor suppressors are harnessed, or by synthetic lethality [90, 91]. Alternatively, immunotherapy exploits the antigenicity of mutations, regardless of their driver status and without relying on the recurrence of mutations in different patients. Where durable clinical responses to immunotherapy have been seen, they are probably brought about by the simultaneous targeting of multiple clonal mutations. Indeed, one of the potential predictive markers of response to immune checkpoint blockade in non-small-cell lung cancer and melanoma is the clonal neoantigen load [92, 93]. If a common mechanism of resistance to an immunotherapy can occur, (epi) genetic variation and selection could drive tumors towards it, even when multi-pronged approaches are used. These mechanisms of immune editing are still a subject of intense study. They include an ability of tumors to reduce their antigen-presenting capability. In melanoma, lung, and ovarian cancer, these changes have been shown to result in part from either somatic (often subclonal) or germline loss of heterozygosity of the HLA locus [94–96]. Equivalent loss of expression of class II MHC may also result in treatment failure after allogeneic bone marrow transplant for acute myeloid leukemia [97]. Without a full and diverse HLA repertoire, many neoantigens cannot be successfully presented on the surface of tumor cells and therefore are not recognized by an adaptive immune response.

Targeting subclonal mutations

The detection of subclonal mutations is still an active research topic and therefore potential strategies for their therapeutic use are only conceptual at present. The simplest approach is to target a combination of multiple subclonal mutations, probably coupled with a clonal therapy. In rare circumstances, such as those recently suggested in pediatric brain tumors, subclonal populations might be highly functionally interdependent [98]. In these circumstances, even subclonal population depletion might have a profound effect on the tumor as a

whole. Alternatively, if the relative importance and the clinical impact of different subclonal populations can be measured, then those causing the greatest symptomatic burden could be prioritized. Implicit in this more strategic approach is the acceptance that other cell populations that cause lower symptomatic burden will not be eradicated, representing a shift to managing cancer as a chronic disease without the intent to cure [99].

A combination of conventional cross-sectional imaging with the monitoring of circulating markers could be used to identify spatially or mutationally distinct metastases. If lesions are spatially segregated, they may be amenable to local therapies: surgery, cryotherapy, focused ultrasound, or stereotactic radiotherapy. If they are characterized by treatable mutations, additional systemic therapies could be used. At present, proofs of this concept are yet to emerge.

Finally, the concept of adaptive therapy has also been proposed [100, 101]. Each of the subclones present in a tumor may be either sensitive or insensitive to a potential therapy. They compete for survival within the tumor environment and a mutation that confers resistance to a treatment, possibly through the loss or alteration of an oncogenic driver, might result in a growth disadvantage when that treatment agent is not present. With an adaptive approach, sensitive subclones can be treated to the point at which tumor size is reduced or growth is suppressed to achieve symptomatic benefit. Response may conceivably be monitored with a non-invasive surrogate biomarker, such as serum prostate-specific antigen (PSA) in prostate cancer. Thereafter, treatment can be reduced or withdrawn to allow the competitive suppression of resistant subclones (Fig. 2c). This approach is currently under evaluation in metastatic prostate cancer with the use of individualized PSA thresholds to guide the use of abiraterone, a CYP17A1 inhibitor [102]. To date, only small numbers of patients have been treated, albeit with good clinical outcome and reduced cumulative exposure to medication. It is worth noting that adaptive therapy is not the same as intermittent therapy, in which treatment may also be used discontinuously and with the monitoring of a biomarker, but without any individualization of treatment duration on the basis of response dynamics. For example, intermittent hormonal therapy has been attempted in prostate cancer. Crucially, trials such as TAP22 used fixed PSA thresholds rather than individualized thresholds [103, 104]. This could result in the depletion of treatment-sensitive clones, reducing their ability to suppress their treatment-resistant cousins.

Therapy monitoring: circulating tumor DNA and circulating tumor cells

Liquid biopsies sample more readily available body fluids, mainly blood, for cellular or genomic material

that has been shed from the tumor. They are heralded for reducing the invasiveness of clinical assays used for diagnosis [105, 106], prognosis [107, 108], molecular profiling [109], and response assessment [110–114]. Monitoring the treatment of more advanced disease may be substantially enhanced by monitoring the dynamics of different tumor cell populations.

The therapeutic approaches discussed above, particularly combination subclonal targeting and adaptive therapy, rely on accurate information about the relative importance of different subclonal populations in space and time. Liquid biopsies allow non-invasive assays that can easily be repeated over time. In particular, ctDNA is relatively stable and simple to handle, and its sequence content can be analyzed using a variety of approaches [115] (Fig. 1).

The detection of early subclinical relapse or minimal residual disease after attempted curative therapy has relied on detecting clonal mutations in circulation. Somatic structural variants are particularly amenable to highly disease-specific PCR-based approaches. Canonical disease-defining genomic rearrangements, such as the *BCR-ABL1* fusion in chronic myeloid leukemia, are routinely monitored in hematological malignancies to assess treatment response [116, 117]. Solid organ malignancies have fewer disease-defining rearrangements, but frequently possess unique somatic rearrangements that can be used to define bespoke monitoring panels [118, 119].

Monitoring of subclonal evolution has focused on evaluating somatic point mutations. Murtaza et al. [120] demonstrated that a dominant subclone, which was responsible for the progression of a chest wall breast cancer metastasis, was detectable by the increasing level of mutations private to that subclone. O'Leary et al. were able to use ctDNA in a small proportion of metastatic breast cancer patients, who were treated with the addition of the CDK4/6 inhibitor palbociclib, to both predict longer progression-free intervals [111] and detect emerging resistant subclones [121]. Furthermore, Abbosh et al. [30] showed that ctDNA was detectable 10–346 days (median 70 days) prior to clinical detection of relapsed lung cancer.

There are several challenges to the adoption of this approach. Clearly, bespoke ctDNA monitoring is costly. Abbosh et al. [30] estimated that even a limited bespoke monitoring panel, based on detected mutations from a single primary tumor region, would cost USD 1750 per patient. In addition, current analyses have only explored minimal numbers of detectable subclones and give an incomplete picture of their number and range. Whether there are substantial biases in the tumor cells that contribute circulating DNA is currently not known. It is likely that highly vascular and necrotic tumors will contribute more to ctDNA than tumors in cryptic sites, such as the central nervous system [122]. The use of other sources of

cell-free DNA, such as stool [123], urine, cerebrospinal fluid, and effusions, may in part compensate for this [124, 125]. There are also likely to be genomic biases because cell-free DNA is predominantly thought to be generated by apoptotic nuclease activity which produces nucleosome-associated DNA fragments [126, 127], resulting in distinct chromatin-associated patterns. These patterns and the degree of apoptosis are likely to vary across tumor cell populations, and result in a bias in circulating tumor DNA.

The detection of subclonal mutations is also limited by the sensitivity of detection assays. Next-generation sequencing approaches that seek to gain an unbiased view of all detectable variants in circulation cannot identify rare subclonal mutations. In the Murtaza et al. [120] study, even clonal mutations had variant allele fractions of 3.8–34.9%. To compensate for this, most approaches, as exemplified by Abbosh et al. [30], use a specific amplification method based on fixed expected mutations that are detected in a sequenced primary tumor. This, by definition, means that de novo mutations that arose subsequent to the sampling of the primary tumor will not be detectable in circulation.

Circulating tumor cells (CTCs) can be analyzed using single-cell sequencing approaches. In a study by Carter et al. [128], the copy number profile of circulating tumor cells at the time of diagnosis of small-cell lung cancer predicted the duration of response to chemotherapy. Cellular approaches are less likely to be confounded by the genomic aberrations that arise in other cells than the index tumor [129, 130]. As a result of the rarity of these tumor cells, they require significant enrichment which is likely to introduce biases, resulting in low sensitivity even for clonal tumor populations [131, 132]. Interestingly, Kwan et al. [133] demonstrated that after some initial filtration, an RNA expression-based signature can be used to detect breast cancer CTCs, and that the presence of these cells carried prognostic information in the setting of neoadjuvant chemotherapy.

Can insight into tumor evolution improve early diagnosis, risk stratification, and cancer prevention?

In order to improve cancer outcomes, it is essential to alter tumor evolution. This can be achieved throughout the evolutionary timeline by preventing etiological factors, screening cell populations on the path to cancer, or stratifying cancers that will pose the greatest threat.

Cancer screening

Cancer screening aims to reduce cancer mortality by increasing detection at a curable stage [134]. This needs to be carefully managed, however, as overtreatment of incidental findings causes unnecessary cost, harm, and

anxiety [135]. This problem has beset the introduction of a prostate cancer screening strategy, as many low-grade prostate cancers can be managed with observation alone [136, 137]. Reliable predictive biomarkers of progression in detected lesions could increase the utility of screening programs. To date, risk stratification has relied almost exclusively on histological staging and grading.

Methods are being developed that recapitulate the early evolution of cancers using sequencing information from later-stage cancers alone, as recently reviewed [42]. In general terms, these methods utilize the number of copies of mutations on gained chromosomal segments to infer whether these mutations happened before or after that gain. For example, if a whole chromosome has been duplicated and there are two copies of a mutation found on that chromosome, then it is likely that the mutation occurred first and was duplicated with the chromosomal gain. By analysis of whole-genome sequencing data from primary and metastatic prostate cancers, Wedge et al. [138] have been able to retrospectively identify chromosomal changes that developed earlier in tumorigenesis. These findings, such as the early gain of chromosome 8q, recapitulated those previously found in prostate intraepithelial neoplasia (PIN), which is thought to be a precursor of prostate adenocarcinoma [139].

These approaches have also been applied to invasive cancers, which have less well characterized precursor lesions [76, 89]. Recently, the Pan-Cancer Analysis of Whole Genomes (PCAWG) initiative leveraged whole-genome sequencing data to infer evolutionary timelines across cancer types [140]. This work reproduced and refined classic models of mutational progression such as for colorectal cancer, in which *APC* mutations precede *KRAS* and *TP53* mutations. This information could define mutations that can be used to risk-stratify those pre-malignant or early invasive lesions that require intervention and those that do not. In addition, large datasets and novel computational methods [141, 142] may be able to detect stereotyped evolutionary patterns and trajectories in cancer evolution that may inform early diagnosis or risk-stratification approaches.

Predicting tumor evolution—implications for risk stratification

A deeper and more comprehensive understanding of tumor evolution should allow us to understand how a cancer will behave in the future. This has specific implications for the risk stratification of established cancers. Incidental findings, such as small renal lesions that are often found during investigations for other conditions, are a clinical challenge because definitive resection is morbid but radiological and histological criteria are unreliable for prognostication [143]. In clear cell renal cell carcinoma, Turajlic et al. [87] have modeled that

analyses of two biopsies can allow the quantification of intra-tumor copy number heterogeneity. This can discriminate lesions of higher and lower risk of progression, thereby potentially assisting in the decision-making process for small renal lesions. In a companion study, the same authors also suggested that richer information gleaned from more thorough tumor sampling can identify evolutionary profiles that are more likely to be associated with the development of metastatic disease [88]. In other cancer types, patterns of heterogeneity, such as copy number diversity in lung cancer [89] and pan-mutational diversity (so-called regional ‘explosions’) in childhood cancers [144], have also been shown to carry prognostic information. More transformative change to cancer prediction strategies will require the development of more complex computational tools and models [141, 142]. Much as weather forecasting models require vast amounts of measured data from the real world, cancer evolution models will require the sequence-based profiling of the evolution of many more cancers. Ultimately, this will allow these forecasts to guide the optimal management for each patient.

Prevention of key early mutagenic processes

The identification of predisposing factors for cancer, whether heritable, environmental, or infectious, has previously relied on a combination of epidemiological and biological evidence. A deeper understanding of tumor evolution can lead to new insights into the impact of these factors on the genome.

Two clear examples of direct impact on the genome are ultraviolet (UV) radiation exposure for sun-induced cancers, such as cutaneous squamous cell cancers, and exposure to tobacco smoke carcinogens for smoking-related airway cancers. The epidemiological evidence for both has long been established, although its popular acceptance took some time [145]. Mechanisms of mutation as a result of each exposure have been identified: misrepair by transcription-coupled nucleotide excision repair of UV-induced pyrimidine photodimers [146] and misrepair of guanine damage by the same mechanism [147], respectively. These specific mutational types can now be detected across the genome as mutational signatures [70, 148], and this allows estimation of the contributions of each mutational signature (and potentially the level of mutagen exposure) in any individual tumor [149].

The accrual of mutations over time can now be explored retrospectively in a whole-genome-sequenced tumor. Nik-Zainal et al. [13] used a mutation timing approach to study changes in mutational processes over the life history of breast cancers. By leveraging the power of a large cohort of tumor samples, it becomes possible to identify mutational processes that act early or late in tumor evolution. In lung cancer, the

proportion of mutations bearing a smoking signature declines later in tumor evolution, despite ongoing smoke exposure [150, 151]. Conversely, mutagenesis that is related to the activity of the APOBEC family of cytidine deaminases increases later in lung tumor evolution. As expected, inherited defects in DNA repair, such as the deficient mismatch repair seen in Lynch syndrome, can lead to steady and ongoing mutation throughout a tumor's lifetime [152].

Many mutational signatures do not have identified etiologies, but direct genomic evidence can provide an objective starting point for both epidemiological and biological study. Identifying causative environmental exposures may suggest preventative measures, akin to smoking cessation and UV protection.

The challenge of somatic variation in normal tissues

The challenge in identifying mutations that are acquired early in tumorigenesis is that many canonical driver mutations, which are thought to be specific and relevant to cancer, may also occur in populations of phenotypically normal cells (Table 1).

Martincorena et al. [77, 129] identified multiple clonal expansions of cells, containing mutations in *TP53*, *NOTCH1*, and other known cancer genes, in both sun-exposed normal eyelids and in aging normal esophagus. Interestingly, mutations were much more common in *NOTCH1* than in *TP53* in normal esophagus, the inverse of the pattern seen in esophageal cancer, suggesting that early *NOTCH1* mutations may protect against cancer development. Demeulemeester et al. [130] analyzed epithelial cells found in bone marrow aspirates of breast cancer patients, identifying cells with copy number aberrations that were completely distinct from the primary breast cancer and therefore from an unknown origin. Gao et al. [153] also detected similar aberrant cells in tissue adjacent to breast tumors that were once again unrelated to tumor cells. Finally, clonal expansions of hematopoietic cells containing leukemia-associated mutations are reported in the circulation of otherwise healthy adults [154, 155]. These confer an increased risk of the subsequent development of a hematological malignancy, but clearly many do not progress [156, 157].

In order to truly reveal the early evolution of cancer, we will need to understand the frequency of these mutational events in the normal tissues in which cancers arise. Cataloguing mutational events in normal tissues, at rare frequencies, will help to identify the cells of origin of cancer as well as the early mutational steps that occur in these cells [158].

Conclusions and future perspectives

Intra-tumor heterogeneity and the ability of cancers to evolve continuously has proved a major challenge to the

implementation of precision anti-cancer medicine. Molecular therapies, predicted to be effective on the basis of the presence of a sensitizing mutation in a single sample, may be of limited clinical benefit. Driver mutations may be subclonal and resistance mechanisms can evolve rapidly [31, 89]. Deeper understanding of this complexity will allow the development of more robust therapeutic strategies. Without doubt, the complexity of tumor evolution is still far from being fully understood, and on an individual basis, tumors will always make unanticipated moves to evade even our best efforts. The recognition that cancer is an evolving system offers a framework on which to hang our clinical and research observations of cancer behavior and biology. We have discussed the more immediate opportunities for translating knowledge of tumor evolution here, but it seems likely that deeper insight will open additional unforeseen avenues.

Insight into the full spectrum of evolutionary paths that cancers can take may lead to the stratification of subsets of cancers that follow specific evolutionary paths. Potentially, the earliest steps or the rate-limiting steps in tumor evolution could be interrupted, either by the identification of preventable etiological factors or by timely medical interventions. These strategies may lead to a significant reduction in the incidence of some cancers or to a high cure rate in early diagnosed cancers, respectively. In addition, once diagnosed, treatment pathways may be matched according to the anticipated evolutionary path of the cancer, as opposed to classification based on traditional histological tumor subtyping. Patients with indolent tumors may be spared therapy altogether. As future therapies emerge, insight into tumor evolution is likely to inform their further development and maximize their impact. Immune checkpoint blockade is possibly the first class of therapy to emerge in this context, reaping the reward of a better understanding of the spectrum of clinical response [92, 94, 159]. Many cancers will probably need an armory of affordable, effective, and tolerable therapies that can be used safely in combination and sequentially. It is likely that conventional therapies—surgery, radiotherapy, and cytotoxic chemotherapy—will continue to have crucial roles in these treatment paradigms, but with a better understanding of the disease, these conventional therapies could be rationally combined with approaches informed by (epi) genomic insights into tumor evolution to achieve improved outcomes for cancer patients.

Abbreviations

CTC: Circulating tumor cell; ctDNA: Circulating tumor DNA; MCRA: Most recent common ancestor; PSA: Prostate-specific antigen

Funding

Research in the PVL lab is supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001202), the UK Medical Research Council (FC001202), and the Wellcome Trust (FC001202).

Authors' contributions

MWF and PVL wrote the manuscript, and both authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK. ²University College London Cancer Institute, 72 Huntley Street, London WC1E 6DD, UK. ³Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK. ⁴University of Leuven, Herestraat 49, B-3000 Leuven, Belgium.

Published online: 29 March 2019

References

1. Yeo B, Turner NC, Jones A. An update on the medical management of breast cancer. *BMJ*. 2014;348:g3608.
2. Rosell R, Karachaliou N. Optimizing lung cancer treatment approaches. *Nat Rev Clin Oncol*. 2014;12:75–6.
3. Nowell PC. The clonal evolution of tumor cell populations. *Science*. 1976; 194:23–8.
4. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet*. 2010;11:685–96.
5. Yates LR, Campbell PJ. Evolution of the cancer genome. *Nat Rev Genet*. 2012;13:795–806.
6. Baslan T, Hicks J. Unravelling biology and shifting paradigms in cancer with single-cell sequencing. *Nat Rev Cancer*. 2017;17:557–69.
7. Lynch M. Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad Sci U S A*. 2010;107:961–8.
8. Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. *Science*. 2015;349:1483–9.
9. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *New Engl J Med*. 2012;366:883–92.
10. Andor N, Graham TA, Jansen M, Xia LC, Aktipis CA, Petritsch C, et al. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat Med*. 2016;22:105–13.
11. Drento SC, Leshchiner I, Haase K, Tarabichi M, Wintersinger J, Deshwar AG, et al. Portraits of genetic intra-tumour heterogeneity and subclonal selection across cancer types. *bioRxiv*. 2018; <https://doi.org/10.1101/312041>.
12. McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell*. 2017;168:613–28.
13. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. *Cell*. 2012;149:994–1007.
14. Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*. 2011;144:27–40.
15. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, et al. Clock-like mutational processes in human somatic cells. *Nat Genet*. 2015;47:1402–7.
16. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, et al. Punctuated evolution of prostate cancer genomes. *Cell*. 2013;153:666–77.
17. Markowitz F. A saltationist theory of cancer evolution. *Nat Genet*. 2016;48:1102–3.
18. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. *Cell*. 2012;149:979–93.
19. Sottoriva A, Kang H, Ma Z, Graham TA, Salomon MP, Zhao J, et al. A big bang model of human colorectal tumor growth. *Nat Genet*. 2015;47:209–16.
20. Feinberg AP, Ohlsson R, Henikoff S. The epigenetic progenitor origin of human cancer. *Nat Rev Genet*. 2006;7:21–33.
21. Prensner JR, Chinnaiyan AM. The emergence of lncRNAs in cancer biology. *Cancer Discov*. 2011;1:391–407.
22. Hanash SM, Pitteri SJ, Faca VM. Mining the plasma proteome for cancer biomarkers. *Nature*. 2008;452:571–9.
23. Sheffield NC, Pierron G, Klughammer J, Datlinger P, Schönegger A, Schuster M, et al. DNA methylation heterogeneity defines a disease spectrum in Ewing sarcoma. *Nat Med*. 2017;23:386–95.
24. Li S, Garrett-Bakelman FE, Chung SS, Sanders MA, Hricik T, Rapaport F, et al. Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. *Nat Med*. 2016;22:792–9.
25. Pan H, Jiang Y, Boi M, Tabbò F, Redmond D, Nie K, et al. Epigenomic evolution in diffuse large B-cell lymphomas. *Nat Commun*. 2015;6:6921.
26. Landau DA, Clement K, Ziller MJ, Boyle P, Fan J, Gu H, et al. Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. *Cancer Cell*. 2014;26:813–25.
27. Baker A-M, Cross W, Curtius K, Al Bakir I, Choi C-HR, Davis HL, et al. Evolutionary history of human colitis-associated colorectal cancer. *Gut*. 2018. <https://doi.org/10.1136/gutjnl-2018-31619>.
28. Angelova M, Mlecnik B, Vasaturo A, Bindea G, Fredriksen T, Lafontaine L, et al. Evolution of metastases in space and time under immune selection. *Cell*. 2018;175:751–65.
29. Cooper CS, Eeles R, Wedge DC, Van Loo P, Gundem G, Alexandrov LB, et al. Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nat Genet*. 2015;47:367–72.
30. Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature*. 2017;545:446–51.
31. Gundem G, Van Loo P, Kremeyer B, Alexandrov LB, Tubio JM, Papaemmanuil E, et al. The evolutionary history of lethal metastatic prostate cancer. *Nature*. 2015;520:353–7.
32. Campbell PJ, Yachida S, Mudie LJ, Stephens PJ, Pleasance ED, Stebbings LA, et al. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature*. 2010;467:1109–13.
33. Yates LR, Knappskog S, Wedge D, Farmery JHR, Gonzalez S, Martincorena I, et al. Genomic evolution of breast cancer metastasis and relapse. *Cancer Cell*. 2017;32:169–84.
34. Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell*. 2013;152:714–26.
35. Rocco JW. Mutant allele tumor heterogeneity (MATH) and head and neck squamous cell carcinoma. *Head Neck Pathol*. 2015;9:1–5.
36. Yates LR, Gerstung M, Knappskog S, Desmedt C, Gundem G, Van Loo P, et al. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat Med*. 2015;21:751–9.
37. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell*. 1990;61:759–67.
38. Shain AH, Joseph NM, Yu R, Benhamida J, Liu S, Prow T, et al. Genomic and transcriptomic analysis reveals incremental disruption of key signaling pathways during melanoma evolution. *Cancer Cell*. 2018;34:45–55.
39. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, et al. Genetic alterations during colorectal-tumor development. *New Engl J Med*. 1988;319:525–32.
40. Teixeira VH, Pipinikas CP, Pennycuik A, Lee-Six H, Chandrasekharan D, Beane J, et al. Deciphering the genomic, epigenomic, and transcriptomic landscapes of pre-invasive lung cancer lesions. *Nat Med*. 2019;25:517–25.
41. Drento SC, Wedge DC, Van Loo P. Principles of reconstructing the subclonal architecture of cancers. *Cold Spring Harb Perspect Med*. 2017;7:a026625.
42. Jolly C, Van Loo P. Timing somatic events in the evolution of cancer. *Genome Biol*. 2018;19:95.
43. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*. 2012;481:506–10.
44. Cortes JE, Talpaz M, Giles F, O'Brien S, Rios MB, Shan J, et al. Prognostic significance of cytogenetic clonal evolution in patients with chronic myelogenous leukemia on imatinib mesylate therapy. *Blood*. 2003;101:3794–800.
45. Jan M, Snyder TM, Corces-Zimmerman MR, Vyas P, Weissman IL, Quake SR, et al. Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci Transl Med*. 2012;4:149a118.
46. Schuh A, Becq J, Humphray S, Alexa A, Burns A, Clifford R, et al. Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. *Blood*. 2012;120:4191–6.
47. Wang Y, Navin NE. Advances and applications of single-cell sequencing technologies. *Mol Cell*. 2015;58:598–609.

48. Navin N, Hicks J. Future medical applications of single-cell sequencing in cancer. *Genome Med.* 2011;3:31.
49. Van Loo P, Voet T. Single cell analysis of cancer genomes. *Curr Opin Genet Dev.* 2014;24:82–91.
50. Navin NE. Cancer genomics: one cell at a time. *Genome Biol.* 2014;15:452.
51. Zheng GK, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun.* 2017;8:14049.
52. Kim C, Gao R, Sei E, Brandt R, Hartman J, Hatschek T, et al. Chemoresistance evolution in triple-negative breast cancer delineated by single-cell sequencing. *Cell.* 2018;173:879–93.
53. Zahn H, Steif A, Laks E, Eirew P, VanInsberghe M, Shah SP, et al. Scalable whole-genome single-cell library preparation without preamplification. *Nat Methods.* 2017;14:167–73.
54. Macaulay IC, Haerty W, Kumar P, Li Yi, Hu TX, Teng MJ, et al. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat Methods.* 2015;12:519–22.
55. Macaulay IC, Ponting CP, Voet T. Single-cell multiomics: multiple measurements from single cells. *Trends Genet.* 2017;33:155–68.
56. Dey SS, Kester L, Spanjaard B, Bienko M, van Oudenaarden A. Integrated genome and transcriptome sequencing of the same cell. *Nat Biotechnol.* 2015;33:285–9.
57. Ryser MD, Yu M, Grady W, Siegmund K, Shibata D. Epigenetic heterogeneity in human colorectal tumors reveals preferential conservation and evidence of immune surveillance. *Sci Rep.* 2018;8:17292.
58. Bian S, Hou Y, Zhou X, Li X, Yong J, Wang Y, et al. Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science.* 2018;362:1060–3.
59. Maley CC, Aktipis A, Graham TA, Sottoriva A, Boddy AM, Janiszewska M, et al. Classifying the evolutionary and ecological features of neoplasms. *Nat Rev Cancer.* 2017;17:605–19.
60. Davis C, Naci H, Gurpinar E, Poplavska E, Pinto A, Aggarwal A. Availability of evidence of benefits on overall survival and quality of life of cancer drugs approved by European medicines agency: retrospective cohort study of drug approvals 2009–13. *BMJ.* 2017;359:j4530.
61. McArthur GA, Chapman PB, Robert C, Larkin J, Haanen JB, Dummer R, et al. Safety and efficacy of vemurafenib in BRAF(V600E) and BRAF(V600K) mutation-positive melanoma (BRIM-3): extended follow-up of a phase 3, randomised, open-label study. *Lancet Oncol.* 2014;15:323–32.
62. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol.* 2012;13:239–46.
63. Jeselsohn R, Buchwalter G, De Angelis C, Brown M, Schiff R. ESR1 mutations—a mechanism for acquired endocrine resistance in breast cancer. *Nat Rev Clin Oncol.* 2015;12:573–83.
64. Watson PA, Arora VK, Sawyers CL. Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nat Rev Cancer.* 2015;15:701–11.
65. Lord CJ, Ashworth A. BRCAness revisited. *Nat Rev Cancer.* 2016;16:110–20.
66. Ashworth A. A synthetic lethal therapeutic approach: poly (ADP) ribose polymerase inhibitors for the treatment of cancers deficient in DNA double-strand break repair. *J Clin Oncol.* 2008;26:3785–90.
67. Patch AM, Christie EL, Etemadmoghadam D, Garsed DW, George J, Fereday S, et al. Whole-genome characterization of chemoresistant ovarian cancer. *Nature.* 2015;521:489–94.
68. Weigelt B, Comino-Mendez I, de Bruijn I, Tian L, Meisel JL, Garcia-Murillas I, et al. Diverse BRCA1 and BRCA2 reversion mutations in circulating cell-free DNA of therapy-resistant breast or ovarian cancer. *Clin Cancer Res.* 2017;23:6708–20.
69. Christie EL, Fereday S, Doig K, Pattnaik S, Dawson SJ, Bowtell DDL. Reversion of BRCA1/2 germline mutations detected in circulating tumor DNA from patients with high-grade serous ovarian cancer. *J Clin Oncol.* 2017;35:1274–80.
70. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature.* 2013;500:415–21.
71. Szikriszt B, Póti Á, Pipek O, Krzystanek M, Kanu N, Molnár J, et al. A comprehensive survey of the mutagenic impact of common cancer cytotoxics. *Genome Biol.* 2016;17:99.
72. Sharma A, Cao EY, Kumar V, Zhang X, Leong HS, Wong AML, et al. Longitudinal single-cell RNA sequencing of patient-derived primary cells reveals drug-induced infidelity in stem cell hierarchy. *Nat Commun.* 2018;9:4931.
73. Shah KN, Bhatt R, Rotow J, Rohrberg J, Olivás V, Wang VE, et al. Aurora kinase a drives the evolution of resistance to third-generation EGFR inhibitors in lung cancer. *Nat Med.* 2019;25:111–8.
74. Glickman MS, Sawyers CL. Converting cancer therapies into cures: lessons from infectious diseases. *Cell.* 2012;148:1089–98.
75. Stewart PS, Franklin MJ. Physiological heterogeneity in biofilms. *Nat Rev Microbiol.* 2008;6:199–210.
76. Mitchell TJ, Turajlic S, Rowan A, Nicol D, Farmery JHR, O'Brien T, et al. Timing the landmark events in the evolution of clear cell renal cell cancer: TRACERx renal. *Cell.* 2018;173:611–23.
77. Martincorena I, Fowler JC, Wabik A, Lawson ARJ, Abascal F, Hall MWJ, et al. Somatic mutant clones colonize the human esophagus with age. *Science.* 2018;362:911–7.
78. Yokoyama A, Kakiuchi N, Yoshizato T, Nannya Y, Suzuki H, Takeuchi Y, et al. Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature.* 2019;565:312–7.
79. Wylie AA, Schoepfer J, Jahnke W, Cowan-Jacob SW, Loo A, Furet P, et al. The allosteric inhibitor ABL001 enables dual targeting of BCR-ABL1. *Nature.* 2017;543:733–7.
80. Dean JL, Thangavel C, McClendon AK, Reed CA, Knudsen ES. Therapeutic CDK4/6 inhibition in breast cancer: key mechanisms of response and failure. *Oncogene.* 2010;29:4018–32.
81. Cristofanilli M, Tumer NC, Bondarenko I, Ro J, Im SA, Masuda N, et al. Fulvestrant plus palbociclib versus fulvestrant plus placebo for treatment of hormone-receptor-positive, HER2-negative metastatic breast cancer that progressed on previous endocrine therapy (PALOMA-3): final analysis of the multicentre, double-blind, phase 3 randomised controlled trial. *Lancet Oncol.* 2016;17:425–39.
82. Nazarian R, Shi H, Wang Q, Kong X, Koya RC, Lee H, et al. Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature.* 2010;468:973–7.
83. Johannessen CM, Boehm JS, Kim SY, Thomas SR, Wardwell L, Johnson LA, et al. COT drives resistance to RAF inhibition through MAP kinase pathway reactivation. *Nature.* 2010;468:968–72.
84. Long GV, Stroyakovskiy D, Gogas H, Levchenko E, de Braud F, Larkin J, et al. Combined BRAF and MEK inhibition versus BRAF inhibition alone in melanoma. *New Engl J Med.* 2014;371:1877–88.
85. Moriceau G, Hugo W, Hong A, Shi H, Kong X, Yu CC, et al. Tunable-combinatorial mechanisms of acquired resistance limit the efficacy of BRAF/MEK cotargeting but result in melanoma drug addiction. *Cancer Cell.* 2015;27:240–56.
86. Wagle N, Van Allen EM, Treacy DJ, Frederick DT, Cooper ZA, Taylor-Weiner A, et al. MAP kinase pathway alterations in BRAF-mutant melanoma patients with acquired resistance to combined RAF/MEK inhibition. *Cancer Discov.* 2014;4:661–8.
87. Turajlic S, Xu H, Litchfield K, Rowan A, Horswell S, Chambers T, et al. Deterministic evolutionary trajectories influence primary tumor growth: TRACERx renal. *Cell.* 2018;173:595–610.
88. Turajlic S, Xu H, Litchfield K, Rowan A, Chambers T, Lopez JL, et al. Tracking cancer evolution reveals constrained routes to metastases: TRACERx renal. *Cell.* 2018;173:581–94.
89. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al. Tracking the evolution of non-small-cell lung cancer. *New Engl J Med.* 2017;376:2109–21.
90. Muller FL, Aquilanti EA, DePinho RA. Collateral lethality: a new therapeutic strategy in oncology. *Trends Cancer.* 2015;1:161–73.
91. Muller FL, Colla S, Aquilanti E, Manzo VE, Genovese G, Lee J, et al. Passenger deletions generate therapeutic vulnerabilities in cancer. *Nature.* 2012;488:337–42.
92. McGranahan N, Furness AJ, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science.* 2016;351:1463–9.
93. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *New Engl J Med.* 2014;371:2189–99.
94. McGranahan N, Rosenthal R, Hiley CT, Rowan AJ, Watkins TBK, Wilson GA, et al. Allele-specific HLA loss and immune escape in lung cancer evolution. *Cell.* 2017;171:1259–71.
95. Chowell D, Morris LGT, Grigg CM, Weber JK, Samstein RM, Makarov V, et al. Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science.* 2018;359:582–7.
96. Zhang AW, McPherson A, Milne K, Kroeger DR, Hamilton PT, Miranda A, et al. Interfaces of malignant and immunologic clonal dynamics in ovarian cancer. *Cell.* 2018;173:1755–69.

97. Christopher MJ, Petti AA, Rettig MP, Miller CA, Chendamarai E, Duncavage EJ, et al. Immune escape of relapsed AML cells after allogeneic transplantation. *New Engl J Med*. 2018;379:2330–41.
98. Vinci M, Burford A, Molinari V, Kessler K, Popov S, Clarke M, et al. Functional diversity and cooperativity between subclonal populations of pediatric glioblastoma and diffuse intrinsic pontine glioma cells. *Nat Med*. 2018;24:1204–15.
99. Beck S, Ng T. C2c: turning cancer into chronic disease. *Genome Med*. 2014;6:38.
100. Enriquez-Navas PM, Kam Y, Das T, Hassan S, Silva A, Foroutan P, et al. Exploiting evolutionary principles to prolong tumor control in preclinical models of breast cancer. *Sci Transl Med*. 2016;8:327ra24.
101. Gatenby RA, Silva AS, Gillies RJ, Frieden BR. Adaptive therapy. *Cancer Res*. 2009;69:4894–903.
102. Zhang J, Cunningham JJ, Brown JS, Gatenby RA. Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. *Nat Commun*. 2017;8:1816.
103. Mottet N, Van Damme J, Loulidi S, Russel C, Leitenberger A, Wolff JM, et al. Intermittent hormonal therapy in the treatment of metastatic prostate cancer: a randomized trial. *BJU Int*. 2012;110:1262–9.
104. Magnan S, Zarychanski R, Pilote L, Bemier L, Shemilt M, Vigneault E, et al. Intermittent vs continuous androgen deprivation therapy for prostate cancer: a systematic review and meta-analysis. *JAMA Oncol*. 2015;1:1261–9.
105. Newman AM, Bratman SV, To J, Wynne JF, Eclow NC, Modlin LA, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med*. 2014;20:548–54.
106. Beaver JA, Jelovac D, Balukrishna S, Cochran R, Croessmann S, Zabransky DJ, et al. Detection of cancer DNA in plasma of patients with early-stage breast cancer. *Clin Cancer Res*. 2014;20:2643–50.
107. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med*. 2014;6:224ra24.
108. Khan KH, Cunningham D, Wemer B, Vlachogiannis G, Spiteri I, Heide T, et al. Longitudinal liquid biopsy and mathematical modeling of clonal evolution forecast time to treatment failure in the PROSPECT-C phase II colorectal cancer clinical trial. *Cancer Discov*. 2018;8:1270–85.
109. Oxnard GR, Thress KS, Alden RS, Lawrance R, Paweletz CP, Cantarini M, et al. Association between plasma genotyping and outcomes of treatment with osimertinib (AZD9291) in advanced non-small-cell lung cancer. *J Clin Oncol*. 2016;34:3375–82.
110. Dawson SJ, Tsui DW, Murtaza M, Biggs H, Rueda OM, Chin SF, et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. *New Engl J Med*. 2013;368:1199–209.
111. O'Leary B, Hrebien S, Morden JP, Beaney M, Friibens C, Huang X, et al. Early circulating tumor DNA dynamics and clonal selection with palbociclib and fulvestrant for breast cancer. *Nat Commun*. 2018;9:896.
112. Spina V, Brusca G, Cuccaro A, Martini M, Di Trani M, Forestieri G, et al. Circulating tumor DNA reveals genetics, clonal evolution, and residual disease in classical Hodgkin lymphoma. *Blood*. 2018;131:2413–25.
113. Barault L, Amatu A, Siravegna G, Ponzetti A, Moran S, Cassingena A, et al. Discovery of methylated circulating DNA biomarkers for comprehensive non-invasive monitoring of treatment response in metastatic colorectal cancer. *Gut*. 2018;67:1995–2005.
114. Agarwal R, Chan Y-C, Tam CS, Hunter T, Vassiliadis D, Teh CE, et al. Dynamic molecular monitoring reveals that SWI-SNF mutations mediate resistance to ibrutinib plus venetoclax in mantle cell lymphoma. *Nat Med*. 2019;25:119–29.
115. Henao Diaz E, Yachnin J, Gronberg H, Lindberg J. The in vitro stability of circulating tumour DNA. *PLoS One*. 2016;11:e0168153.
116. Marin D, Kaeda J, Szydlo R, Saunders S, Fleming A, Howard J, et al. Monitoring patients in complete cytogenetic remission after treatment of CML in chronic phase with imatinib: patterns of residual leukaemia and prognostic factors for cytogenetic relapse. *Leukemia*. 2005;19:507–12.
117. Branford S, Rudzki Z, Parkinson I, Grigg A, Taylor K, Seymour JF, et al. Real-time quantitative PCR analysis can be used as a primary screen to identify patients with CML treated with imatinib who have BCR-ABL kinase domain mutations. *Blood*. 2004;104:2926–32.
118. McBride DJ, Orpana AK, Sotiropoulos C, Joensuu H, Stephens PJ, Mudie LJ, et al. Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. *Genes Chromosomes Cancer*. 2010;49:1062–9.
119. Olsson E, Winter C, George A, Chen Y, Howlin J, Tang MH, et al. Serial monitoring of circulating tumor DNA in patients with primary breast cancer for detection of occult metastatic disease. *EMBO Mol Med*. 2015;7:1034–47.
120. Murtaza M, Dawson SJ, Pogrebniak K, Rueda OM, Provenzano E, Grant J, et al. Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. *Nat Commun*. 2015;6:8760.
121. O'Leary B, Cutts RJ, Liu Y, Hrebien S, Huang X, Fenwick K, et al. The genetic landscape and clonal evolution of breast cancer resistance to palbociclib plus fulvestrant in the PALOMA-3 trial. *Cancer Discov*. 2018;8:1390–403.
122. De Mattos-Arruda L, Mayor R, Ng CK, Weigelt B, Martinez-Ricarte F, Torrejon D, et al. Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. *Nat Commun*. 2015;6:8839.
123. Imperiale TF, Ransohoff DF, Itzkowitz SH. Multitarget stool DNA testing for colorectal-cancer screening. *New Engl J Med*. 2014;371:187–8.
124. Peng M, Chen C, Hulbert A, Brock MW, Yu F. Non-blood circulating tumor DNA detection in cancer. *Oncotarget*. 2017;8:69162–73.
125. Miller AM, Shah RH, Pentsova EJ, Poumaleki M, Briggs S, Distefano N, et al. Tracking tumour evolution in glioma through liquid biopsies of cerebrospinal fluid. *Nature*. 2019;565:654–8.
126. Thierry AR, Moulriere F, Gongora C, Ollier J, Robert B, Ychou M, et al. Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts. *Nucleic Acids Res*. 2010;38:6159–75.
127. Giacomini MB, Ruben GC, Iczkowski KA, Roos TB, Porter DM, Sorenson GD. Cell-free DNA in human blood plasma: length measurements in patients with pancreatic cancer and healthy controls. *Pancreas*. 1998;17:89–97.
128. Carter L, Rothwell DG, Mesquita B, Smowton C, Leong HS, Fernandez-Gutierrez F, et al. Molecular analysis of circulating tumor cells identifies distinct copy-number profiles in patients with chemosensitive and chemorefractory small-cell lung cancer. *Nat Med*. 2017;23:114–9.
129. Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S, et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science*. 2015;348:880–6.
130. Demeulemeester J, Kumar P, Moller EK, Nord S, Wedge DC, Peterson A, et al. Tracing the origin of disseminated tumor cells in breast cancer using single-cell sequencing. *Genome Biol*. 2016;17:250.
131. Alix-Panabieres C, Pantel K. Challenges in circulating tumour cell research. *Nat Rev Cancer*. 2014;14:623–31.
132. Krebs MG, Metcalf RL, Carter L, Brady G, Blackhall FH, Dive C. Molecular analysis of circulating tumour cells-biology and biomarkers. *Nat Rev Clin Oncol*. 2014;11:129–44.
133. Kwan TT, Bardia A, Spring LM, Giobbie-Hurder A, Kalinich M, Dubash T, et al. A digital RNA signature of circulating tumor cells predicting early therapeutic response in localized and metastatic breast cancer. *Cancer Discov*. 2018;8:1286–99.
134. Gray JA. New concepts in screening. *Br J Gen Pract*. 2004;54:292–8.
135. Esserman LJ, Thompson IM Jr, Reid B. Overdiagnosis and overtreatment in cancer: an opportunity for improvement. *JAMA*. 2013;310:797–8.
136. Eggener SE, Gfu AS, Nabhan C. Prostate cancer screening. *JAMA*. 2015;314:825–6.
137. Hugosson J, Carlsson S, Aus G, Bergdahl S, Khatami A, Lodding P, et al. Mortality results from the Goteborg randomised population-based prostate-cancer screening trial. *Lancet Oncol*. 2010;11:725–32.
138. Wedge DC, Gundem G, Mitchell T, Woodcock DJ, Martincorena I, Ghori M, et al. Sequencing of prostate cancers identifies new cancer genes, routes of progression and drug targets. *Nat Genet*. 2018;50:682–92.
139. Jung SH, Shin S, Kim MS, Baek IP, Lee JY, Lee SH, et al. Genetic progression of high grade prostatic intraepithelial neoplasia to prostate cancer. *Eur Urol*. 2016;69:823–30.
140. Gerstung M, Jolly C, Leshchiner I, Dentro SC, Gonzalez S, Mitchell TJ, et al. The evolutionary history of 2,658 cancers. *bioRxiv*. 2018; doi: <https://doi.org/10.1101/161562>.
141. Caravagna G, Giarratano Y, Ramazzotti D, Tomlinson I, Graham TA, Sanguinetti G, et al. Detecting repeated cancer evolution from multi-region tumor sequencing data. *Nat Methods*. 2018;15:707–14.
142. Gerhauser C, Favero F, Risch T, Simon R, Feuerbach L, Assenov Y, et al. Molecular evolution of early-onset prostate cancer identifies molecular risk markers and clinical trajectories. *Cancer Cell*. 2018;34:996–1011.
143. Ljungberg B, Bensalah K, Canfield S, Dabestani S, Hofmann F, Hora M, et al. EAU guidelines on renal cell carcinoma: 2014 update. *Eur Urol*. 2015;67:913–24.
144. Carlsson J, Valind A, Holmquist Mengelbier L, Bredin S, Cormark L, Jansson C, et al. Four evolutionary trajectories underlie genetic intratumoral variation in childhood cancer. *Nat Genet*. 2018;50:944–50.
145. Doll R, Hill AB. Smoking and carcinoma of the lung; preliminary report. *BMJ*. 1950;2:739–48.

146. Brash DE, Haseltine WA. UV-induced mutation hotspots occur at DNA damage hotspots. *Nature*. 1982;298:189–92.
147. Denissenko MF, Pao A, Tang M, Pfeifer GP. Preferential formation of benzo [a] pyrene adducts at lung cancer mutational hotspots in P53. *Science*. 1996;274:430–2.
148. Alexandrov LB, Ju YS, Haase K, Van Loo P, Martincorena I, Nik-Zainal S, et al. Mutational signatures associated with tobacco smoking in human cancer. *Science*. 2016;354:618–22.
149. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*. 2016;534:47–54.
150. Swanton C, McGranahan N, Starrett GJ, Harris RS. APOBEC enzymes: mutagenic fuel for cancer evolution and heterogeneity. *Cancer Discov*. 2015;5:704–12.
151. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science*. 2014;346:251–6.
152. Campbell BB, Light N, Fabrizio D, Zatzman M, Fuligni F, De Borja R, et al. Comprehensive analysis of hypermutation in human cancer. *Cell*. 2017;171:1042–56.
153. Gao R, Davis A, McDonald TO, Sei E, Shi X, Wang Y, et al. Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat Genet*. 2016;48:1119–30.
154. Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med*. 2014;20:1472–8.
155. Genovese G, Kahler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *New Engl J Med*. 2014;371:2477–87.
156. Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC, et al. The origin and evolution of mutations in acute myeloid leukemia. *Cell*. 2012;150:264–78.
157. Gibson CJ, Steensma DP. New insights from studies of clonal hematopoiesis. *Clin Cancer Res*. 2018;24:4633–42.
158. Young MD, Mitchell TJ, Vieira Braga FA, Tran MGB, Stewart BJ, Ferdinand JR, et al. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science*. 2018;361:594–9.
159. Snyder A, Nathanson T, Funt SA, Ahuja A, Buros Novik J, Hellmann MD, et al. Contribution of systemic and somatic factors to clinical response and resistance to PD-L1 blockade in urothelial cancer: an exploratory multi-omic analysis. *PLoS Med*. 2017;14:e1002309.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



ARTICLE

DOI: 10.1038/s41467-018-04530-z

OPEN

Recurrent rearrangements of *FOS* and *FOSB* define osteoblastoma

Matthew W. Fittall^{1,2,3}, William Mifsud^{2,3,4}, Nischalan Pillay^{2,5}, Hongtao Ye⁵, Anna-Christina Strobl⁵, Annelien Verfaillie¹, Jonas Demeulemeester^{1,6}, Lei Zhang⁷, Fitim Berisha⁵, Maxime Tarabichi^{1,3}, Matthew D. Young³, Elena Miranda², Patrick S. Tarpey³, Roberto Tirabosco⁵, Fernanda Amary⁵, Agamemnon E. Grigoriadis⁸, Michael R. Stratton³, Peter Van Loo^{1,6}, Cristina R. Antonescu⁷, Peter J. Campbell³, Adrienne M. Flanagan^{2,5} & Sam Behjati^{3,9}

The transcription factor *FOS* has long been implicated in the pathogenesis of bone tumours, following the discovery that the viral homologue, *v-fos*, caused osteosarcoma in laboratory mice. However, mutations of *FOS* have not been found in human bone-forming tumours. Here, we report recurrent rearrangement of *FOS* and its paralogue, *FOSB*, in the most common benign tumours of bone, osteoblastoma and osteoid osteoma. Combining whole-genome DNA and RNA sequences, we find rearrangement of *FOS* in five tumours and of *FOSB* in one tumour. Extending our findings into a cohort of 55 cases, using FISH and immunohistochemistry, provide evidence of ubiquitous mutation of *FOS* or *FOSB* in osteoblastoma and osteoid osteoma. Overall, our findings reveal a human bone tumour defined by mutations of *FOS* and *FOSB*.

¹The Francis Crick Institute, London NW1 1AT, UK. ²University College London Cancer Institute, London WC1E 6DD, UK. ³Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, UK. ⁴Great Ormond Street Hospital for Children NHS Foundation Trust, London WC1N 3JH, UK. ⁵Department of Histopathology, Royal National Orthopaedic Hospital NHS Trust, Stanmore, Middlesex HA7 4LP, UK. ⁶Department of Human Genetics, University of Leuven, Leuven 3000, Belgium. ⁷Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. ⁸Centre for Craniofacial and Regenerative Biology, King's College London, Guy's Hospital, London SE1 9RT, UK. ⁹Department of Paediatrics, University of Cambridge, Cambridge CB2 0QQ, UK. These authors contributed equally: Matthew W. Fittall, William Mifsud. These authors jointly supervised this work: Adrienne M. Flanagan, Sam Behjati. Correspondence and requests for materials should be addressed to A.M.F. (email: a.flanagan@ucl.ac.uk) or to S.B. (email: sb31@sanger.ac.uk)

Osteoblastoma is the most common benign bone-forming tumour. It typically occurs in the medulla of long bones and the neural arch from where it may extend into the vertebral body¹. Osteoid osteoma is thought to represent a variant of osteoblastoma. The two entities are distinguished arbitrarily by size, with osteoblastoma measuring more than 2 cm in diameter. Large, inaccessible and recurrent tumours can cause considerable morbidity¹. Treatment is by surgical resection. The genetic changes underpinning osteoblastoma have been studied at the resolution of karyotypes and copy number arrays. Copy number losses involving chromosome 22 and rearrangements involving chromosome 14 have been reported in rare cases only^{2,3}.

Diagnosis of osteoblastoma is currently based on histological assessment. Occasionally this can be challenging, as osteoblastoma has to be distinguished from osteoblastic osteosarcoma, an aggressive bone cancer that requires extensive, sometimes disabling, multimodal treatment⁴.

Here, we sought to define the somatic changes that underpin osteoblastoma. Our starting point was a discovery cohort of six tumours, five osteoblastomas and one osteoid osteoma, that we subjected to RNA and whole-genome DNA sequencing. Tissue was obtained from frozen tumour and corresponding germline DNA sequences derived from blood samples. Using the analysis pipeline of the Cancer Genome Project ('Methods'), we generated catalogues of all classes of somatic mutations: substitutions, indels, structural variants (rearrangements) and copy number changes. Transcriptome sequences were analysed to corroborate DNA changes and to call gene fusions.

Our key finding was recurrent, disease-defining structural variation of the *FOS* and *FOSB* oncogenes in osteoblastoma and osteoid osteoma.

Results

Osteoblastoma harbours few somatic alterations. Overall, there was a paucity of somatic alterations in osteoblastoma, with a median mutation burden of 319 substitutions per genome (range, 123–700) and 28 indels per genome (range, 14–50; Supplementary Data 1–3). Similarly, copy number analyses demonstrated diploid tumours with few aberrations (Supplementary Fig. 1 and Supplementary Data 4). The previously reported losses in chromosome 22 were not seen in our cases². Only a small number of mutations affected the coding sequence of genes, none of which were plausible driver events.

Recurrent *FOS* and *FOSB* rearrangements. Against this backdrop of a quiet somatic architecture, analysis of structural variants revealed break points in the AP-1 transcription factor *FOS*, in 5/6 cases, and its paralogue *FOSB* in the sixth case (Figs. 1 and 2; and Supplementary Data 5). We analysed and validated these rearrangements at the DNA level by local assembly, copy number analyses and at the RNA level by identification of break point spanning cDNA reads (Supplementary Data 6–8). A single *FOS* or *FOSB* break point was confirmed in each sample, suggesting that these were mono-allelic rearrangements. There was no evidence of similar rearrangements in paired normal tissue samples, confirming that they were somatic. *FOS* rearrangements were also validated with Sanger sequencing of cDNA (Supplementary Fig. 2).

Unusually for structural variants, all *FOS* break points were exonic, residing within a narrow genomic window of exon 4 (Fig. 1a). The rearrangements comprised both interchromosomal and intrachromosomal events. The rearrangement partners were introns of other genes (3/5 cases) or intergenic regions (2/5 cases). There was evidence of expression of the fusion transcript, visible as aberrant spikes in RNA-Seq read coverage adjacent to the break point in the fusion partners. However, these aberrantly

transcribed sequences did not include any known exonic sequence. Indeed, stop codons were encountered at, or immediately after the break points (Fig. 1d and Supplementary Fig. 2).

FOSB rearrangements have been described in two different types of vascular tumours, namely pseudomyogenic haemangiopericytoma and epithelioid haemangioma^{5,6}. The interchromosomal translocation, found in PD7525a, occurred in the same region of exon 1 (Fig. 2). cDNA reads spanning the fusion junction support the expression of a fusion gene, connecting, in frame, *PPP1R10* to *FOSB*. Consequently, the expression of the *FOSB* fusion gene would be brought under the control of the *PPP1R10* promoter (Supplementary Fig. 3).

In contrast to the *FOSB* genomic alteration, the rearrangements of *FOS* do not involve coding sequence of other genes. Transcription remained under the control of its native promoter. Furthermore, in 2/5 cases the fusion partner did not lie within a gene. These observations are supported by re-analyses of RNA sequences of epithelioid haemangioma harbouring *FOS* rearrangements (Fig. 1d)^{7,8}. Similarly to osteoblastoma, the break points in these vascular tumours clustered within the same narrow 200 bp window of exon 4. Furthermore, stop codons were again found in the immediate vicinity of the *FOS* break point.

***FOS* and *FOSB* alterations are ubiquitous in osteoblastoma.** To validate our findings, we examined by fluorescence in situ hybridisation (FISH) an extension cohort of 55 formalin-fixed paraffin-embedded (FFPE) histologically typical cases of osteoblastoma and osteoid osteoma (Supplementary Data 1). In these 55 samples, we found *FOSB* and *FOS* breakapart signals in 1 and 48 tumours, respectively (89% in total; Supplementary Data 1).

We speculated that the six FISH-negative cases may also harbour *FOS* or *FOSB* rearrangements that were not detected because FISH analysis is hampered in tumours of low cellularity, a frequent feature of osteoblastoma¹. FISH may also miss cases with short distance intrachromosomal rearrangements, such as tandem duplications, that insufficiently separate probe target sequences. Since sufficient tissue was available for 3/6 negative cases, we sought alternative evidence for *FOS* dysregulation by immunohistochemistry. All three samples demonstrated strong nuclear *FOS* immunoreactivity, supporting the notion that alterations in *FOS* or *FOSB* underpin every case of osteoblastoma and osteoid osteoma (Supplementary Fig. 4b). *FOSB* immunohistochemistry was uninformative in osteoblastoma, consistent with previous experience with decalcified tumours (Supplementary Fig. 4c)⁹.

***FOS* and *FOSB* alterations are specific to benign bone tumours.**

To explore the utility of our findings as diagnostic markers of osteoblastoma, we assessed their specificity across different tumour sets. We examined *FOS* immunoreactivity in 183 cases of osteosarcoma, including 97 cases of osteoblastic osteosarcoma, and 17 cases of angiosarcoma. In keeping with previous reports, *FOS* immunoreactivity was seen in osteosarcoma samples but only one had a distribution and intensity of immunoreactivity comparable with osteoblastoma¹⁰. While there were no breakapart signals in *FOS* or *FOSB* on FISH testing, copy number gains were noted (Supplementary Fig. 4d). We then examined 55 whole-genome sequences of two published osteosarcoma series, none of which harboured break points in *FOS* or *FOSB*^{11,12}. Finally, we could not find similar *FOS* and *FOSB* rearrangements in whole-genome sequences in 2652 non-osteoblastoma tumours¹³. Taken together, our findings indicate that *FOS* and *FOSB* alterations may be exploited as diagnostic markers for osteoblastoma and osteoid osteoma. We also demonstrate for the first time that both tumour types are similar at a molecular level.

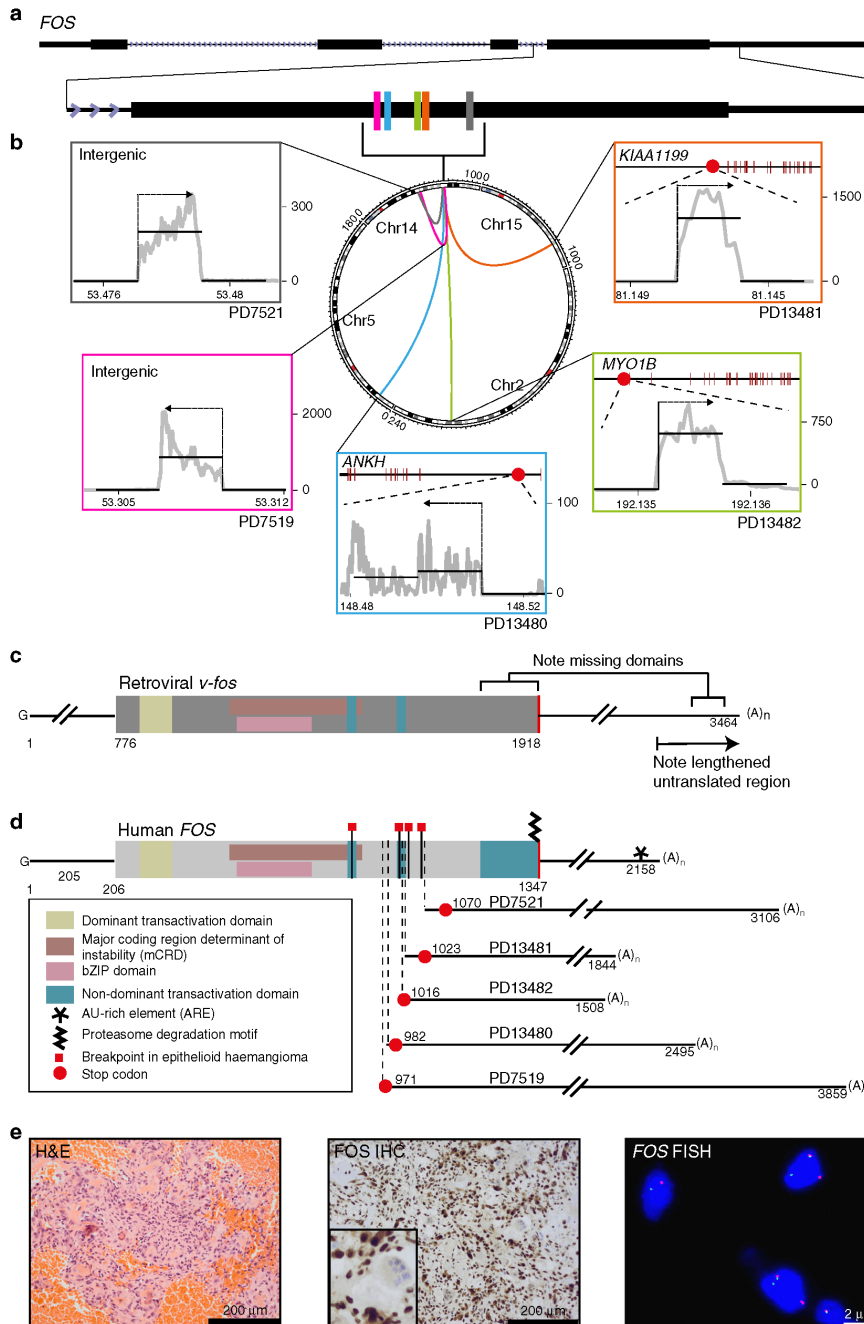


Fig. 1 *FOS* fusions in osteblastoma. **a** Clustered break points in *FOS*. **b** Central Circos plot showing the clustering of break points in *FOS*-mutant samples. All structural variants involving chromosome 14 are shown, demonstrating the paucity of genomic rearrangements. Surrounding panels demonstrate normalised RNA-Seq read counts for each fusion partner. Horizontal line segments reflect mean sequencing counts. The arrow above shows the direction of transcription of the fusion. **c** Retroviral *v-fos*. **d** Schematic of *FOS* break points in benign bone and vascular tumours generating similarity with the murine retroviral transforming *v-fos*. Subscript numbers from left to right report the length of the transcript to the stop codon and the predicted cleavage and polyadenylation site, respectively. **e** *FOS* immunohistochemistry demonstrating strong nuclear immunoreactivity in *FOS*-mutant osteblastoma, PD13482 (centre with zoom inset), Haematoxylin and eosin (H&E) (left), and a clear breakapart of FISH probes surrounding the *FOS* locus (right)

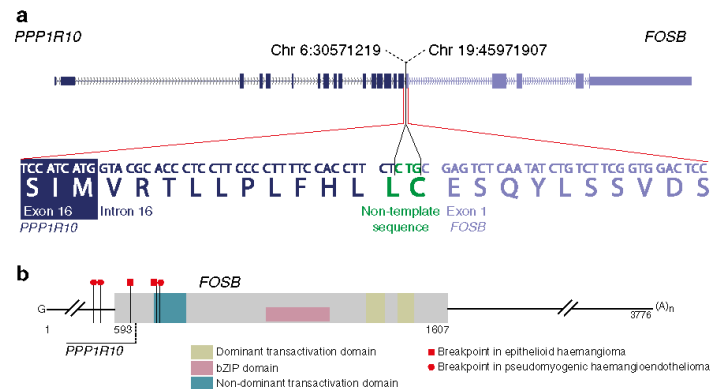


Fig. 2 *FOSB* fusion in PD7525. **a** Schematic of the *PPP1R10*-*FOSB* fusion. The *PPP1R10* exon 16 splice donor site is skipped, resulting in exonisation of part of the intron. A three-nucleotide non-template sequence at the break point then allows an in-frame *PPP1R10*-*FOSB* fusion transcript to be produced. **b** Annotated schematic of the wild-type *FOSB* transcript with the published break points in the vascular tumours (pseudomyogenic haemangioendothelioma (PHE), epithelioid haemangioma (EH))

Discussion

Rearrangements of *FOS* moulded a mutant transcript that lacks regulatory elements. This configuration bears a striking resemblance to the retroviral oncogene, *v-fos*, identified in the FBJ murine osteosarcoma virus (Fig. 1c, d). Dysregulated expression of the murine orthologue, *c-fos*, can cause osteosarcoma in model systems but requires fusion with a highly active promoter and the *v-fos* 3' untranslated region¹⁴.

FOS levels are tightly regulated by both transcript and protein degradation. Two translation-dependent mechanisms ensure rapid mRNA degradation: a length-dependent interaction between the poly-A tail and an exon 3 domain (known as the major coding region determinant of instability)¹⁵, and an independent AU-rich element in the 3' untranslated region¹⁶. Both mechanisms are likely to be disrupted by the rearrangements we have found. Furthermore, ubiquitin-independent proteasomal degradation rapidly depletes the wild-type *FOS* protein¹⁷. The C-terminal truncations seen in epithelioid haemangioma have recently been shown to protect *FOS* from degradation¹⁸. While the break points disrupt components of the C-terminal transactivation domain, this is not required for in vitro transformation by *v-fos*^{19,20}. While we cannot also exclude alteration of AP-1 activity we would expect increased *FOS* concentration in osteoblastoma cells. Consistent with this prediction, we observed intense nuclear immunoreactivity of *FOS* in osteoblastoma cells (Fig. 1f and Supplementary Data 1). Our findings may explain the absence of nonsense mutations in *FOS*, as only rearrangements could abolish both levels of regulation.

Fifty years after the identification of *v-fos* we report human bone-forming tumours, osteoblastoma and osteoid osteoma, that are predominantly characterised by an aberrant *FOS* homologue resembling the viral *fos* oncogene. This shifts our understanding of *FOS*/AP-1 dysregulation in human bone tumours. Our findings also draw an intriguing parallel between bone-forming tumours and a subset of vascular tumours, suggesting possible shared developmental pathways. Patients are likely to benefit from our findings, as they can be readily translated into routine diagnostic practice.

Methods

Patient samples. Patients provided their written and informed consent to provide samples for this study, which was approved by the National Research Ethics Service (NRES) Committee Yorkshire and The Humber – Leeds East (15/YH/0311).

Sequencing. Tumour DNA and RNA were derived from fresh-frozen tissue reviewed by bone pathologists (A.M.F./R.T./F.A.). Matched normal DNA was acquired from blood samples. Whole-genome sequencing was performed using the Illumina HiSeq 2000 or 2500 platform, using 100 bp paired-end sequencing. For whole-genome sequencing, we followed the Illumina no-PCR library protocol to construct short insert 500 bp libraries, prepare flowcells and generate clusters. The average coverage of tumours was at least 40× and of normal DNA at least 30× after alignment with BWA-Mem (2.0.54)²¹ (Supplementary Data 9). Poly-A RNA was sequenced on an Illumina HiSeq 2000 (75 bp paired-end). Sequenced RNA libraries were aligned with STAR (2.0.42)²².

Variant detection. The Cancer Genome Project (Wellcome Trust Sanger Institute) variant calling pipeline was used to call somatic mutations. The following algorithms, with standard settings, and no additional post-processing was used on aligned DNA BAM files: CaVEMan (1.11.0)²³ for substitutions; Pindel (2.1.0)²⁴ for indels; BRASS (5.3.3 <https://github.com/cancerit/BRASS>) for rearrangements, and ASCAT NGS (4.0.0)²⁵ for copy number aberrations. Aligned RNA BAM files, including those realigned from published data, were run through the RNA-Seq analysis pipeline (<https://github.com/cancerit/cgpRna/wiki>), which includes HTSeq (0.6.1)²⁶ for gene feature counts, and the combination of STAR (2.5.0c), TopHat2 (2.1.0)²⁷ and deFuse (0.7.0)²⁸ fusion discovery protocols.

Variant validation. The precision of Cancer Genome Project (Wellcome Trust Sanger Institute) variant calling pipeline has been determined in multiple studies²⁹. We confirmed this through manual inspection of raw sequencing reads. As for rearrangements, we only included break points in this data set that had been validated by reconstruction at base pair resolution.

Analysis of mutations in cancer genes. We analysed variants using a previously documented strategy¹². In brief, we considered variants as potential drivers if they presented in established cancer genes (COSMIC v82). Tumour suppressor coding variants were considered if they were annotated as functionally deleterious by the in-house algorithm, VAGrENT (<http://cancerit.github.io/VAGrENT/>). Disruptive rearrangement break points in or homozygous deletions of tumour suppressors were also considered. Additionally, homozygous deletions were required to be focal (<1 Mb in size). Mutations in oncogenes were considered driver events if they were located at previously reported hot spots (point mutations) or amplified the intact gene. Amplifications also had to be focal (<1 Mb) result in at least five copies in diploid genomes.

Fusion detection. Rearrangements in *FOS* and *FOSB* were analysed using the DNA structural rearrangement caller, BRASS and the in-house RNA fusion detection algorithm, infuse. Fusions were considered if break points and orientations were supported by both algorithms. All reads supporting the break points were manually inspected. In sample PD13482, in which neither algorithm identified the fusion, both split reads and discordant read pairs spanning the fusion were identified in the DNA- and RNA-Seq data.

All *FOS* fusion partner break points were located in genomic regions not normally represented in RNA sequencing libraries as they were intergenic or intronic segments. The per-base coverage in these regions therefore reveals a clear peak, present only in that tumour sample, demonstrating expression of aberrant

transcripts (normalised by the mean of HTSeq counts $\times 10^3$). The end of the transcript was considered to be 10–30 bp downstream of the cleavage and polyadenylation signal ('AATAAA') with the greatest drop in coverage in the surrounding 200 bp. For schematic purposes, mean normalised coverage was plotted as a segment, as Fig. 1: the 'mate transcript segment' is between the break point (grey vertical dashed line) and the poly-adenylation cleavage site; surrounding segments are the mean sequencing coverage over a genomic range of equal length to the 'mate transcript segment'.

FOS fusion validation. cDNA was synthesised from 1 μ g of total RNA was using the ProtoScript[®] II First-Strand cDNA Synthesis Kit (NEB). PCR was performed with Phusion high-fidelity PCR master mix (HF buffer, NEB) using the primers listed in Supplementary Data 10. Amplified products were size selected using gel electrophoresis and then Sanger sequenced using an internal primer listed in Supplementary Data 10.

Allele-specific expression analysis. We analysed allele-specific expression in *FOS* and *FOSB* using allele counts at heterozygous single-nucleotide polymorphisms (SNPs). To allow for poor alignment in RNA-Seq data close to break points, allele counts at heterozygous SNPs were computed manually. Heterozygous SNPs were identified from DNA sequencing data. Allele counts were measured from RNA-Seq reads using GATK ASEReadCounter³⁰.

Fluorescence in situ hybridisation (FISH) for *FOS* and *FOSB*. A cohort of 55 informative cases of osteoblastoma/osteoid osteoma was examined by FISH for *FOS* breakpoint. *FOSB* probes were custom designed with Agilent SureDesign to flank the breakpoint region. *FOS* probes and methods have been described previously³ (Supplementary Data 11). In brief, deparaffinised sections were pretreated by pressure cooking for 5 min and subsequently incubated in pepsin solution at 37 °C for 50 min. Probes were applied to tissue sections and denatured at 72 °C, followed by hybridisation overnight at 37 °C. After hybridisation, the sections were washed and mounted with 4',6-diamidino-2-phenylindole and coverslips.

Immunohistochemistry for *FOS* and *FOSB*. Deparaffinised hydrated tissue sections underwent antigen unmasking in Tris-EDTA pH 9 (DAKO S2367 - Agilent Technologies LDA UK Limited, Cheshire, UK) at high pressure for 2 min. After washing and quenching, sections were blocked in 2.5% horse serum (Vector ImmPRESS Kit) for 20 min at room temperature. Incubation with primary antibodies was for 60 min, secondary antibodies for 30 min, and DAB + substrate/ chromagen (Dako, K3468) for 5 min, all at room temperature, prior to counterstaining and mounting. *FOS* antibodies were EMD Millipore ABE457 (Rabbit Polyclonal, used at 1 or 0.5 μ g mL⁻¹) and ImmPRESS Horse Radish Peroxidase Anti-Rabbit IgG (Peroxidase) Polymer Detection Kit, made in Horse (MP-7401, Vector Laboratories, Peterborough, UK) while *FOSB* antibodies (clone 5G4, dilution 1:100, Cell Signaling Technology, Danvers, MA and rabbit polyclonal CAMTA1 antibody Atlas Antibodies, Stockholm, Sweden) were as previously described³¹.

Data availability. The authors declare that all data supporting the findings of this study are available within the article and its supplementary files or from the corresponding author on reasonable request. Sequencing data have been deposited at the European Genome-Phenome Archive (<http://www.ebi.ac.uk/ega/>) that is hosted by the European Bioinformatics Institute. DNA (<https://www.ebi.ac.uk/ega/datasets/EGAD00001000785>; <https://www.ebi.ac.uk/ega/datasets/EGAD00001000147>) accession numbers: EGAN00001100713, EGAN00001100730, EGAN00001100714, EGAN00001100731, EGAN00001100715, EGAN00001100732, EGAN00001031765, EGAN00001032117, EGAN00001031767, EGAN00001032119, EGAN00001036773, EGAN00001036983. RNA (<https://www.ebi.ac.uk/ega/studies/EGAS00001000763>) accession numbers: EGAN00001196539, EGAN00001196540, EGAN00001209957, EGAN00001196544, EGAN00001196545, EGAN00001209959.

Received: 23 November 2017 Accepted: 8 May 2018

Published online: 01 June 2018

References

- Czerniak, B. & Dorfman, H. D. *Dorfman and Czerniak's Bone Tumors* (Elsevier, Philadelphia, 2016).
- Nord, K. H. et al. Recurrent chromosome 22 deletions in osteoblastoma affect inhibitors of the Wnt/beta-catenin signaling pathway. *PLoS ONE* **8**, e80725 (2013).
- Giannico, G. et al. Osteoblastoma characterized by a three-way translocation: report of a case and review of the literature. *Cancer Genet. Cytogenet.* **195**, 168–171 (2009).
- Zhang, Y. & Rosenberg, A. E. Bone-forming tumors. *Surg. Pathol. Clin.* **10**, 513–535 (2017).
- Walther, C. et al. A novel SERPINE1-FOSB fusion gene results in transcriptional up-regulation of FOSE in pseudomyogenic haemangioid endothelioma. *J. Pathol.* **232**, 534–540 (2014).
- Antonescu, C. R. et al. ZFP36-FOSB fusion defines a subset of epithelioid hemangioma with atypical features. *Genes Chromosomes Cancer* **53**, 951–959 (2014).
- van Ijzendoorn, D. G. et al. Fusion events lead to truncation of FOS in epithelioid hemangioma of bone. *Genes Chromosomes Cancer* **54**, 565–574 (2015).
- Huang, S. C. et al. Frequent FOS gene rearrangements in epithelioid hemangioma: a molecular study of 58 cases with morphologic reappraisal. *Am. J. Surg. Pathol.* **39**, 1313–1321 (2015).
- Hung, Y. P., Fletcher, C. D. & Hornick, J. L. FOSE is a useful diagnostic marker for pseudomyogenic hemangioid endothelioma. *Am. J. Surg. Pathol.* **41**, 596–606 (2017).
- Weekes, D. et al. Regulation of osteosarcoma cell lung metastasis by the c-Fos/AP-1 target PGPRI. *Oncogene* **35**, 2948 (2016).
- Chen, K. S. et al. A novel TP53-KPNA3 translocation defines a de novo treatment-resistant clone in osteosarcoma. *Cold Spring Harb. Mol. Case Stud.* **2**, a000992 (2016).
- Behjati, S. et al. Recurrent mutation of IGF signalling genes and distinct patterns of genomic rearrangement in osteosarcoma. *Nat. Commun.* **8**, 15936 (2017).
- Li, Y. et al. Patterns of structural variation in human cancer. Preprint at *bioRxiv* <https://www.biorxiv.org/content/early/2017/08/27/181339> (2017).
- Grigoriadis, A. E., Schellander, K., Wang, Z. Q. & Wagner, E. F. Osteoblasts are target cells for transformation in c-fos transgenic mice. *J. Cell Biol.* **122**, 685–701 (1993).
- Grosset, C. et al. A mechanism for translationally coupled mRNA turnover: interaction between the poly(A) tail and a c-fos RNA coding determinant via a protein complex. *Cell* **103**, 29–40 (2000).
- Chen, C. Y., Chen, T. M. & Shyu, A. B. Interplay of two functionally and structurally distinct domains of the c-fos AU-rich element specifies its mRNA-destabilizing function. *Mol. Cell Biol.* **14**, 416–426 (1994).
- Ferrara, P. et al. The structural determinants responsible for c-Fos protein proteasomal degradation differ according to the conditions of expression. *Oncogene* **22**, 1461–1474 (2003).
- van Ijzendoorn, D. G. P. et al. Functional analyses of a human vascular tumor FOS variant identify a novel degradation mechanism and a link to tumorigenesis. *J. Biol. Chem.* **292**, 21282–21290 (2017).
- Jooss, K. U., Funk, M. & Müller, R. An autonomous N-terminal transactivation domain in Fos protein plays a crucial role in transformation. *EMBO J.* **13**, 1467–1475 (1994).
- Sutherland, J. A., Cook, A., Bannister, A. J. & Kouzarides, T. Conserved motifs in Fos and Jun define a new class of activation domain. *Genes Dev.* **6**, 1810–1819 (1992).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
- Jones, D. et al. cgpCaVEManWrapper: simple execution of CaVEMan in order to detect somatic single nucleotide variants in NGS data. *Curr. Protoc. Bioinformatics* **56**, 15.10.1–15.10.18 (2016).
- Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865–2871 (2009).
- Van Loo, P. et al. Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci. USA* **107**, 16910–16915 (2010).
- Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
- Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
- McPherson, A. et al. deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. *PLoS Comput. Biol.* **7**, e1001138 (2011).
- Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
- McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
- Sugita, S. et al. Diagnostic utility of FOSE immunohistochemistry in pseudomyogenic hemangioid endothelioma and its histological mimics. *Diagn. Pathol.* **11**, 75 (2016).

Acknowledgements

This work was funded by The Wellcome Trust; Skeletal Cancer Action Trust UK; the Royal National Orthopaedic Hospital NHS Trust; Rosetrees Trust. A.M.F. is an NIHR

senior investigator and was also supported by the National Institute for Health Research, UCLH Biomedical Research Centre and the UCL Experimental Cancer Centre. M.W.F., A.V., J.D., M.T. and P.V.L. are supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001202), the UK Medical Research Council (FC001202) and the Wellcome Trust (FC001202). Personal fellowships have been granted to S.B. (Wellcome Trust Intermediate Clinical Research Fellowship; St. Baldrick's Foundation Robert J. Arceci International Innovation Award); P.J.C. (Wellcome Trust Senior Clinical Research Fellowship); N.P. (CRUK Clinician Scientist Fellowship) and E.M. (CRUK Career Development Fellow); C.R.A. is supported by NIH grants (P50 CA140146-01 and P30 CA008748), M.W.F. (Crick-CRUK Doctoral Fellowship); J.D. (Research Foundation – Flanders, FWO Postdoctoral Fellowship; European Union's Horizon 2020 Research and Innovation programme, MSCA 703594-DECODE); M.T. (European Union's Horizon 2020 Research and Innovation programme postdoctoral fellowship, MSCA 747852-SIOMICS). P.V.L. is a Winton Group Leader in recognition of the Winton Charitable Foundation's support towards the establishment of The Francis Crick Institute. We are grateful to Donna Magsumbol for technical work, the RNOH Musculoskeletal Pathology Biobank biobank team for consenting patients and accessing samples and the UCL CRUK Cancer Centre Pathology Core Facility for immunohistochemistry. We thank patients for participating in our research and the clinical teams involved in their care.

Author contributions

M.W.F. and W.M. performed data analyses. M.D.Y. and P.S.T. contributed to data analysis. C.R.A., L.Z., A.M.F., R.T., F.A. and W.M. contributed to design of FISH experiments and supplied BAC clones. H.Y., F.B. and A.-C.S. performed FISH analyses. E.M. and F.B. performed immunostaining. A.V. performed the fusion validation with PCR and Sanger sequencing. J.D. contributed to allele-specific expression analysis. A.M.F., W.M., R.T. and F.A. curated and reviewed the samples, clinical data and/or provided clinical expertise. M.T., N.P., J.D. and A.E.G. contributed to discussions. M.W.F., S.B. and A.V. contributed to figure design. A.M.F., P.V.L., S.B., M.R.S. and P.J.C. directed the research. M.W.F., S.B. and A.M.F. wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-04530-z>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

Pan-cancer analysis of whole genomes

<https://doi.org/10.1038/s41586-020-1969-6>

The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium

Received: 29 July 2018

Accepted: 11 December 2019

Published online: 5 February 2020

Open access

Cancer is driven by genetic change, and the advent of massively parallel sequencing has enabled systematic documentation of this variation at the whole-genome scale^{1–3}. Here we report the integrative analysis of 2,658 whole-cancer genomes and their matching normal tissues across 38 tumour types from the Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium of the International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA). We describe the generation of the PCAWG resource, facilitated by international data sharing using compute clouds. On average, cancer genomes contained 4–5 driver mutations when combining coding and non-coding genomic elements; however, in around 5% of cases no drivers were identified, suggesting that cancer driver discovery is not yet complete. Chromothripsis, in which many clustered structural variants arise in a single catastrophic event, is frequently an early event in tumour evolution; in acral melanoma, for example, these events precede most somatic point mutations and affect several cancer-associated genes simultaneously. Cancers with abnormal telomere maintenance often originate from tissues with low replicative activity and show several mechanisms of preventing telomere attrition to critical levels. Common and rare germline variants affect patterns of somatic mutation, including point mutations, structural variants and somatic retrotransposition. A collection of papers from the PCAWG Consortium describes non-coding mutations that drive cancer beyond those in the *TERT* promoter⁴; identifies new signatures of mutational processes that cause base substitutions, small insertions and deletions and structural variation^{5,6}; analyses timings and patterns of tumour evolution⁷; describes the diverse transcriptional consequences of somatic mutation on splicing, expression levels, fusion genes and promoter activity^{8,9}; and evaluates a range of more-specialized features of cancer genomes^{8,10–18}.

Cancer is the second most-frequent cause of death worldwide, killing more than 8 million people every year; the incidence of cancer is expected to increase by more than 50% over the coming decades^{19,20}. ‘Cancer’ is a catch-all term used to denote a set of diseases characterized by autonomous expansion and spread of a somatic clone. To achieve this behaviour, the cancer clone must co-opt multiple cellular pathways that enable it to disregard the normal constraints on cell growth, modify the local microenvironment to favour its own proliferation, invade through tissue barriers, spread to other organs and evade immune surveillance²¹. No single cellular program directs these behaviours. Rather, there is a large pool of potential pathogenic abnormalities from which individual cancers draw their own combinations: the commonalities of macroscopic features across tumours belie a vastly heterogeneous landscape of cellular abnormalities.

This heterogeneity arises from the stochastic nature of Darwinian evolution. There are three preconditions for Darwinian evolution: characteristics must vary within a population; this variation must be heritable from parent to offspring; and there must be competition for survival within the population. In the context of somatic cells, heritable variation arises from mutations acquired stochastically throughout life, notwithstanding additional contributions from germline and epigenetic variation. A subset of these mutations alter the cellular phenotype, and a small subset of those variants confer an advantage

on clones during the competition to escape the tight physiological controls wired into somatic cells. Mutations that provide a selective advantage to the clone are termed driver mutations, as opposed to selectively neutral passenger mutations.

Initial studies using massively parallel sequencing demonstrated the feasibility of identifying every somatic point mutation, copy-number change and structural variant (SV) in a given cancer^{1–3}. In 2008, recognizing the opportunity that this advance in technology provided, the global cancer genomics community established the ICGC with the goal of systematically documenting the somatic mutations that drive common tumour types²².

The pan-cancer analysis of whole genomes

The expansion of whole-genome sequencing studies from individual ICGC and TCGA working groups presented the opportunity to undertake a meta-analysis of genomic features across tumour types. To achieve this, the PCAWG Consortium was established. A Technical Working Group implemented the informatics analyses by aggregating the raw sequencing data from different working groups that studied individual tumour types, aligning the sequences to the human genome and delivering a set of high-quality somatic mutation calls for downstream analysis (Extended Data Fig. 1). Given the recent meta-analysis

A list of members and their affiliations appears in the online version of the paper and lists of working groups appear in the Supplementary Information.

Box 1

Online resources for data access, visualization and analysis

The PCAWG landing page (<http://docs.icgc.org/pcawg>) provides links to several data resources for interactive online browsing, analysis and download of PCAWG data and results (Supplementary Table 4).

Direct download of PCAWG data

Aligned PCAWG read data in BAM format are also available at the European Genome Phenome Archive (EGA; <https://www.ebi.ac.uk/ega/search/site/pcawg> under accession number EGAS00001001692). In addition, all open-tier PCAWG genomics data, as well as reference datasets used for analysis, can be downloaded from the ICGC Data Portal at <http://docs.icgc.org/pcawg/data/>. Controlled-tier genomic data, including SNVs and indels that originated from TCGA projects (in VCF format) and aligned reads (in BAM format) can be downloaded using the Score (<https://www.overture.bio/>) software package, which has accelerated and secure file transfer, as well as BAM slicing facilities to selectively download defined regions of genomic alignments.

PCAWG computational pipelines

The core alignment, somatic variant-calling, quality-control and variant consensus-generation pipelines used by PCAWG have each been packaged into portable cross-platform images using the Dockstore system⁸⁴ and released under an Open Source licence that enables unrestricted use and redistribution. All PCAWG Dockstore images are available to the public at <https://dockstore.org/organizations/PCAWG/collections/PCAWG>.

ICGC Data Portal

The ICGC Data Portal⁸⁵ (<https://dcc.icgc.org>) serves as the main entry point for accessing PCAWG datasets with a single uniform web interface and a high-performance data-download client. This uniform interface provides users with easy access to the myriad of PCAWG sequencing data and variant calls that reside in many repositories and compute clouds worldwide. Streaming technology⁸⁶ provides users with high-level visualizations in real time of BAM and VCF files stored remotely on the Cancer Genome Collaboratory.

UCSC Xena

UCSC Xena⁸⁷ (<https://pcawg.xenahubs.net>) visualizes all PCAWG primary results, including copy-number, gene-expression, gene-fusion and promoter-usage alterations, simple somatic mutations, large somatic structural variations, mutational signatures and phenotypic data. These open-access data are available through a public Xena hub, and consensus simple somatic mutations can be loaded to the local computer of a user via a private Xena hub. Kaplan–Meier plots, histograms, box plots, scatter plots and transcript-specific views offer additional visualization options and statistical analyses.

The Expression Atlas

The Expression Atlas (<https://www.ebi.ac.uk/gxa/home>) contains RNA-sequencing and expression microarray data for querying gene expression across tissues, cell types, developmental stages and/or experimental conditions⁸⁸. Two different views of the data are provided: summarized expression levels for each tumour type and gene expression at the level of individual samples, including reference-gene expression datasets for matching normal tissues.

PCAWG Scout

PCAWG Scout (<http://pcawgscout.bsc.es/>) provides a framework for -omics workflow and website templating to generate on-demand, in-depth analyses of the PCAWG data that are openly available to the whole research community. Views of protected data are available that still safeguard sensitive data. Through the PCAWG Scout web interface, users can access an array of reports and visualizations that leverage on-demand bioinformatic computing infrastructure to produce results in real time, allowing users to discover trends as well as form and test hypotheses.

Chromothrips Explorer

Chromothrips Explorer (<http://compbio.med.harvard.edu/chromothripsis/>) is a portal that allows structural variation in the PCAWG dataset to be explored on an individual patient basis through the use of circo plots. Patterns of chromothripsis can also be explored in aggregated formats.

of exome data from the TCGA Pan-Cancer Atlas^{23–25}, scientific working groups concentrated their efforts on analyses best-informed by whole-genome sequencing data.

We collected genome data from 2,834 donors (Extended Data Table 1), of which 176 were excluded after quality assurance. A further 75 had minor issues that could affect some of the analyses (grey-listed donors) and 2,583 had data of optimal quality (white-listed donors) (Supplementary Table 1). Across the 2,658 white- and grey-listed donors, whole genome sequencing data were available from 2,605 primary tumours and 173 metastases or local recurrences. Mean read coverage was 39× for normal samples, whereas tumours had a bimodal coverage distribution with modes at 38× and 60× (Supplementary Fig. 1). RNA-sequencing data were available for 1,222 donors. The final cohort comprised 1,469 men (55%) and 1,189 women (45%), with a mean age of 56 years (range, 1–90 years) across 38 tumour types (Extended Data Table 1 and Supplementary Table 1).

To identify somatic mutations, we analysed all 6,835 samples using a uniform set of algorithms for alignment, variant calling and quality control (Extended Data Fig. 1, Supplementary Fig. 2 and Supplementary Methods 2). We used three established pipelines to call somatic single-nucleotide variations (SNVs), small insertions and deletions (indels), copy-number alterations (CNAs) and SVs. Somatic retrotransposition events, mitochondrial DNA mutations and telomere lengths were also called by bespoke algorithms. RNA-sequencing data were uniformly

processed to call transcriptomic alterations. Germline variants identified by the three separate pipelines included single-nucleotide polymorphisms, indels, SVs and mobile-element insertions (Supplementary Table 2).

The requirement to uniformly realign and call variants on approximately 5,800 whole genomes presented considerable computational challenges, and raised ethical issues owing to the use of data from different jurisdictions (Extended Data Table 2). We used cloud computing^{26,27} to distribute alignment and variant calling across 13 data centres on 3 continents (Supplementary Table 3). Core pipelines were packaged into Docker containers²⁸ as reproducible, stand-alone packages, which we have made available for download. Data repositories for raw and derived datasets, together with portals for data visualization and exploration, have also been created (Box 1 and Supplementary Table 4).

Benchmarking of genetic variant calls

To benchmark mutation calling, we ran the 3 core pipelines, together with 10 additional pipelines, on 63 representative tumour–normal genome pairs (Supplementary Note 1). For 50 of these cases, we performed validation by hybridization of tumour and matched normal DNA to a custom bait set with deep sequencing²⁹. The 3 core somatic variant-calling pipelines had individual estimates of sensitivity of 80–90% to detect a true somatic SNV called by any of the 13 pipelines; more

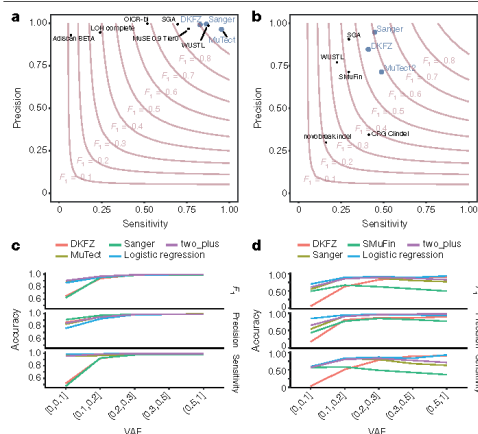


Fig. 1 | Validation of variant-calling pipelines in PCAWG. **a**, Scatter plot of estimated sensitivity and precision for somatic SNVs across individual algorithms assessed in the validation exercise across $n = 63$ PCAWG samples. Core algorithms included in the final PCAWG call set are shown in blue. **b**, Sensitivity and precision estimates across individual algorithms for somatic indels. **c**, Accuracy (precision, sensitivity and F_1 score, defined as $2 \times \text{sensitivity} \times \text{precision} / (\text{sensitivity} + \text{precision})$) of somatic SNV calls across variant allele fractions (VAFs) for the core algorithms. The accuracy of two methods of combining variant calls (two-plus, which was used in the final dataset, and logistic regression) is also shown. **d**, Accuracy of indel calls across variant allele fractions.

than 95% of SNV calls made by each of the core pipelines were genuine somatic variants (Fig. 1a). For indels—a more-challenging class of variants to identify with short-read sequencing—the 3 core algorithms had individual sensitivity estimates in the range of 40–50%, with precision of 70–95% (Fig. 1b). For individual SV algorithms, we estimated precision to be in the range 80–95% for samples in the 63-sample pilot dataset.

Next, we defined a strategy to merge results from the three pipelines into one final call set to be used for downstream scientific analyses (Methods and Supplementary Note 2). Sensitivity and precision of consensus somatic variant calls were 95% (90% confidence interval, 88–98%) and 95% (90% confidence interval, 71–99%), respectively, for SNVs (Extended Data Fig. 2). For somatic indels, sensitivity and precision were 60% (34–72%) and 91% (73–96%), respectively (Extended Data Fig. 2). Regarding somatic SVs, we estimate the sensitivity of merged calls to be 9.0% for true calls generated by any one pipeline; precision was estimated as 97.5%. The improvement in calling accuracy from combining different pipelines was most noticeable in variants with low variant allele fractions, which probably originate from tumour subclones (Fig. 1c, d). Germline variant calls, phased using a haplotype-reference panel, displayed a precision of more than 99% and a sensitivity of 92–98% (Supplementary Note 2).

Analysis of PCAWG data

The uniformly generated, high-quality set of variant calls across more than 2,500 donors provided the springboard for a series of scientific working groups to explore the biology of cancer. A comprehensive suite of companion papers that describe the analyses and discoveries across these thematic areas is copublished with this paper^{4–18} (Extended Data Table 3).

Pan-cancer burden of somatic mutations

Across the 2,583 white-listed PCAWG donors, we called 43,778,859 somatic SNVs, 410,123 somatic multinucleotide variants, 2,418,247 somatic indels, 288,416 somatic SVs, 19,166 somatic retrotransposition events and 8,185 de novo mitochondrial DNA mutations (Supplementary Table 1). There was considerable heterogeneity in the burden of somatic mutations across patients and tumour types, with a broad correlation in mutation burden among different classes of somatic variation (Extended Data Fig. 3). Analysed at a per-patient level, this correlation held, even when considering tumours with similar purity and ploidy (Supplementary Fig. 3). Why such correlation should apply on a pan-cancer basis is unclear. It is likely that age has some role, as we observe a correlation between most classes of somatic mutation and age at diagnosis (around 19.0 SNVs per year, $P = 0.02$; about 22 indels per year, $P = 5 \times 10^{-3}$; 1.5 SVs per year, $P < 2 \times 10^{-16}$; linear regression with likelihood ratio tests; Supplementary Fig. 4). Other factors are also likely to contribute to the correlations among classes of somatic mutation, as there is evidence that some DNA repair defects can cause multiple types of somatic mutation³⁰, and a single carcinogen can cause a range of DNA lesions³¹.

Panorama of driver mutations in cancer

We extracted the subset of somatic mutations in PCAWG tumours that have high confidence to be driver events on the basis of current knowledge. One challenge to pinpointing the specific driver mutations in an individual tumour is that not all point mutations in recurrently mutated cancer-associated genes are drivers³². For genomic elements significantly mutated in PCAWG data, we developed a ‘rank-and-cut’ approach to identify the probable drivers (Supplementary Methods 8.1). This approach works by ranking the observed mutations in a given genomic element based on recurrence, estimated functional consequence and expected pattern of drivers in that element. We then estimate the excess burden of somatic mutations in that genomic element above that expected for the background mutation rate, and cut the ranked mutations at this level. Mutations in each element with the highest driver ranking were then assigned as probable drivers; those below the threshold will probably have arisen through chance and were assigned as probable passengers. Improvements to features that are used to rank the mutations and the methods used to measure them will contribute to further development of the rank-and-cut approach.

We also needed to account for the fact that some bona fide cancer genomic elements were not rediscovered in PCAWG data because of low statistical power. We therefore added previously known cancer-associated genes to the discovery set, creating a ‘compendium of mutational driver elements’ (Supplementary Methods 8.2). Then, using stringent rules to nominate driver point mutations that affect these genomic elements on the basis of prior knowledge³³, we separated probable driver from passenger point mutations. To cover all classes of variant, we also created a compendium of known driver SVs, using analogous rules to identify which somatic CNAs and SVs are most likely to act as drivers in each tumour. For probable pathogenic germline variants, we identified all truncating germline point mutations and SVs that affect high-penetrance germline cancer-associated genes.

This analysis defined a set of mutations that we could confidently assert, based on current knowledge, drove tumorigenesis in the more than 2,500 tumours of PCAWG. We found that 91% of tumours had at least one identified driver mutation, with an average of 4.6 drivers per tumour identified, showing extensive variation across cancer types (Fig. 2a). For coding point mutations, the average was 2.6 drivers per tumour, similar to numbers estimated in known cancer-associated genes in tumours in the TCGA using analogous approaches³².

To address the frequency of non-coding driver point mutations, we combined promoters and enhancers that are known targets of

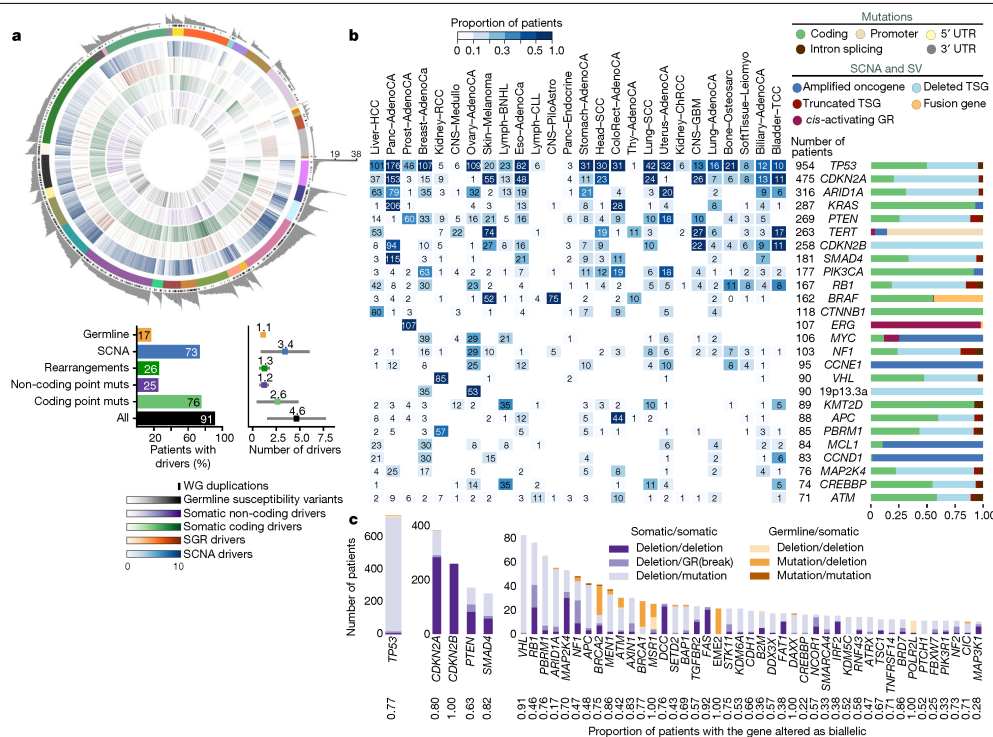


Fig. 2 | Panorama of driver mutations in PCAWG. **a**, Top, putative driver mutations in PCAWG, represented as a circos plot. Each sector represents a tumour in the cohort. From the periphery to the centre of the plot the concentric rings represent: (1) the total number of driver alterations; (2) the presence of whole-genome (WG) duplication; (3) the tumour type; (4) the number of driver CNAs; (5) the number of driver genomic rearrangements; (6) driver coding point mutations; (7) driver non-coding point mutations; and (8) pathogenic germline variants. Bottom, snapshots of the panorama of driver mutations. The horizontal bar plot (left) represents the proportion of patients with different types of drivers. The dot plot (right) represents the mean number of each type of driver mutation across tumours with at least one event (the square dot) and the standard deviation (grey whiskers), based on $n = 2,583$

patients. **b**, Genomic elements targeted by different types of mutations in the cohort altered in more than 65 tumours. Both germline and somatic variants are included. Left, the heatmap shows the recurrence of alterations across cancer types. The colour indicates the proportion of mutated tumours and the number indicates the absolute count of mutated tumours. Right, the proportion of each type of alteration that affects each genomic element. **c**, Tumour-suppressor genes with biallelic inactivation in 10 or more patients. The values included under the gene labels represent the proportions of patients who have biallelic mutations in the gene out of all patients with a somatic mutation in that gene. GR, genomic rearrangement; SCNA, somatic copy-number alteration; SGR, somatic genome rearrangement; TSG, tumour suppressor gene; UTR, untranslated region.

non-coding drivers^{34–37} with those newly discovered in PCAWG data; this is reported in a companion paper⁴. Using this approach, only 13% (785 out of 5,913) of driver point mutations were non-coding in PCAWG. Nonetheless, 25% of PCAWG tumours bear at least one putative non-coding driver point mutation, and one third (237 out of 785) affected the *TERT* promoter (9% of PCAWG tumours). Overall, non-coding driver point mutations are less frequent than coding driver mutations. With the exception of the *TERT* promoter, individual enhancers and promoters are only infrequent targets of driver mutations⁴.

Across tumour types, SVs and point mutations have different relative contributions to tumorigenesis. Driver SVs are more prevalent in breast adenocarcinomas (6.4 ± 3.7 SVs (mean \pm s.d.) compared with 2.2 ± 1.3 point mutations; $P < 1 \times 10^{-16}$, Mann-Whitney *U*-test) and ovary adenocarcinomas (5.8 ± 2.6 SVs compared with 1.9 ± 1.0 point mutations; $P < 1 \times 10^{-16}$), whereas driver point mutations have

a larger contribution in colorectal adenocarcinomas (2.4 ± 1.4 SVs compared with 7.4 ± 7.0 point mutations; $P = 4 \times 10^{-10}$) and mature B cell lymphomas (2.2 ± 1.3 SVs compared with 6 ± 3.8 point mutations; $P < 1 \times 10^{-16}$), as previously shown³⁸. Across tumour types, there are differences in which classes of mutation affect a given genomic element (Fig. 2b).

We confirmed that many driver mutations that affect tumour-suppressor genes are two-hit inactivation events (Fig. 2c). For example, of the 954 tumours in the cohort with driver mutations in *TP53*, 736 (77%) had both alleles mutated, 96% of which (707 out of 736) combined a somatic point mutation that affected one allele with somatic deletion of the other allele. Overall, 17% of patients had rare germline protein-truncating variants (PTVs) in cancer-predisposition genes³⁹, DNA-damage response genes⁴⁰ and somatic driver genes. Biallelic inactivation due to somatic alteration on top of a germline PTV was observed in 4.5% of patients overall, with 81% of

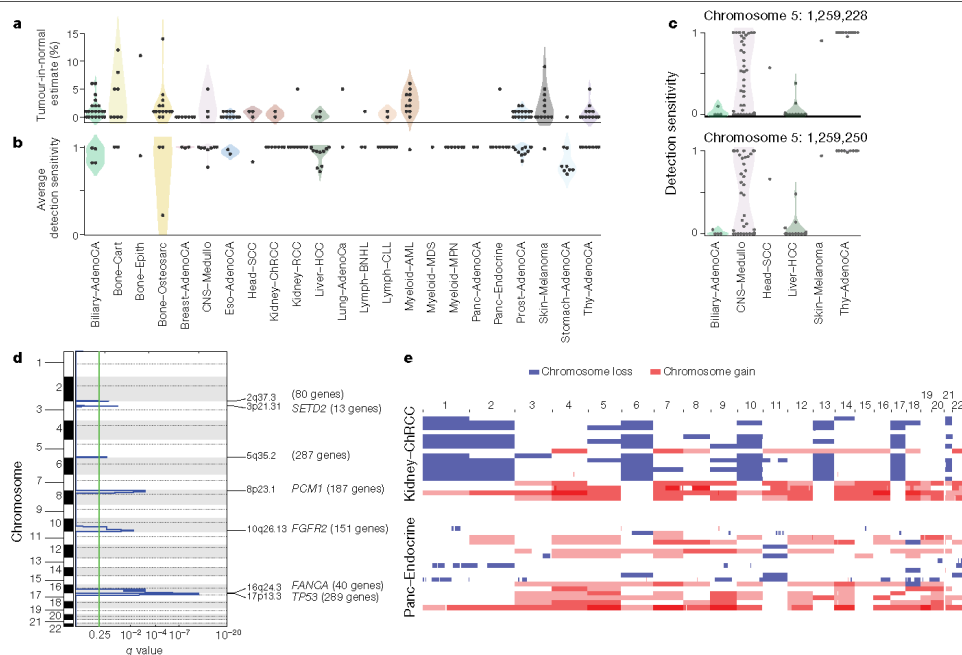


Fig. 3 | Analysis of patients with no detected driver mutations. **a**, Individual estimates of the percentage of tumour-in-normal contamination across patients with no driver mutations in PCAWG ($n = 181$). No data were available for myelodysplastic syndromes and acute myeloid leukaemia. Points represent estimates for individual patients, and the coloured areas are estimated density distributions (violin plots). Abbreviations of the tumour types are defined in Extended Data Table 1. **b**, Average detection sensitivity by tumour type for tumours without known drivers ($n = 181$). Each dot represents a given sample and is the average sensitivity of detecting clonal substitutions across the genome, taking into account purity and ploidy. Coloured areas are estimated density distributions, shown for cohorts with at least five cases. **c**, Detection

sensitivity for *TERT* promoter hotspots in tumour types in which *TERT* is frequently mutated. Coloured areas are estimated density distributions. **d**, Significant copy-number losses identified by two-sided hypothesis testing using GISTIC2.0, corrected for multiple-hypothesis testing. Numbers in parentheses indicate the number of genes in significant regions when analysing medulloblastomas without known drivers ($n = 42$). Significant regions with known cancer-associated genes are labelled with the representative cancer-associated gene. **e**, Aneuploidy in chromophobe renal cell carcinomas and pancreatic neuroendocrine tumours without known drivers. Patients are ordered on the y-axis by tumour type and then by presence of whole-genome duplication (bottom) or not (top).

these affecting known cancer-predisposition genes (such as *BRCA1*, *BRCA2* and *ATM*).

PCAWG tumours with no apparent drivers

Although more than 90% of PCAWG cases had identified drivers, we found none in 181 tumours (Extended Data Fig. 4a). Reasons for missing drivers have not yet been systematically evaluated in a pan-cancer cohort, and could arise from either technical or biological causes.

Technical explanations could include poor-quality samples, inadequate sequencing or failures in the bioinformatic algorithms used. We assessed the quality of the samples and found that 4 of the 181 cases with no known drivers had more than 5% tumour DNA contamination in their matched normal sample (Fig. 3a). Using an algorithm designed to correct for this contamination⁴¹, we identified previously missed mutations in genes relevant to the respective cancer types. Similarly, if the fraction of tumour cells in the cancer sample is low through stromal contamination, the detection of driver mutations can be impaired. Most tumours with no known drivers had an average power to detect mutations close to 100%; however, a few had power in the 70–90% range (Fig. 3b and Extended Data Fig. 4b). Even

in adequately sequenced genomes, lack of read depth at specific driver loci can impair mutation detection. For example, only around 50% of PCAWG tumours had sufficient coverage to call a mutation ($\geq 90\%$ power) at the two *TERT* promoter hotspots, probably because the high GC content of this region causes biased coverage (Fig. 3c). In fact, 6 hepatocellular carcinomas and 2 biliary cholangiocarcinomas among the 181 cases with no known drivers actually did contain *TERT* mutations, which were discovered after deep targeted sequencing⁴².

Finally, technical reasons for missing driver mutations include failures in the bioinformatic algorithms. This affected 35 myeloproliferative neoplasms in PCAWG, in which the *JAK2*^{V617F} driver mutation should have been called. Our somatic variant-calling algorithms rely on ‘panels of normals’, typically from blood samples, to remove recurrent sequencing artefacts. As 2–5% of healthy individuals carry occult haematopoietic clones⁴³, recurrent driver mutations in these clones can enter panels of normals.

With regard to biological causes, tumours may be driven by mutations in cancer-associated genes that are not yet described for that tumour type. Using driver discovery algorithms on tumours with no known drivers, no individual genes reached significance for point mutations. However, we identified a recurrent CNA that spanned *SETD2* in

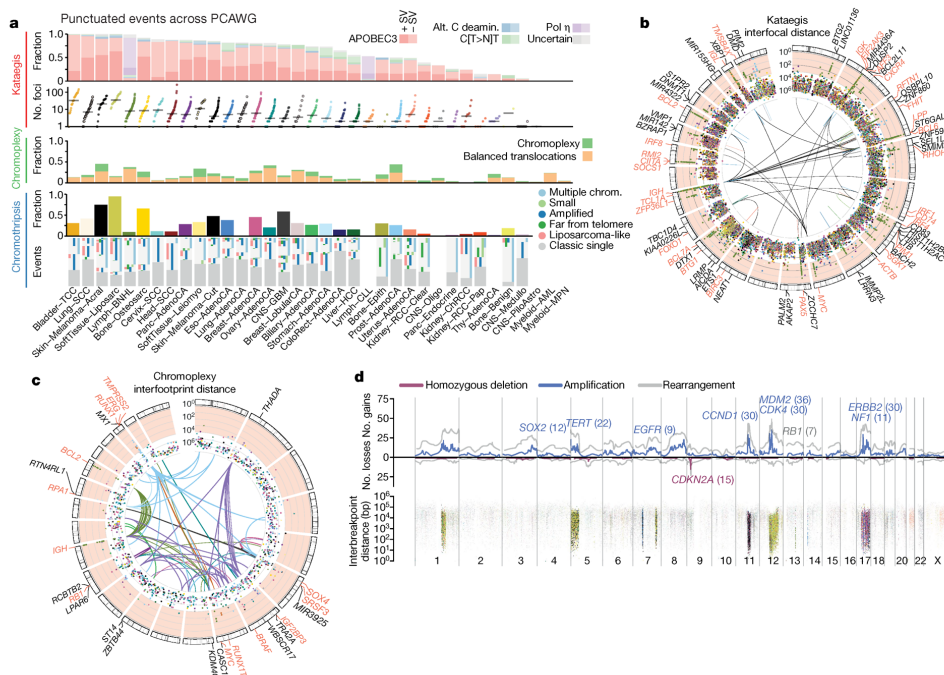


Fig. 4 | Patterns of clustered mutational processes in PCAWG. **a**, Kataegis. Top, prevalence of different types of kataegis and their association with SVs (≤ 1 kb from the focus). Bottom, the distribution of the number of foci of kataegis per sample. Chromoplexy. Prevalence of chromoplexy across cancer types, subdivided into balanced translocations and more complex events. Chromothripsis. Top, frequency of chromothripsis across cancer types. Bottom, for each cancer type a column is shown, in which each row is a chromothripsis region represented by five coloured rectangles relating to its categorization. **b**, Circos rainfall plot showing the distances between consecutive kataegis events across PCAWG compared with their genomic position. Lymphoid tumours (khaki, B cell non-Hodgkin's lymphoma; orange, chronic lymphocytic leukaemia) have hypermutation hot spots (≥ 3 foci with distance ≤ 1 kb; pale red zone), many of which are near known cancer-associated genes (red annotations) and have associated SVs (≤ 10 kb from the focus; shown as arcs in the centre). **c**, Circos rainfall plot as in **b** that shows the distance versus

the position of consecutive chromoplexy and reciprocal translocation footprints across PCAWG. Lymphoid, prostate and thyroid cancers exhibit recurrent events (≥ 2 footprints with distance ≤ 10 kb; pale red zone) that are likely to be driver SVs and are annotated with nearby genes and associated SVs, which are shown as bold and thin arcs for chromoplexy and reciprocal translocations, respectively (colours as in **a**). **d**, Effect of chromothripsis along the genome and involvement of PCAWG driver genes. Top, number of chromothripsis-induced gains or losses (grey) and amplifications (blue) or deletions (red). Within the identified chromothripsis regions, selected recurrently rearranged (light grey), amplified (blue) and homozygously deleted (magenta) driver genes are indicated. Bottom, interbreakpoint distance between all subsequent breakpoints within chromothripsis regions across cancer types, coloured by cancer type. Regions with an average interbreakpoint distance < 10 kb are highlighted. C[T>N]T, kataegis with a pattern of thymine mutations in a Cp TpT context.

medulloblastomas that lacked known drivers (Fig. 3d), indicating that restricting hypothesis testing to missing-driver cases can improve power if undiscovered genes are enriched in such tumours. Inactivation of *SETD2* in medulloblastoma significantly decreased gene expression ($P = 0.002$) (Extended Data Fig. 4c). Notably, *SETD2* mutations occurred exclusively in medulloblastoma group-4 tumours ($P < 1 \times 10^{-4}$). Group-4 medulloblastomas are known for frequent mutations in other chromatin-modifying genes⁴⁴, and our results suggest that *SETD2* loss of function is an additional driver that affects chromatin regulators in this subgroup.

Two tumour types had a surprisingly high fraction of patients without identified driver mutations: chromophobe renal cell carcinoma (44%; 19 out of 43) and pancreatic neuroendocrine cancers (22%; 18 out of 81) (Extended Data Fig. 4a). A notable feature of the missing-driver cases in both tumour types was a remarkably consistent

profile of chromosomal aneuploidy—patterns that have previously been reported^{45,46} (Fig. 3e). The absence of other identified driver mutations in these patients raises the possibility that certain combinations of whole-chromosome gains and losses may be sufficient to initiate a cancer in the absence of more-targeted driver events such as point mutations or fusion genes of focal CNAs.

Even after accounting for technical issues and novel drivers, 5.3% of PCAWG tumours still had no identifiable driver events. In a research setting, in which we are interested in drawing conclusions about populations of patients, the consequences of technical issues that affect occasional samples will be mitigated by sample size. In a clinical setting, in which we are interested in the driver mutations in a specific patient, these issues become substantially more important. Careful and critical appraisal of the whole pipeline—including sample acquisition, genome sequencing, mapping, variant calling and driver annotation, as done

Article

here—should be required for laboratories that offer clinical sequencing of cancer genomes.

Patterns of clustered mutations and SVs

Some somatic mutational processes generate multiple mutations in a single catastrophic event, typically clustered in genomic space, leading to substantial reconfiguration of the genome. Three such processes have previously been described: (1) **chromoplexy**, in which repair of co-occurring double-stranded DNA breaks—typically on different chromosomes—results in shuffled chains of rearrangements^{47,48} (Extended Data Fig. 5a); (2) **kataegis**, a focal hypermutation process that leads to locally clustered nucleotide substitutions, biased towards a single DNA strand^{49–51} (Extended Data Fig. 5b); and (3) **chromothripsis**, in which tens to hundreds of DNA breaks occur simultaneously, clustered on one or a few chromosomes, with near-random stitching together of the resulting fragments^{52–55} (Extended Data Fig. 5c). We characterized the PCAWG genomes for these three processes (Fig. 4).

Chromoplexy events and reciprocal translocations were identified in 467 (17.8%) samples (Fig. 4a, c). Chromoplexy was prominent in prostate adenocarcinoma and lymphoid malignancies, as previously described^{47,48}, and—unexpectedly—thyroid adenocarcinoma. Different genomic loci were recurrently rearranged by chromoplexy across the three tumour types, mediated by positive selection for particular fusion genes or enhancer hijacking events. Of 13 fusion genes or enhancer hijacking events in 48 thyroid adenocarcinomas, at least 4 (31%) were caused by chromoplexy, with a further 4 (31%) part of complexes that contained chromoplexy footprints (Extended Data Fig. 5a). These events generated fusion genes that involved *RET* (two cases) and *NTRK3* (one case)⁵⁶, and the juxtaposition of the oncogene *IGF2BP3* with regulatory elements from highly expressed genes (five cases).

Kataegis events were found in 60.5% of all cancers, with particularly high abundance in lung squamous cell carcinoma, bladder cancer, acral melanoma and sarcomas (Fig. 4a, b). Typically, kataegis comprises C > N mutations in a TpC context, which are probably caused by APOBEC activity^{49–51}, although a T > N conversion in a TpT or CpT process (the affected T is highlighted in bold) attributed to error-prone polymerases has recently been described⁵⁷. The APOBEC signature accounted for 81.7% of kataegis events and correlated positively with *APOBEC3B* expression levels, somatic SV burden and age at diagnosis (Supplementary Fig. 5). Furthermore, 5.7% of kataegis events involved the T > N error-prone polymerase signature and 2.3% of events, most notably in sarcomas, showed cytidine deamination in an alternative CpC or CpC context.

Kataegis events were frequently associated with somatic SV breakpoints (Fig. 4a and Supplementary Fig. 6a), as previously described^{50,52}. Deletions and complex rearrangements were most strongly associated with kataegis, whereas tandem duplications and other simple SV classes were only infrequently associated (Supplementary Fig. 6b). Kataegis inducing predominantly T > N mutations in CpTpT context was enriched near deletions, specifically those in the 10–25 kilobase (kb) range (Supplementary Fig. 6c).

Samples with extreme kataegis burden (more than 30 foci) comprise four types of focal hypermutation (Extended Data Fig. 6): (1) off-target somatic hypermutation and foci of T > N CpTpT, found in B cell non-Hodgkin lymphoma and oesophageal adenocarcinomas, respectively; (2) APOBEC kataegis associated with complex rearrangements, notably found in sarcoma and melanoma; (3) rearrangement-independent APOBEC kataegis on the lagging strand and in early-replicating regions, mainly found in bladder and head and neck cancer; and (4) a mix of the last two types. Kataegis only occasionally led to driver mutations (Supplementary Table 5).

We identified chromothripsis in 587 samples (22.3%), most frequently among sarcoma, glioblastoma, lung squamous cell carcinoma, melanoma and breast adenocarcinoma⁵⁸. Chromothripsis

increased with whole-genome duplications in most cancer types (Extended Data Fig. 7a), as previously shown in medulloblastoma⁵⁸. The most recurrently associated driver was *TP53*⁵² (pan-cancer odds ratio = 3.22; pan-cancer $P = 8.3 \times 10^{-35}$; $q < 0.05$ in breast lobular (odds ratio = 13), colorectal (odds ratio = 25), prostate (odds ratio = 2.6) and hepatocellular (odds ratio = 3.9) cancers; Fisher–Boschloo tests). In two cancer types (osteosarcoma and B cell lymphoma), women had a higher incidence of chromothripsis than men (Extended Data Fig. 7b). In prostate cancer, we observed a higher incidence of chromothripsis in patients with late-onset than early-onset disease⁵⁹ (Extended Data Fig. 7c).

Chromothripsis regions coincided with 3.6% of all identified drivers in PCAWG and around 7% of copy-number drivers (Fig. 4d). These proportions are considerably enriched compared to expectation if selection were not acting on these events (Extended Data Fig. 7d). The majority of coinciding driver events were amplifications (58%), followed by homozygous deletions (34%) and SVs within genes or promoter regions (8%). We frequently observed a ≥ 2 -fold increase or decrease in expression of amplified or deleted drivers, respectively, when these loci were part of a chromothripsis event, compared with samples without chromothripsis (Extended Data Fig. 7e).

Chromothripsis manifested in diverse patterns and frequencies across tumour types, which we categorized on the basis of five characteristics (Fig. 4a). In liposarcoma, for example, chromothripsis events often involved multiple chromosomes, with universal *MDM2* amplification⁶⁰ and co-amplification of *TERT* in 4 of 19 cases (Fig. 4d). By contrast, in glioblastoma the events tended to affect a smaller region on a single chromosome that was distant from the telomere, resulting in focal amplification of *EGFR* and *MDM2* and loss of *CDKN2A*. Acral melanomas frequently exhibited *CCND1* amplification, and lung squamous cell carcinomas *SOX2* amplifications. In both cases, these drivers were more frequently altered by chromothripsis compared with other drivers in the same cancer type and to other cancer types for the same driver (Fig. 4d and Extended Data Fig. 7f). Finally, in chromophobe renal cell carcinoma, chromothripsis nearly always affected chromosome 5 (Supplementary Fig. 7): these samples had breakpoints immediately adjacent to *TERT*, increasing *TERT* expression by 80-fold on average compared with samples without rearrangements ($P = 0.0004$; Mann–Whitney *U*-test).

Timing clustered mutations in evolution

An unanswered question for clustered mutational processes is whether they occur early or late in cancer evolution. To address this, we used molecular clocks to define broad epochs in the life history of each tumour^{49,61}. One transition point is between clonal and subclonal mutations: clonal mutations occurred before, and subclonal mutations after, the emergence of the most-recent common ancestor. In regions with copy-number gains, molecular time can be further divided according to whether mutations preceded the copy-number gain (and were themselves duplicated) or occurred after the gain (and therefore present on only one chromosomal copy)⁷.

Chromothripsis tended to have greater relative odds of being clonal than subclonal, suggesting that it occurs early in cancer evolution, especially in liposarcomas, prostate adenocarcinoma and squamous cell lung cancer (Fig. 5a). As previously reported, chromothripsis was especially common in melanomas⁶². We identified 89 separate chromothripsis events that affected 66 melanomas (61%); 47 out of 89 events affected genes known to be recurrently altered in melanoma⁶³ (Supplementary Table 6). Involvement of a region on chromosome 11 that includes the cell-cycle regulator *CCND1* occurred in 21 cases (10 out of 86 cutaneous, and 11 out of 21 acral or mucosal melanomas), typically combining chromothripsis with amplification (19 out of 21 cases) (Extended Data Fig. 8). Co-involvement of other cancer-associated genes in the same chromothripsis event was also frequent, including

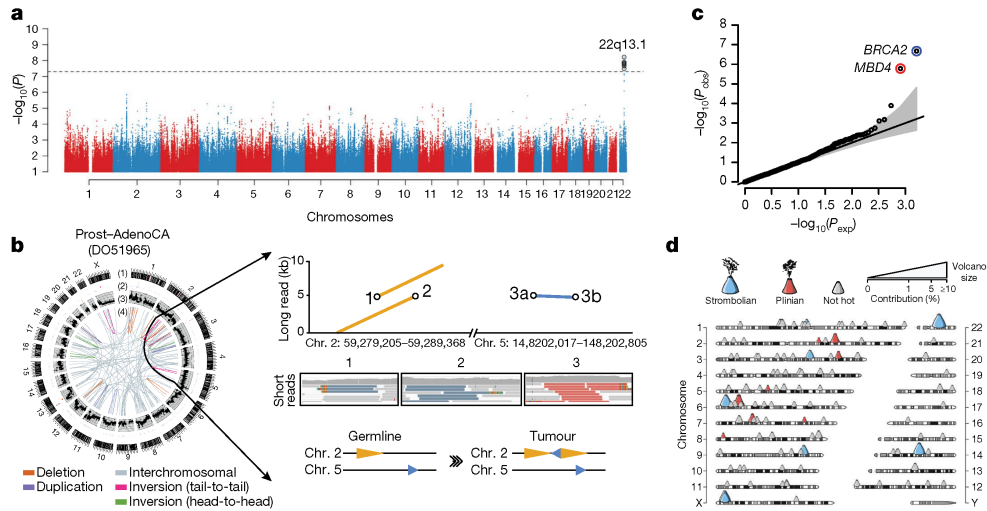


Fig. 6 | Germline determinants of the somatic mutation landscape.

a, Association between common (MAF > 5%) germline variants and somatic APOBEC3B-like mutagenesis in individuals of European ancestry ($n = 1,201$). Two-sided hypothesis testing was performed with PLINK v1.9. To mitigate multiple-hypothesis testing, the significance threshold was set to genome-wide significance ($P < 5 \times 10^{-8}$). **b**, Templated insertion SVs in a *BRCA1*-associated prostate cancer. Left, chromosome bands (1); SVs ≤ 10 megabases (Mb) (2); 1-kb read depth corrected to copy number 0–6 (3); inter- and intrachromosomal SVs > 10 Mb (4). Right, a complex somatic SV composed of a 2.2-kb tandem duplication on chromosome 2 together with a 232-base-pair (bp) inverted templated insertion SV that is derived from chromosome 5 and inserted in between the tandem duplication (bottom). Consensus sequence alignment of locally assembled Oxford Nanopore Technologies long sequencing reads to chromosomes 2 and 5 of the human reference genome (top). Breakpoints are circled and marked as 1 (beginning of tandem duplication), 2 (end of tandem duplication) or 3 (inverted templated insertion). For each breakpoint, the middle panel shows Illumina short reads at SV

breakpoints. **c**, Association between rare germline PTVs (MAF < 0.5%) and somatic CpG mutagenesis (approximately with signature 1) in individuals of European ancestry ($n = 1,201$). Genes highlighted in blue or red were associated with lower or higher somatic mutation rates. Two-sided hypothesis testing was performed using linear-regression models with sex, age at diagnosis and cancer project as variables. To mitigate multiple-hypothesis testing, the significance threshold was set to exome-wide significance ($P < 2.5 \times 10^{-6}$). The black line represents the identity line that would be followed if the observed P values followed the null expectation; the shaded area shows the 95% confidence intervals. **d**, Catalogue of polymorphic germline L1 source elements that are active in cancer. The chromosomal map shows germline source L1 elements as volcano symbols. Each volcano is colour-coded according to the type of source L1 activity. The contribution of each source locus (expressed as a percentage) to the total number of transductions identified in PCAWG tumours is represented as a gradient of volcano size, with top contributing elements exhibiting larger sizes.

Germline effects on somatic mutations

We integrated the set of 88 million germline genetic variant calls with somatic mutations in PCAWG, to study germline determinants of somatic mutation rates and patterns. First, we performed a genome-wide association study of somatic mutational processes with common germline variants (minor allele frequency (MAF) > 5%) in individuals with inferred European ancestry. An independent genome-wide association study was performed in East Asian individuals from Asian cancer genome projects. We focused on two prevalent endogenous mutational processes: spontaneous deamination of 5-methylcytosine at CpG dinucleotides⁵ (signature 1) and activity of the APOBEC3 family of cytidine deaminases⁶⁴ (signatures 2 and 13). No locus reached genome-wide significance ($P < 5 \times 10^{-8}$) for signature 1 (Extended Data Fig. 10a, b). However, a locus at 22q13.1 predicted an APOBEC3B-like mutagenesis at the pan-cancer level⁶⁵ (Fig. 6a). The strongest signal at 22q13.1 was driven by rs12628403, and the minor (non-reference) allele was protective against APOBEC3B-like mutagenesis ($\beta = -0.43$, $P = 5.6 \times 10^{-9}$, MAF = 8.2%, $n = 1,201$ donors) (Extended Data Fig. 10c). This variant tags a common, approximately 30-kb germline SV that deletes the *APOBEC3B* coding sequence and fuses the *APOBEC3B* 3' untranslated region with the coding sequence of *APOBEC3A*. The deletion is known

to increase breast cancer risk and APOBEC mutagenesis in breast cancer genomes^{66,67}. Here, we found that rs12628403 reduces APOBEC3B-like mutagenesis specifically in cancer types with low levels of APOBEC mutagenesis ($\beta_{low} = -0.50$, $P_{low} = 1 \times 10^{-8}$; $\beta_{high} = 0.17$, $P_{high} = 0.2$), and increases APOBEC3A-like mutagenesis in cancer types with high levels of APOBEC mutagenesis ($\beta_{high} = 0.44$, $P_{high} = 8 \times 10^{-4}$; $\beta_{low} = -0.21$, $P_{low} = 0.02$). Moreover, we identified a second, novel locus at 22q13.1 that was associated with APOBEC3B-like mutagenesis across cancer types (rs2142833, $\beta = 0.23$, $P = 1.3 \times 10^{-8}$). We independently validated the association between both loci and APOBEC3B-like mutagenesis using East Asian individuals from Asian cancer genome projects ($\beta_{rs12628403} = 0.57$, $P_{rs12628403} = 4.2 \times 10^{-12}$; $\beta_{rs2142833} = 0.58$, $P_{rs2142833} = 8 \times 10^{-15}$) (Extended Data Fig. 10d). Notably, in a conditional analysis that accounted for rs12628403, we found that rs2142833 and rs12628403 are inherited independently in Europeans ($r^2 < 0.1$), and rs2142833 remained significantly associated with APOBEC3B-like mutagenesis in Europeans ($\beta_{EUR} = 0.17$, $P_{EUR} = 3 \times 10^{-5}$) and East Asians ($\beta_{ASN} = 0.25$, $P_{ASN} = 2 \times 10^{-3}$) (Extended Data Fig. 10e, f). Analysis of donor-matched expression data further suggests that rs2142833 is a *cis*-expression quantitative trait locus (eQTL) for *APOBEC3B* at the pan-cancer level ($\beta = 0.19$, $P = 2 \times 10^{-6}$) (Extended Data Fig. 10g, h), consistent with *cis*-eQTL studies in normal cells^{68,69}.

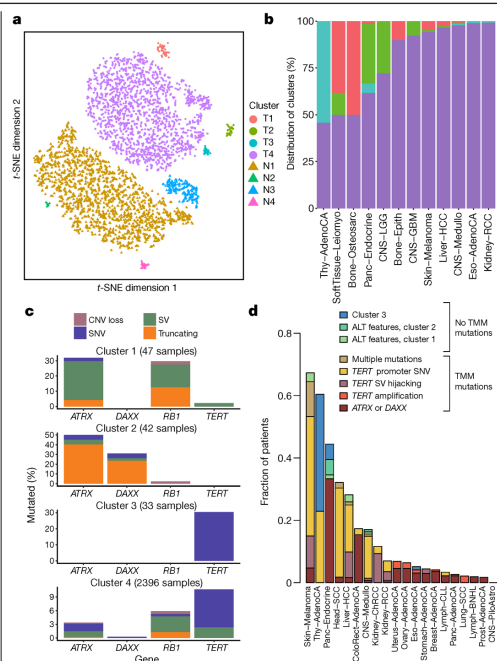


Fig. 7 | Telomere sequence patterns across PCAWG. **a**, Scatter plot of the clusters of telomere patterns identified across PCAWG using *t*-distributed stochastic neighbour embedding (*t*-SNE), based on $n = 2,518$ tumour samples and their matched normal samples. Axes have arbitrary dimensions such that samples with similar telomere profiles are clustered together and samples with dissimilar telomere profiles are far apart with high probability. **b**, Distribution of the four tumour-specific clusters of telomere patterns in selected tumour types from PCAWG. **c**, Distribution of relevant driver mutations associated with alternative lengthening of telomere and normal telomere maintenance across the four clusters. **d**, Distribution of telomere maintenance abnormalities across tumour types with more than 40 patients in PCAWG. Samples were classified as tumour clusters 1–3 if they fell into a relevant cluster without mutations in *TERT*, *ATRX* or *DAXX* and had no ALT phenotype. TMM, telomere maintenance mechanisms.

Second, we performed a rare-variant association study (MAF < 0.5%) to investigate the relationship between germline PTVs and somatic DNA rearrangements in individuals with European ancestry (Extended Data Fig. 11a–c). Germline *BRCA2* and *BRCA1* PTVs were associated with an increased burden of small (less than 10 kb) somatic SV deletions ($P = 1 \times 10^{-9}$) and tandem duplications ($P = 6 \times 10^{-13}$), respectively, corroborating recent studies in breast and ovarian cancer^{30,70}. In PCAWG data, this pattern also extends to other tumour types, including adenocarcinomas of the prostate and pancreas⁶, typically in the setting of biallelic inactivation. In addition, tumours with high levels of small SV tandem duplications frequently exhibited a novel and distinct class of SVs termed ‘cycles of templated insertions’⁶. These complex SV events consist of DNA templates that are copied from across the genome, joined into one contiguous sequence and inserted into a single derivative chromosome. We found a significant association between germline *BRCA1* PTVs and templated insertions at the pan-cancer level ($P = 4 \times 10^{-15}$) (Extended Data Fig. 11d, e). Whole-genome

long-read sequencing data generated for a *BRCA1*-deficient PCAWG prostate tumour verified the small tandem-duplication and templated-insertion SV phenotypes (Fig. 6b). Almost all (20 out of 21) of *BRCA1*-associated tumours with a templated-insertion SV phenotype displayed combined germline and somatic hits in the gene. Together, these data suggest that biallelic inactivation of *BRCA1* is a driver of the templated-insertion SV phenotype.

Third, rare-variant association analysis revealed that patients with germline *MBD4* PTVs had increased rates of somatic C > T mutation rates at CpG dinucleotides ($P < 2.5 \times 10^{-6}$) (Fig. 6c and Extended Data Fig. 11f, g). Analysis of previously published whole-exome sequencing samples from the TCGA ($n = 8,134$) replicated the association between germline *MBD4* PTVs and increased somatic CpG mutagenesis at the pan-cancer level ($P = 7.1 \times 10^{-4}$) (Extended Data Fig. 11h). Moreover, gene-expression profiling revealed a significant but modest correlation between *MBD4* expression and somatic CpG mutation rates between and within PCAWG tumour types (Extended Data Fig. 11i–k). *MBD4* encodes a DNA-repair gene that removes thymidines from T:G mismatches within methylated CpG sites⁷¹, a functionality that would be consistent with a CpG mutational signature in cancer.

Fourth, we assessed long interspersed nuclear elements (LINE-1; L1 hereafter) that mediate somatic retrotransposition events^{72–74}. We identified 114 germline source L1 elements capable of active somatic retrotransposition, including 70 that represent insertions with respect to the human reference genome (Fig. 6d and Supplementary Table 7), and 53 that were tagged by single-nucleotide polymorphisms in strong linkage disequilibrium (Supplementary Table 7). Only 16 germline L1 elements accounted for 67% (2,440 out of 3,669) of all L1-mediated transductions¹⁰ detected in the PCAWG dataset (Extended Data Fig. 12a). These 16 hot-L1 elements followed two broad patterns of somatic activity (8 of each), which we term Strombolian and Plinian in analogy to patterns of volcanic activity. Strombolian L1s are frequently active in cancer, but mediate only small-to-moderate eruptions of somatic L1 activity in cancer samples (Extended Data Fig. 12b). By contrast, Plinian L1s are more rarely seen, but display aggressive somatic activity. Whereas Strombolian elements are typically relatively common (MAF > 2%) and sometimes even fixed in the human population, all Plinian elements were infrequent (MAF ≤ 2%) in PCAWG donors (Extended Data Fig. 12c; $P = 0.001$, Mann–Whitney *U*-test). This dichotomous pattern of activity and allele frequency may reflect differences in age and selective pressures, with Plinian elements potentially inserted into the human germline more recently. PCAWG donors bear on average between 50 and 60 L1 source elements and between 5 and 7 elements with hot activity (Extended Data Fig. 12d), but only 38% (1,075 out of 2,814) of PCAWG donors carried ≥ 1 Plinian element. Some L1 germline source loci caused somatic loss of tumour-suppressor genes (Extended Data Fig. 12e). Many are restricted to individual continental population ancestries (Extended Data Fig. 12f–j).

Replicative immortality

One of the hallmarks of cancer is the ability of cancer to evade cellular senescence²¹. Normal somatic cells typically have finite cell division potential; telomere attrition is one mechanism to limit numbers of mitoses⁷⁵. Cancers enlist multiple strategies to achieve replicative immortality. Overexpression of the telomerase gene, *TERT*, which maintains telomere lengths, is especially prevalent. This can be achieved through point mutations in the promoter that lead to de novo transcription factor binding^{24,37}; hitching *TERT* to highly active regulatory elements elsewhere in the genome^{46,76}; insertions of viral enhancers upstream of the gene^{77,78}; and increased dosage through chromosomal amplification, as we have seen in melanoma (Fig. 5b). In addition, there is an ‘alternative lengthening of telomeres’ (ALT) pathway, in which telomeres are lengthened through homologous recombination, mediated by loss-of-function mutations in the *ATRX* and *DAXX* genes⁷⁹.

Article

As reported in a companion paper¹³, 16% of tumours in the PCAWG dataset exhibited somatic mutations in at least one of *ATRX*, *DAXX* and *TERT*. *TERT* alterations were detected in 270 samples, whereas 128 tumours had alterations in *ATRX* or *DAXX*, of which 71 were protein-truncating. In the companion paper, which focused on describing patterns of ALT and *TERT*-mediated telomere maintenance¹³, 12 features of telomeric sequence were measured in the PCAWG cohort. These included counts of nine variants of the core hexameric sequence, the number of ectopic telomere-like insertions within the genome, the number of genomic breakpoints and telomere length as a ratio between tumour and normal. Here we used the 12 features as an overview of telomere integrity across all tumours in the PCAWG dataset.

On the basis of these 12 features, tumour samples formed 4 distinct subclusters (Fig. 7a and Extended Data Fig. 13a), suggesting that telomere maintenance mechanisms are more diverse than the well-established *TERT* and ALT dichotomy. Clusters C1 (47 tumours) and C2 (42 tumours) were enriched for traits of the ALT pathway—having longer telomeres, more genomic breakpoints, more ectopic telomere insertions and variant telomere sequence motifs (Supplementary Fig. 9). C1 and C2 were distinguished from one another by the latter having a considerable increase in the number of TTCGGG and TGAGGG variant motifs among the telomeric hexamers. Thyroid adenocarcinomas were markedly enriched among C3 samples (26 out of 33 C3 samples; $P < 10^{-36}$); the C1 cluster (ALT subtype 1) was common among sarcomas; and both pancreatic endocrine neoplasms and low-grade gliomas had a high proportion of samples in the C2 cluster (ALT subtype 2) (Fig. 7b). Notably, some of the thyroid adenocarcinomas and pancreatic neuroendocrine tumours that cluster together (cluster C3) had matched normal samples that also cluster together (normal cluster N3) (Extended Data Fig. 13a) and which share common properties. For example, the GTAGGG repeat was overrepresented among samples in this group (Supplementary Fig. 10).

Somatic driver mutations were also unevenly distributed across the four clusters (Fig. 7c). C1 tumours were enriched for *RBI* mutations or SVs ($P = 3 \times 10^{-5}$), as well as frequent SVs that affected *ATRX* ($P = 6 \times 10^{-4}$), but not *DAXX*. *RBI* and *ATRX* mutations were largely mutually exclusive (Extended Data Fig. 13b). By contrast, C2 tumours were enriched for somatic point mutations in *ATRX* and *DAXX* ($P = 6 \times 10^{-5}$), but not *RBI*. The enrichment of *RBI* mutations in C1 remained significant when only leiomyosarcomas and osteosarcomas were considered, confirming that this enrichment is not merely a consequence of the different distribution of tumour types across clusters. C3 samples had frequent *TERT* promoter mutations (30%; $P = 2 \times 10^{-6}$).

There was a marked predominance of *RBI* mutations in C1. Nearly a third of the samples in C1 contained an *RBI* alteration, which were evenly distributed across truncating SNVs, SVs and shallow deletions (Extended Data Fig. 13c). Previous research has shown that *RBI* mutations are associated with long telomeres in the absence of *TERT* mutations and *ATRX* inactivation⁸⁰, and studies using mouse models have shown that knockout of Rb-family proteins causes elongated telomeres⁸¹. The association with the C1 cluster here suggests that *RBI* mutations can represent another route to activating the ALT pathway, which has subtly different properties of telomeric sequence compared with the inactivation of *DAXX*—these fall almost exclusively in cluster C2.

Tumour types with the highest rates of abnormal telomere maintenance mechanisms often originate in tissues that have low endogenous replicative activity (Fig. 7d). In support of this, we found an inverse correlation between previously estimated rates of stem cell division across tissues⁸² and the frequency of telomere maintenance abnormalities ($P = 0.01$, Poisson regression) (Extended Data Fig. 13d). This suggests that restriction of telomere maintenance is an important tumour-suppression mechanism, particularly in tissues with low steady-state cellular proliferation, in which a clone must overcome this constraint to achieve replicative immortality.

Conclusions and future perspectives

The resource reported in this paper and its companion papers has yielded insights into the nature and timing of the many mutational processes that shape large- and small-scale somatic variation in the cancer genome; the patterns of selection that act on these variations; the widespread effect of somatic variants on transcription; the complementary roles of the coding and non-coding genome for both germline and somatic mutations; the ubiquity of intratumoral heterogeneity; and the distinctive evolutionary trajectory of each cancer type. Many of these insights can be obtained only from an integrated analysis of all classes of somatic mutation on a whole-genome scale, and would not be accessible with, for example, targeted exome sequencing.

The promise of precision medicine is to match patients to targeted therapies using genomics. A major barrier to its evidence-based implementation is the daunting heterogeneity of cancer chronicled in these papers, from tumour type to tumour type, from patient to patient, from clone to clone and from cell to cell. Building meaningful clinical predictors from genomic data can be achieved, but will require knowledge banks comprising tens of thousands of patients with comprehensive clinical characterization⁸³. As these sample sizes will be too large for any single funding agency, pharmaceutical company or health system, international collaboration and data sharing will be required. The next phase of ICGC, ICGC-ARGO (<https://www.icgc-argo.org/>), will bring the cancer genomics community together with healthcare providers, pharmaceutical companies, data science and clinical trials groups to build comprehensive knowledge banks of clinical outcome and treatment data from patients with a wide variety of cancers, matched with detailed molecular profiling.

Extending the story begun by TCGA, ICGC and other cancer genomics projects, the PCAWG has brought us closer to a comprehensive narrative of the causal biological changes that drive cancer phenotypes. We must now translate this knowledge into sustainable, meaningful clinical treatments.


Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-1969-6>.

1. Pleasance, E. D. et al. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196 (2010).
2. Pleasance, E. D. et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184–190 (2010).
3. Ley, T. J. et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66–72 (2008).
4. Rheinbay, E. et al. Analyses of non-coding somatic drivers in 2,693 cancer whole genomes. *Nature* <https://doi.org/10.1038/s41586-020-1965-x> (2020).
5. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* <https://doi.org/10.1038/s41586-020-1943-3> (2020).
6. Li, Y. et al. Patterns of somatic structural variation in human cancer genomes. *Nature* <https://doi.org/10.1038/s41586-019-1913-9> (2020).
7. Gerstung, M. et al. The evolutionary history of 2,658 cancers. *Nature* <https://doi.org/10.1038/s41586-019-1907-7> (2020).
8. PCAWG Transcriptome Core Group et al. Genomic basis of RNA alterations in cancer. *Nature* <https://doi.org/10.1038/s41586-020-1970-0> (2020).
9. Zhang, Y. et al. High-coverage whole-genome analysis of 1,220 cancers reveals hundreds of genes deregulated by rearrangement-mediated cis-regulatory alterations. *Nat. Commun.* <https://doi.org/10.1038/s41467-019-1388-5-w> (2020).
10. Rodriguez-Martin, B. et al. Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nat. Genet.* <https://doi.org/10.1038/s41588-019-0562-0> (2020).
11. Zapatka, M. et al. The landscape of viral associations in human cancers. *Nat. Genet.* <https://doi.org/10.1038/s41588-019-0558-9> (2020).
12. Jiao, W. et al. A deep learning system can accurately classify primary and metastatic cancers based on patterns of passenger mutations. *Nat. Commun.* <https://doi.org/10.1038/s41467-019-13825-8> (2020).

13. Sieverling, L. et al. Genomic footprints of activated telomere maintenance mechanisms in cancer. *Nat. Commun.* <https://doi.org/10.1038/s41467-019-13824-9> (2020).
14. Yuan, Y. et al. Comprehensive molecular characterization of mitochondrial genomes in human cancers. *Nat. Genet.* <https://doi.org/10.1038/s41588-019-0557-x> (2020).
15. Alkderri, K. C. et al. Chromatin folding domains disruptions by somatic genomic rearrangements in human cancers. *Nat. Genet.* <https://doi.org/10.1038/s41588-019-0564-y> (2020).
16. Reysa, M. A. et al. Pathway and network analysis of more than 2,500 whole cancer genomes. *Nat. Commun.* <https://doi.org/10.1038/s41467-020-14351-8> (2020).
17. Bailey, M. H. et al. Retrospective evaluation of whole exome and genome mutation calls in 746 cancer samples. *Nat. Commun.* (2020).
18. Cortes-Ciriano, J. et al. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nat. Genet.* <https://doi.org/10.1038/s41588-019-0576-7> (2020).
19. Bray, F., Ren, J.-S., Masuyer, E. & Ferlay, J. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. *Int. J. Cancer* **132**, 1133–1145 (2013).
20. Tarver, T. Cancer Facts & Figures 2012. American Cancer Society (ACS). *J. Consum. Health Internet* **16**, 366–367 (2012).
21. Hanahan, D. & Weinberg, R. A. Hallmarks of cancers: the next generation. *Cell* **144**, 646–674 (2011).
22. International Cancer Genome Consortium. International network of cancer genome projects. *Nature* **464**, 993–998 (2010).
23. Bailey, M. H. et al. Comprehensive characterization of cancer driver genes and mutations. *Cell* **173**, 371–385 (2018).
24. Sanchez-Vega, F. et al. Onco-genic signaling pathways in The Cancer Genome Atlas. *Cell* **173**, 321–337 (2018).
25. Hoadley, K. A. et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* **173**, 291–304 (2018).
26. Stein, L. D., Knoppers, B. M., Campbell, P., Getz, G. & Korbel, J. O. Data analysis: create a cloud commons. *Nature* **523**, 149–151 (2015).
27. Phillips, M. et al. Genomics: data sharing needs international code of conduct. *Nature* <https://doi.org/10.1038/d41586-020-00082-9> (2020).
28. Krochmalski, J. *Developing with Docker* (Packt Publishing, 2016).
29. Welch, J. S. et al. The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264–278 (2012).
30. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
31. Meier, B. et al. C. elegans whole-genome sequencing reveals mutational signatures related to carcinogens and DNA repair deficiency. *Genome Res.* **24**, 1624–1636 (2014).
32. Martincorena, I. et al. Universal patterns of selection in cancer and somatic tissues. *Cell* **171**, 1029–1041 (2017).
33. Tamborero, D. et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med.* **10**, 25 (2018).
34. Huang, F. W. et al. Highly recurrent TERT promoter mutations in human melanoma. *Science* **339**, 957–959 (2013).
35. Rheinbay, E. et al. Recurrent and functional regulatory mutations in breast cancer. *Nature* **547**, 55–60 (2017).
36. Fredriksson, M. J., Ny, L., Nilsson, J. A. & Larsson, E. Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types. *Nat. Genet.* **46**, 1258–1263 (2014).
37. Horn, S. et al. TERT promoter mutations in familial and sporadic melanoma. *Science* **339**, 959–961 (2013).
38. Ciriello, G. et al. Emerging landscape of oncogenic signatures across human cancers. *Nat. Genet.* **45**, 1127–1133 (2013).
39. Rahman, N. Realizing the promise of cancer predisposition genes. *Nature* **505**, 302–308 (2014).
40. Pearl, L. H., Schierz, A. C., Ward, S. E., Al-Lazikani, B. & Pearl, F. M. G. Therapeutic opportunities within the DNA damage response. *Nat. Rev. Cancer* **15**, 166–180 (2015).
41. Taylor-Weiner, A. et al. DeFIN: overcoming tumor-in-normal contamination. *Nat. Methods* **15**, 531–534 (2018).
42. Fujimoto, A. et al. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat. Genet.* **48**, 500–509 (2016).
43. Shlush, L. I. Age-related clonal hematopoiesis. *Blood* **131**, 496–504 (2018).
44. Northcott, P. A. et al. The whole-genome landscape of medulloblastoma subtypes. *Nature* **547**, 311–317 (2017).
45. Scarpa, A. et al. Whole-genome landscape of pancreatic neuroendocrine tumours. *Nature* **549**, 65–71 (2017).
46. Davis, C. F. et al. The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell* **26**, 319–330 (2014).
47. Berger, M. F. et al. The genomic complexity of primary human prostate cancer. *Nature* **470**, 214–220 (2011).
48. Baca, S. C. et al. Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666–677 (2013).
49. Nik-Zainal, S. et al. The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).
50. Nik-Zainal, S. et al. Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
51. Roberts, S. A. et al. Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Mol. Cell* **46**, 424–435 (2012).
52. Rausch, T. et al. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell* **148**, 59–71 (2012).
53. Stephens, P. J. et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).
54. Korbel, J. O. & Campbell, P. J. Criteria for inference of chromothripsis in cancer genomes. *Cell* **152**, 1226–1236 (2013).
55. Zhang, C.-Z. et al. Chromothripsis from DNA damage in micronuclei. *Nature* **522**, 179–184 (2015).
56. The Cancer Genome Atlas Research Network. Integrated genomic characterization of papillary thyroid carcinoma. *Cell* **159**, 676–690 (2014).
57. Supek, F. & Lehner, B. Clustered mutation signatures reveal that error-prone DNA repair targets mutations to active genes. *Cell* **170**, 534–547 (2017).
58. Mardin, B. R. et al. A cell-based model system links chromothripsis with hyperploidy. *Mol. Syst. Biol.* **11**, 828 (2015).
59. Wesschenfeldt, J. et al. Integrative genomic analyses reveal an androgen-driven somatic alteration landscape in early-onset prostate cancer. *Cancer Cell* **23**, 159–170 (2013).
60. Garsed, D. W. et al. The architecture and evolution of cancer neochromosomes. *Cancer Cell* **26**, 653–667 (2014).
61. Durinok, S. et al. Temporal dissection of tumorigenesis in primary cancers. *Cancer Discov.* **1**, 137–143 (2011).
62. Hayward, N. K. et al. Whole-genome landscapes of major melanoma subtypes. *Nature* **545**, 175–180 (2017).
63. The Cancer Genome Atlas Network. Genomic classification of cutaneous melanoma. *Cell* **161**, 1681–1696 (2015).
64. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
65. Chan, K. et al. An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. *Nat. Genet.* **47**, 1067–1072 (2015).
66. Nik-Zainal, S. et al. Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nat. Genet.* **46**, 487–491 (2014).
67. Middlebrooks, C. D. et al. Association of germline variants in the APOBEC3 region with cancer risk and enrichment with APOBEC-signature mutations in tumors. *Nat. Genet.* **48**, 1330–1338 (2016).
68. Westra, H.-J. et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
69. Stranger, B. E. et al. Population genomics of human gene expression. *Nat. Genet.* **39**, 1217–1224 (2007).
70. Menghi, F. et al. The tandem duplicator phenotype as a distinct genomic configuration in cancer. *Proc. Natl. Acad. Sci. USA* **113**, E2373–E2382 (2016).
71. Hendrich, B., Hardeland, J., Ng, H. H., Irlin, J. & Bird, A. The thymine glycosylase MBD4 can bind to the product of desamination at methylated CpG sites. *Nature* **401**, 301–304 (1999).
72. Lee, E. et al. Landscape of somatic retrotransposition in human cancers. *Science* **337**, 967–971 (2012).
73. Tubio, J. M. C. et al. Extensive transduction of nonreplicative DNA mediated by L1 retrotransposition in cancer genomes. *Science* **345**, 1251343–1251343 (2014).
74. Helman, E. et al. Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. *Genome Res.* **24**, 1053–1063 (2014).
75. Shay, J. W. & Wright, W. E. Hayflick, his limit, and cellular ageing. *Nat. Rev. Mol. Cell Biol.* **1**, 72–76 (2000).
76. Peifer, M. et al. Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature* **526**, 700–704 (2015).
77. Totoki, Y. et al. Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. *Nat. Genet.* **46**, 1267–1273 (2014).
78. Patelin-Béchet, P. et al. Hepatitis B virus-related insertional mutagenesis occurs frequently in human liver cancers and recurrently targets human telomerase gene. *Oncogene* **22**, 3911–3916 (2003).
79. Heaphy, C. M. et al. Prevalence of the alternative lengthening of telomeres telomere maintenance mechanism in human cancer subtypes. *Am. J. Pathol.* **179**, 1608–1615 (2011).
80. Barthel, F. P. et al. Systematic analysis of telomere length and somatic alterations in 31 cancer types. *Nat. Genet.* **49**, 349–357 (2017).
81. Garcia-Cao, M., Gonzalo, S., Dean, D. & Blasco, M. A. A role for the Rb family of proteins in controlling telomere length. *Nat. Genet.* **32**, 415–419 (2002).
82. Tomasetti, C. & Vogelstein, B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78–81 (2015).
83. Gerstung, M. et al. Precision oncology for acute myeloid leukemia using a knowledge bank approach. *Nat. Genet.* **49**, 332–340 (2017).
84. O'Connor, B. D. et al. The Dockstore: enabling modular, community-focused sharing of Docker-based genomics tools and workflows. *FAOOpenRes.* **6**, 52 (2017).
85. Zhang, J. et al. The International Cancer Genome Consortium Data Portal. *Nat. Biotechnol.* **37**, 367–369 (2019).
86. Miller, C. A., Qiao, Y., DiSera, T., D'Astous, B. & Marth, G. T. bamJobio: a web-based, real-time, sequence alignment file inspector. *Nat. Methods* **11**, 1189–1189 (2014).
87. Goldman, M. et al. The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. Preprint at <https://www.biorxiv.org/content/10.1101/326470v6> (2019).
88. Papatheodorou, I. et al. Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Res.* **46**, D246–D251 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

Reference List

- A A SANDBERG, A. & HOSSFELD, D. K. 1970. Chromosomal Abnormalities in Human Neoplasia. *Annual Review of Medicine*, 21, 379-408.
- ABBOSH, C., BIRKBAK, N. J., WILSON, G. A., JAMAL-HANJANI, M., CONSTANTIN, T., SALARI, R., LE QUESNE, J., MOORE, D. A., VEERIAH, S., ROSENTHAL, R., MARAFIOTI, T., KIRKIZLAR, E., WATKINS, T. B. K., MCGRANAHAN, N., WARD, S., MARTINSON, L., RILEY, J., FRAIOLI, F., AL BAKIR, M., GRONROOS, E., ZAMBRANA, F., ENDOZO, R., BI, W. L., FENNESSY, F. M., SPONER, N., JOHNSON, D., LAYCOCK, J., SHAFI, S., CZYZEWSKA-KHAN, J., ROWAN, A., CHAMBERS, T., MATTHEWS, N., TURAJLIC, S., HILEY, C., LEE, S. M., FORSTER, M. D., AHMAD, T., FALZON, M., BORG, E., LAWRENCE, D., HAYWARD, M., KOLVEKAR, S., PANAGIOTOPOULOS, N., JANES, S. M., THAKRAR, R., AHMED, A., BLACKHALL, F., SUMMERS, Y., HAFEZ, D., NAIK, A., GANGULY, A., KAREHT, S., SHAH, R., JOSEPH, L., MARIE QUINN, A., CROSBIE, P. A., NAIDU, B., MIDDLETON, G., LANGMAN, G., TROTTER, S., NICOLSON, M., REMMEN, H., KERR, K., CHETTY, M., GOMERSALL, L., FENNELL, D. A., NAKAS, A., RATHINAM, S., ANAND, G., KHAN, S., RUSSELL, P., EZHIL, V., ISMAIL, B., IRVIN-SELLERS, M., PRAKASH, V., LESTER, J. F., KORNASZEWSKA, M., ATTANOOS, R., ADAMS, H., DAVIES, H., OUKRIF, D., AKARCA, A. U., HARTLEY, J. A., LOWE, H. L., LOCK, S., ILES, N., BELL, H., NGAI, Y., ELGAR, G., SZALLASI, Z., SCHWARZ, R. F., HERRERO, J., STEWART, A., QUEZADA, S. A., PEGGS, K. S., VAN LOO, P., DIVE, C., LIN, C. J., RABINOWITZ, M., AERTS, H., et al. 2017. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature*, 545, 446-451.
- ABELSON, S., COLLORD, G., NG, S. W. K., WEISSBROD, O., MENDELSON COHEN, N., NIEMEYER, E., BARDA, N., ZUZARTE, P. C., HEISLER, L., SUNDARAVADANAM, Y., LUBEN, R., HAYAT, S., WANG, T. T., ZHAO, Z., CIRLAN, I., PUGH, T. J., SOAVE, D., NG, K., LATIMER, C., HARDY, C., RAINE, K., JONES, D., HOULT, D., BRITTEN, A., MCPHERSON, J. D., JOHANSSON, M., MBABAALI, F., EAGLES, J., MILLER, J. K., PASTERNAK, D., TIMMS, L., KRZYZANOWSKI, P., AWADALLA, P., COSTA, R., SEGAL, E., BRATMAN, S. V., BEER, P., BEHJATI, S., MARTINCORENA, I., WANG, J. C. Y., BOWLES, K. M., QUIROS, J. R., KARAKATSANI, A., LA VECCHIA, C., TRICHOPOULOU, A., SALAMANCA-FERNANDEZ, E., HUERTA, J. M., BARRICARTE, A., TRAVIS, R. C., TUMINO, R., MASALA, G., BOEING, H., PANICO, S., KAKS, R., KRAMER, A., SIERI, S., RIBOLI, E., VINEIS, P., FOLL, M., MCKAY, J., POLIDORO, S., SALA, N., KHAW, K. T., VERMEULEN, R., CAMPBELL, P. J., PAPAEMMANUIL, E., MINDEN, M. D., TANAY, A., BALICER, R. D., WAREHAM, N. J., GERSTUNG, M., DICK, J. E., BRENNAN, P., VASSILIOU, G. S. & SHLUSH, L. I. 2018. Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature*, 559, 400-404.
- ABLA, O., ROLLINS, B. & LADISCH, S. 2019. Langerhans cell histiocytosis: progress and controversies. *Br J Haematol*, 187, 559-562.

- AGARAM, N. P., ZHANG, L., COTZIA, P. & ANTONESCU, C. R. 2018. Expanding the Spectrum of Genetic Alterations in Pseudomyogenic Hemangioendothelioma With Recurrent Novel ACTB-FOSB Gene Fusions. *Am J Surg Pathol*, 42, 1653-1661.
- AGARWAL, R., CHAN, Y.-C., TAM, C. S., HUNTER, T., VASSILIADIS, D., TEH, C. E., THIJSEN, R., YEH, P., WONG, S. Q., FTOUNI, S., LAM, E. Y. N., ANDERSON, M. A., POTT, C., GILAN, O., BELL, C. C., KNEZEVIC, K., BLOMBERG, P., RAYEROUX, K., ZORDAN, A., LI, J., HUANG, D. C. S., WALL, M., SEYMOUR, J. F., GRAY, D. H. D., ROBERTS, A. W., DAWSON, M. A. & DAWSON, S.-J. 2019. Dynamic molecular monitoring reveals that SWI-SNF mutations mediate resistance to ibrutinib plus venetoclax in mantle cell lymphoma. *Nature Medicine*, 25, 119-129.
- AL-IBRAHEEMI, A., INWARDS, C. Y., ZREIK, R. T., WENGER, D. E., JENKINS, S. M., CARTER, J. M., BOLAND, J. M., ROSE, P. S., JIN, L., OLIVEIRA, A. M. & FRITCHIE, K. J. 2016. Histologic Spectrum of Giant Cell Tumor (GCT) of Bone in Patients 18 Years of Age and Below: A Study of 63 Patients. *Am J Surg Pathol*, 40, 1702-1712.
- ALBERGHINI, M., KLISKEY, K., KRENACS, T., PICCI, P., KINDBLOM, L., FORSYTH, R. & ATHANASOU, N. A. 2010. Morphological and immunophenotypic features of primary and metastatic giant cell tumour of bone. *Virchows Archiv*, 456, 97-103.
- ALEXANDROV, L. B., KIM, J., HARADHVALA, N. J., HUANG, M. N., NG, A. W., WU, Y., BOOT, A., COVINGTON, K. R., GORDENIN, D. A., BERGSTROM, E. N., ISLAM, S. M. A., LOPEZ-BIGAS, N., KLIMCZAK, L. J., MCPHERSON, J. R., MORGANELLA, S., SABARINATHAN, R., WHEELER, D. A., MUSTONEN, V., GETZ, G., ROZEN, S. G. & STRATTON, M. R. 2019. The Repertoire of Mutational Signatures in Human Cancer. *bioRxiv*, 322859.
- ALEXANDROV, L. B., NIK-ZAINAL, S., WEDGE, D. C., APARICIO, S. A., BEHJATI, S., BIANKIN, A. V., BIGNELL, G. R., BOLLI, N., BORG, A., BORRESEN-DALE, A. L., BOYAUULT, S., BURKHARDT, B., BUTLER, A. P., CALDAS, C., DAVIES, H. R., DESMEDT, C., EILS, R., EYFJORD, J. E., FOEKENS, J. A., GREAVES, M., HOSODA, F., HUTTER, B., ILICIC, T., IMBEAUD, S., IMIELINSKI, M., JAGER, N., JONES, D. T., JONES, D., KNAPPSKOG, S., KOOL, M., LAKHANI, S. R., LOPEZ-OTIN, C., MARTIN, S., MUNSHI, N. C., NAKAMURA, H., NORTHCOTT, P. A., PAJIC, M., PAPAEMMANUIL, E., PARADISO, A., PEARSON, J. V., PUENTE, X. S., RAINE, K., RAMAKRISHNA, M., RICHARDSON, A. L., RICHTER, J., ROSENSTIEL, P., SCHLESNER, M., SCHUMACHER, T. N., SPAN, P. N., TEAGUE, J. W., TOTOKI, Y., TUTT, A. N., VALDES-MAS, R., VAN BUUREN, M. M., VAN 'T VEER, L., VINCENT-SALOMON, A., WADDELL, N., YATES, L. R., AUSTRALIAN PANCREATIC CANCER GENOME, I., CONSORTIUM, I. B. C., CONSORTIUM, I. M.-S., PEDBRAIN, I., ZUCMAN-ROSSI, J., FUTREAL, P. A., MCDERMOTT, U., LICHTER, P., MEYERSON, M., GRIMMOND, S. M., SIEBERT, R., CAMPO, E., SHIBATA, T., PFISTER, S. M., CAMPBELL, P. J. & STRATTON, M. R. 2013. Signatures of mutational processes in human cancer. *Nature*, 500, 415-21.
- ALIX-PANABIÈRES, C. & PANTEL, K. 2014. Challenges in circulating tumour cell research. *Nat Rev Cancer*, 14, 623-31.

- AMARY, F., BERISHA, F., YE, H., GUPTA, M., GUTTERIDGE, A., BAUMHOER, D., GIBBONS, R., TIRABOSCO, R., O'DONNELL, P. & FLANAGAN, A. M. 2017. H3F3A (Histone 3.3) G34W Immunohistochemistry: A Reliable Marker Defining Benign and Malignant Giant Cell Tumor of Bone. *Am J Surg Pathol*, 41, 1059-1068.
- AMARY, F., MARKERT, E., BERISHA, F., YE, H., GERRAND, C., COOL, P., TIRABOSCO, R., LINDSAY, D., PILLAY, N., O'DONNELL, P., BAUMHOER, D. & FLANAGAN, A. M. 2019a. FOS Expression in Osteoid Osteoma and Osteoblastoma: A Valuable Ancillary Diagnostic Tool. *Am J Surg Pathol*, 43, 1661-1667.
- AMARY, F., PEREZ-CASANOVA, L., YE, H., COTTONE, L., STROBL, A. C., COOL, P., MIRANDA, E., BERISHA, F., ASTON, W., ROCHA, M., O'DONNELL, P., PILLAY, N., TIRABOSCO, R., BAUMHOER, D., HOOKWAY, E. S. & FLANAGAN, A. M. 2019b. Synovial chondromatosis and soft tissue chondroma: extraosseous cartilaginous tumor defined by FN1 gene rearrangement. *Mod Pathol*, 32, 1762-1771.
- AMARY, M. F., BACSI, K., MAGGIANI, F., DAMATO, S., HALAI, D., BERISHA, F., POLLOCK, R., O'DONNELL, P., GRIGORIADIS, A., DISS, T., ESKANDARPOUR, M., PRESNEAU, N., HOGENDOORN, P. C., FUTREAL, A., TIRABOSCO, R. & FLANAGAN, A. M. 2011. IDH1 and IDH2 mutations are frequent events in central chondrosarcoma and central and periosteal chondromas but not in other mesenchymal tumours. *J Pathol*, 224, 334-43.
- AMARY, M. F., BERISHA, F., MOZELA, R., GIBBONS, R., GUTTRIDGE, A., O'DONNELL, P., BAUMHOER, D., TIRABOSCO, R. & FLANAGAN, A. M. 2016. The H3F3 K36M mutant antibody is a sensitive and specific marker for the diagnosis of chondroblastoma. *Histopathology*, 69, 121-7.
- ANDERS, S., PYL, P. T. & HUBER, W. 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31, 166-9.
- ANDERSON, N. D., DE BORJA, R., YOUNG, M. D., FULIGNI, F., ROSIC, A., ROBERTS, N. D., HAJJAR, S., LAYEGHIFARD, M., NOVOKMET, A., KOWALSKI, P. E., ANAKA, M., DAVIDSON, S., ZARREI, M., ID SAID, B., SCHREINER, L. C., MARCHAND, R., SITTE, J., GOKGOZ, N., BRUNGA, L., GRAHAM, G. T., FULLAM, A., PILLAY, N., TORETSKY, J. A., YOSHIDA, A., SHIBATA, T., METZLER, M., SOMERS, G. R., SCHERER, S. W., FLANAGAN, A. M., CAMPBELL, P. J., SCHIFFMAN, J. D., SHAGO, M., ALEXANDROV, L. B., WUNDER, J. S., ANDRULIS, I. L., MALKIN, D., BEHJATI, S. & SHLIEN, A. 2018. Rearrangement bursts generate canonical gene fusions in bone and soft tissue tumors. *Science*, 361, eaam8419.
- ANTONESCUCU, C. R., CHEN, H. W., ZHANG, L., SUNG, Y. S., PANICEK, D., AGARAM, N. P., DICKSON, B. C., KRAUSZ, T. & FLETCHER, C. D. 2014. ZFP36-FOSB fusion defines a subset of epithelioid hemangioma with atypical features. *Genes Chromosomes Cancer*, 53, 951-9.
- AOKI, J., TANIKAWA, H., ISHII, K., SEO, G. S., KARAKIDA, O., SONE, S., ICHIKAWA, T. & KACHI, K. 1996. MR findings indicative of hemosiderin in giant-cell tumor of bone: frequency, cause, and diagnostic significance. *American Journal of Roentgenology*, 166, 145-148.
- ARYEE, M. J., JAFFE, A. E., CORRADA-BRAVO, H., LADD-ACOSTA, C., FEINBERG, A. P., HANSEN, K. D. & IRIZARRY, R. A. 2014. Minfi: a flexible

- and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*, 30, 1363-1369.
- ASHWORTH, A. 2008. A synthetic lethal therapeutic approach: poly(ADP) ribose polymerase inhibitors for the treatment of cancers deficient in DNA double-strand break repair. *J Clin Oncol*, 26, 3785-90.
- ATESOK, K. I., ALMAN, B. A., SCHEMITSCH, E. H., PEYSER, A. & MANKIN, H. 2011. Osteoid osteoma and osteoblastoma. *J Am Acad Orthop Surg*, 19, 678-89.
- ATKINS, G. J., BOURALEXIS, S., HAYNES, D. R., GRAVES, S. E., GEARY, S. M., EVDOKIOU, A., ZANNETTINO, A. C., HAY, S. & FINDLAY, D. M. 2001. Osteoprotegerin inhibits osteoclast formation and bone resorbing activity in giant cell tumors of bone. *Bone*, 28, 370-7.
- ATKINS, G. J., HAYNES, D. R., GRAVES, S. E., EVDOKIOU, A., HAY, S., BOURALEXIS, S. & FINDLAY, D. M. 2000. Expression of osteoclast differentiation signals by stromal elements of giant cell tumors. *J Bone Miner Res*, 15, 640-9.
- AURIAS, A., RIMBAUT, C., BUFFE, D., ZUCKER, J. M. & MAZABRAUD, A. 1984. Translocation involving chromosome 22 in Ewing's sarcoma. A cytogenetic study of four fresh tumors. *Cancer Genet Cytogenet*, 12, 21-5.
- BACA, SYLVAN C., PRANDI, D., LAWRENCE, MICHAEL S., MOSQUERA, JUAN M., ROMANEL, A., DRIER, Y., PARK, K., KITABAYASHI, N., MACDONALD, THERESA Y., GHANDI, M., VAN ALLEN, E., KRYUKOV, GREGORY V., SBONER, A., THEURILLAT, J.-P., SOONG, T. D., NICKERSON, E., AUCLAIR, D., TEWARI, A., BELTRAN, H., ONOFRIO, ROBERT C., BOYSEN, G., GUIDUCCI, C., BARBIERI, CHRISTOPHER E., CIBULSKIS, K., SIVACHENKO, A., CARTER, SCOTT L., SAKSENA, G., VOET, D., RAMOS, ALEX H., WINCKLER, W., CIPICCHIO, M., ARDLIE, K., KANTOFF, PHILIP W., BERGER, MICHAEL F., GABRIEL, STACEY B., GOLUB, TODD R., MEYERSON, M., LANDER, ERIC S., ELEMENTO, O., GETZ, G., DEMICHELIS, F., RUBIN, MARK A. & GARRAWAY, LEVI A. 2013. Punctuated Evolution of Prostate Cancer Genomes. *Cell*, 153, 666-677.
- BAILEY, M. H., TOKHEIM, C., PORTA-PARDO, E., SENGUPTA, S., BERTRAND, D., WEERASINGHE, A., COLAPRICO, A., WENDL, M. C., KIM, J., REARDON, B., NG, P. K., JEONG, K. J., CAO, S., WANG, Z., GAO, J., GAO, Q., WANG, F., LIU, E. M., MULARONI, L., RUBIO-PEREZ, C., NAGARAJAN, N., CORTES-CIRIANO, I., ZHOU, D. C., LIANG, W. W., HESS, J. M., YELLAPANTULA, V. D., TAMBORERO, D., GONZALEZ-PEREZ, A., SUPHAVILAI, C., KO, J. Y., KHURANA, E., PARK, P. J., VAN ALLEN, E. M., LIANG, H., LAWRENCE, M. S., GODZIK, A., LOPEZ-BIGAS, N., STUART, J., WHEELER, D., GETZ, G., CHEN, K., LAZAR, A. J., MILLS, G. B., KARCHIN, R. & DING, L. 2018. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell*, 173, 371-385.e18.
- BAKER, A. C., REZEANU, L., KLEIN, M. J., PITT, M. J., BUECKER, P., HERSH, J. H., BUCHINO, J. J. & SIEGAL, G. P. 2010. Aggressive osteoblastoma: a case report involving a unique chromosomal aberration. *Int J Surg Pathol*, 18, 219-24.
- BAKHOUM, S. F. & LANDAU, D. A. 2017. Cancer Evolution: No Room for Negative Selection. *Cell*, 171, 987-989.
- BARAULT, L., AMATU, A., SIRAVEGNA, G., PONZETTI, A., MORAN, S., CASSINGENA, A., MUSSOLIN, B., FALCOMATA, C., BINDER, A. M.,

- CRISTIANO, C., ODDO, D., GUARRERA, S., CANCELLIERE, C., BUSTREO, S., BENCARDINO, K., MADEN, S., VANZATI, A., ZAVATTARI, P., MATULLO, G., TRUINI, M., GRADY, W. M., RACCA, P., MICHELS, K. B., SIENA, S., ESTELLER, M., BARDELLI, A., SARTORE-BIANCHI, A. & DI NICOLANTONIO, F. 2018. Discovery of methylated circulating DNA biomarkers for comprehensive non-invasive monitoring of treatment response in metastatic colorectal cancer. *Gut*, 67, 1995-2005.
- BARLOW, E., DAVIES, A. M., COOL, W. P., BARLOW, D. & MANGHAM, D. C. 2013. Osteoid osteoma and osteoblastoma: novel histological and immunohistochemical observations as evidence for a single entity. *J Clin Pathol*, 66, 768-74.
- BARTHEL, F. P., WEI, W., TANG, M., MARTINEZ-LEDESMA, E., HU, X., AMIN, S. B., AKDEMIR, K. C., SETH, S., SONG, X., WANG, Q., LICHTENBERG, T., HU, J., ZHANG, J., ZHENG, S. & VERHAAK, R. G. W. 2017. Systematic analysis of telomere length and somatic alterations in 31 cancer types. *Nature Genetics*, 49, 349.
- BARUFFI, M. R., VOLPON, J. B., NETO, J. B. & CASARTELLI, C. 2001. Osteoid osteomas with chromosome alterations involving 22q. *Cancer Genetics and Cytogenetics*, 124, 127-131.
- BAUMHOER, D., KOVAC, M., SPERVESLAGE, J., AMELINE, B., STROBL, A. C., KRAUSE, A., TRAUTMANN, M., WARDELMANN, E., NATHRATH, M., HOLLER, S., HARDES, J., GOSHEGER, G., KRIEG, A. H., VIETH, V., TIRABOSCO, R., AMARY, F., FLANAGAN, A. M. & HARTMANN, W. 2019. Activating mutations in the MAP-kinase pathway define non-ossifying fibroma of bone. *J Pathol*, 248, 116-122.
- BEAVER, J. A., JELOVAC, D., BALUKRISHNA, S., COCHRAN, R., CROESSMANN, S., ZABRANSKY, D. J., WONG, H. Y., TORO, P. V., CIDADO, J., BLAIR, B. G., CHU, D., BURNS, T., HIGGINS, M. J., STEARNS, V., JACOBS, L., HABIBI, M., LANGE, J., HURLEY, P. J., LAURING, J., VANDENBERG, D., KESSLER, J., JETER, S., SAMUELS, M. L., MAAR, D., COPE, L., CIMINO-MATHEWS, A., ARGANI, P., WOLFF, A. C. & PARK, B. H. 2014. Detection of cancer DNA in plasma of patients with early-stage breast cancer. *Clin Cancer Res*, 20, 2643-2650.
- BEEBE-DIMMER, J. L., CETIN, K., FRYZEK, J. P., SCHUETZE, S. M. & SCHWARTZ, K. 2009. The epidemiology of malignant giant cell tumors of bone: an analysis of data from the Surveillance, Epidemiology and End Results Program (1975-2004). *Rare Tumors*, 1, e52.
- BEHJATI, S., TARPEY, P. S., HAASE, K., YE, H., YOUNG, M. D., ALEXANDROV, L. B., FARNDON, S. J., COLLORD, G., WEDGE, D. C., MARTINCORENA, I., COOKE, S. L., DAVIES, H., MIFSUD, W., LIDGREN, M., MARTIN, S., LATIMER, C., MADDISON, M., BUTLER, A. P., TEAGUE, J. W., PILLAY, N., SHLIEN, A., MCDERMOTT, U., FUTREAL, P. A., BAUMHOER, D., ZAIKOVA, O., BJERKEHAGEN, B., MYKLEBOST, O., AMARY, M. F., TIRABOSCO, R., VAN LOO, P., STRATTON, M. R., FLANAGAN, A. M. & CAMPBELL, P. J. 2017. Recurrent mutation of IGF signalling genes and distinct patterns of genomic rearrangement in osteosarcoma. *Nat Commun*, 8, 15936.
- BEHJATI, S., TARPEY, P. S., PRESNEAU, N., SCHEIPL, S., PILLAY, N., VAN LOO, P., WEDGE, D. C., COOKE, S. L., GUNDEM, G., DAVIES, H., NIK-ZAINAL,

- S., MARTIN, S., MCLAREN, S., GOODY, V., ROBINSON, B., BUTLER, A., TEAGUE, J. W., HALAI, D., KHATRI, B., MYKLEBOST, O., BAUMHOER, D., JUNDT, G., HAMOUDI, R., TIRABOSCO, R., AMARY, M. F., FUTREAL, P. A., STRATTON, M. R., CAMPBELL, P. J. & FLANAGAN, A. M. 2013. Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. *Nat Genet*, 45, 1479-82.
- BEHJATI, S., TARPEY, P. S., SHELDON, H., MARTINCORENA, I., VAN LOO, P., GUNDEM, G., WEDGE, D. C., RAMAKRISHNA, M., COOKE, S. L., PILLAY, N., VOLLAN, H. K., PAPAEMMANUIL, E., KOSS, H., BUNNEY, T. D., HARDY, C., JOSEPH, O. R., MARTIN, S., MUDIE, L., BUTLER, A., TEAGUE, J. W., PATIL, M., STEERS, G., CAO, Y., GUMBS, C., INGRAM, D., LAZAR, A. J., LITTLE, L., MAHADESHWAR, H., PROTOPOPOV, A., AL SANNAA, G. A., SETH, S., SONG, X., TANG, J., ZHANG, J., RAVI, V., TORRES, K. E., KHATRI, B., HALAI, D., ROXANIS, I., BAUMHOER, D., TIRABOSCO, R., AMARY, M. F., BOSHOFF, C., MCDERMOTT, U., KATAN, M., STRATTON, M. R., FUTREAL, P. A., FLANAGAN, A. M., HARRIS, A. & CAMPBELL, P. J. 2014. Recurrent PTPRB and PLCG1 mutations in angiosarcoma. *Nat Genet*, 46, 376-9.
- BENDER, S., TANG, Y., LINDROTH, A. M., HOVESTADT, V., JONES, D. T., KOOL, M., ZAPATKA, M., NORTHCOTT, P. A., STURM, D., WANG, W., RADLWIMMER, B., HOJFELDT, J. W., TRUFFAUX, N., CASTEL, D., SCHUBERT, S., RYZHOVA, M., SEKER-CIN, H., GRONYCH, J., JOHANN, P. D., STARK, S., MEYER, J., MILDE, T., SCHUHMANN, M., EBINGER, M., MONORANU, C. M., PONNUSWAMI, A., CHEN, S., JONES, C., WITT, O., COLLINS, V. P., VON DEIMLING, A., JABADO, N., PUGET, S., GRILL, J., HELIN, K., KORSHUNOV, A., LICHTER, P., MONJE, M., PLASS, C., CHO, Y. J. & PFISTER, S. M. 2013. Reduced H3K27me3 and DNA hypomethylation are major drivers of gene expression in K27M mutant pediatric high-grade gliomas. *Cancer Cell*, 24, 660-72.
- BERGER, M. F., LAWRENCE, M. S., DEMICHELIS, F., DRIER, Y., CIBULSKIS, K., SIVACHENKO, A. Y., SBONER, A., ESGUEVA, R., PFLUEGER, D., SOUGNEZ, C., ONOFRIO, R., CARTER, S. L., PARK, K., HABEGGER, L., AMBROGIO, L., FENNELL, T., PARKIN, M., SAKSENA, G., VOET, D., RAMOS, A. H., PUGH, T. J., WILKINSON, J., FISHER, S., WINCKLER, W., MAHAN, S., ARDLIE, K., BALDWIN, J., SIMONS, J. W., KITABAYASHI, N. & MACDONALD, T. Y. 2011. The genomic complexity of primary human prostate cancer. *Nature*, 470, 214-20.
- BERTONI, F., BACCHINI, P. & STAALS, E. L. 2003. Malignancy in giant cell tumor of bone. *Cancer*, 97, 2520-9.
- BETTEGOWDA, C., SAUSEN, M., LEARY, R. J., KINDE, I., WANG, Y., AGRAWAL, N., BARTLETT, B. R., WANG, H., LUBER, B., ALANI, R. M., ANTONARAKIS, E. S., AZAD, N. S., BARDELLI, A., BREM, H., CAMERON, J. L., LEE, C. C., FECHER, L. A., GALLIA, G. L., GIBBS, P., LE, D., GIUNTOLI, R. L., GOGGINS, M., HOGARTY, M. D., HOLDHOFF, M., HONG, S. M., JIAO, Y., JUHL, H. H., KIM, J. J., SIRAVEGNA, G., LAHERU, D. A., LAURICELLA, C., LIM, M., LIPSON, E. J., MARIE, S. K., NETTO, G. J., OLINER, K. S., OLIVI, A., OLSSON, L., RIGGINS, G. J., SARTORE-BIANCHI, A., SCHMIDT, K., SHIH, L. M., OBA-SHINJO, S. M., SIENA, S.,

- THEODORESCU, D., TIE, J., HARKINS, T. T., VERONESE, S., WANG, T. L., WEINGART, J. D., WOLFGANG, C. L., WOOD, L. D., XING, D., HRUBAN, R. H., WU, J., ALLEN, P. J., SCHMIDT, C. M., CHOTI, M. A., VELCULESCU, V. E., KINZLER, K. W., VOGELSTEIN, B., PAPADOPOULOS, N. & DIAZ, L. A., JR. 2014. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med*, 6, 224ra24.
- BIAN, S., HOU, Y., ZHOU, X., LI, X., YONG, J., WANG, Y., WANG, W., YAN, J., HU, B., GUO, H., WANG, J., GAO, S., MAO, Y., DONG, J., ZHU, P., XIU, D., YAN, L., WEN, L., QIAO, J., TANG, F. & FU, W. 2018. Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science*, 362, 1060-1063.
- BIELSKI, C. M., ZEHIR, A., PENSON, A. V., DONOGHUE, M. T. A., CHATILA, W., ARMENIA, J., CHANG, M. T., SCHRAM, A. M., JONSSON, P., BANDLAMUDI, C., RAZAVI, P., IYER, G., ROBSON, M. E., STADLER, Z. K., SCHULTZ, N., BASELGA, J., SOLIT, D. B., HYMAN, D. M., BERGER, M. F. & TAYLOR, B. S. 2018. Genome doubling shapes the evolution and prognosis of advanced cancers. *Nature Genetics*, 50, 1189-1195.
- BOS, J. L., FEARON, E. R., HAMILTON, S. R., VERLAAN-DE VRIES, M., VAN BOOM, J. H., VAN DER EB, A. J. & VOGELSTEIN, B. 1987. Prevalence of ras gene mutations in human colorectal cancers. *Nature*, 327, 293-7.
- BOVEE, J. V., CLETON-JANSEN, A. M., WUYTS, W., CAETHOVEN, G., TAMINIAU, A. H., BAKKER, E., VAN HUL, W., CORNELISSE, C. J. & HOGENDOORN, P. C. 1999. EXT-mutation analysis and loss of heterozygosity in sporadic and hereditary osteochondromas and secondary chondrosarcomas. *Am J Hum Genet*, 65, 689-98.
- BOVERI, T. 2008. Concerning the Origin of Malignant Tumours by Theodor Boveri. Translated and annotated by Henry Harris. *Journal of Cell Science*, 121, 1-84.
- BOVERI, T. & BOVERI, M. 1914. [*Zur Frage der Entstehung der malignen Tumoren.*] *The Origin of Malignant Tumors ... Translated by Marcella Boveri.*
- BRYAN, T. M., ENGLEZOU, A., DALLA-POZZA, L., DUNHAM, M. A. & REDDEL, R. R. 1997. Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nature Medicine*, 3, 1271-1274.
- BURSTONE, M. S. 1959. Histochemical demonstration of acid phosphatase activity in osteoclasts. *J Histochem Cytochem*, 7, 39-41.
- CAMERON, D. L., DI STEFANO, L. & PAPPENFUSS, A. T. 2019. Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. *Nature Communications*, 10, 3240.
- CAMPANACCI, M., BALDINI, N., BORIANI, S. & SUDANESE, A. 1987. Giant-cell tumor of bone. *J Bone Joint Surg Am*, 69, 106-14.
- CAMPBELL, B. B., LIGHT, N., FABRIZIO, D., ZATZMAN, M., FULIGNI, F., DE BORJA, R., DAVIDSON, S., EDWARDS, M., ELVIN, J. A., HODEL, K. P., ZAHURANCIK, W. J., SUO, Z., LIPMAN, T., WIMMER, K., KRATZ, C. P., BOWERS, D. C., LAETSCH, T. W., DUNN, G. P., JOHANNIS, T. M., GRIMMER, M. R., SMIRNOV, I. V., LAROUCHE, V., SAMUEL, D., BRONSEMA, A., OSBORN, M., STEARNS, D., RAMAN, P., COLE, K. A., STORM, P. B., YALON, M., OPOCHER, E., MASON, G., THOMAS, G. A., SABEL, M., GEORGE, B., ZIEGLER, D. S., LINDHORST, S., ISSAI, V. M., CONSTANTINI, S., TOLEDANO, H., ELHASID, R., FARAH, R., DVIR, R.,

- DIRKS, P., HUANG, A., GALATI, M. A., CHUNG, J., RAMASWAMY, V., IRWIN, M. S., ARONSON, M., DURNO, C., TAYLOR, M. D., RECHAVI, G., MARIS, J. M., BOUFFET, E., HAWKINS, C., COSTELLO, J. F., MEYN, M. S., PURSELL, Z. F., MALKIN, D., TABORI, U. & SHLIEN, A. 2017. Comprehensive Analysis of Hypermutation in Human Cancer. *Cell*, 171, 1042-1056.e10.
- CAMPBELL, P. J., PLEASANCE, E. D., STEPHENS, P. J., DICKS, E., RANCE, R., GOODHEAD, I., FOLLOWS, G. A., GREEN, A. R., FUTREAL, P. A. & STRATTON, M. R. 2008. Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proc Natl Acad Sci U S A*, 105, 13081-6.
- CAMPBELL, P. J., YACHIDA, S., MUDIE, L. J., STEPHENS, P. J., PLEASANCE, E. D., STEBBINGS, L. A., MORSBERGER, L. A., LATIMER, C., MCLAREN, S., LIN, M.-L., MCBRIDE, D. J., VARELA, I., NIK-ZAINAL, S. A., LEROY, C., JIA, M., MENZIES, A., BUTLER, A. P., TEAGUE, J. W., GRIFFIN, C. A., BURTON, J., SWERDLOW, H., QUAIL, M. A., STRATTON, M. R., IACOBUZIO-DONAHUE, C. & FUTREAL, P. A. 2010. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature*, 467, 1109-1113.
- CARAVAGNA, G., GIARRATANO, Y., RAMAZZOTTI, D., TOMLINSON, I., GRAHAM, T. A., SANGUINETTI, G. & SOTTORIVA, A. 2018. Detecting repeated cancer evolution from multi-region tumor sequencing data. *Nature Methods*, 15, 707-714.
- CARNEY, J. A., BOCCON-GIBOD, L., JARKA, D. E., TANAKA, Y., SWEE, R. G., UNNI, K. K. & STRATAKIS, C. A. 2001. Osteochondromyxoma of bone: a congenital tumor associated with lentiginos and other unusual disorders. *Am J Surg Pathol*, 25, 164-76.
- CARTER, S. L., CIBULSKIS, K., HELMAN, E., MCKENNA, A., SHEN, H., ZACK, T., LAIRD, P. W., ONOFRIO, R. C., WINCKLER, W., WEIR, B. A., BEROUKHIM, R., PELLMAN, D., LEVINE, D. A., LANDER, E. S., MEYERSON, M. & GETZ, G. 2012. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol*, 30, 413-21.
- CASULA, S. & BIANCO, A. C. 2012. Thyroid hormone deiodinases and cancer. *Frontiers in endocrinology*, 3, 74-74.
- CELESTINO, R., SIGSTAD, E., LOVF, M., THOMASSEN, G. O., GROHOLT, K. K., JORGENSEN, L. H., BERNER, A., CASTRO, P., LOTHE, R. A., BJORO, T., SOBRINHO-SIMÕES, M., SOARES, P. & SKOTHEIM, R. I. 2012. Survey of 548 oncogenic fusion transcripts in thyroid tumors supports the importance of the already established thyroid fusions genes. *Genes Chromosomes Cancer*, 51, 1154-64.
- CESARE, A. J. & REDDEL, R. R. 2010. Alternative lengthening of telomeres: models, mechanisms and implications. *Nature Reviews Genetics*, 11, 319.
- CHAKARUN, C. J., FORRESTER, D. M., GOTTSEGEN, C. J., PATEL, D. B., WHITE, E. A. & MATCUK, G. R. 2013. Giant Cell Tumor of Bone: Review, Mimics, and New Developments in Treatment. *RadioGraphics*, 33, 197-211.
- CHAN, C. M., ADLER, Z., REITH, J. D. & GIBBS, C. P., JR. 2015. Risk factors for pulmonary metastases from giant cell tumor of bone. *J Bone Joint Surg Am*, 97, 420-8.

- CHAWLA, S., HENSHAW, R., SEEGER, L., CHOY, E., BLAY, J. Y., FERRARI, S., KROEP, J., GRIMER, R., REICHARDT, P., RUTKOWSKI, P., SCHUETZE, S., SKUBITZ, K., STADDON, A., THOMAS, D., QIAN, Y. & JACOBS, I. 2013. Safety and efficacy of denosumab for adults and skeletally mature adolescents with giant cell tumour of bone: interim analysis of an open-label, parallel-group, phase 2 study. *Lancet Oncol*, 14, 901-8.
- CHEN, C. Y., CHEN, T. M. & SHYU, A. B. 1994. Interplay of two functionally and structurally distinct domains of the c-fos AU-rich element specifies its mRNA-destabilizing function. *Mol Cell Biol*, 14, 416-26.
- CHEN, K. S., KWON, W. S., KIM, J., HEO, S. J., KIM, H. S., KIM, H. K., KIM, S. H., LEE, W. S., CHUNG, H. C., RHA, S. Y. & HWANG, T. H. 2016. A novel TP53-KPNA3 translocation defines a de novo treatment-resistant clone in osteosarcoma. *Cold Spring Harb Mol Case Stud*, 2, a000992.
- CHOU, A., FRASER, S., TOON, C. W., CLARKSON, A., SIOSON, L., FARZIN, M., CUSSIGH, C., ANISS, A., O'NEILL, C., WATSON, N., CLIFTON-BLIGH, R. J., LEAROYD, D. L., ROBINSON, B. G., SELINGER, C. I., DELBRIDGE, L. W., SIDHU, S. B., O'TOOLE, S. A., SYWAK, M. & GILL, A. J. 2015. A detailed clinicopathologic study of ALK-translocated papillary thyroid carcinoma. *Am J Surg Pathol*, 39, 652-9.
- CHRISTIE, E. L., FEREDAY, S., DOIG, K., PATTNAIK, S., DAWSON, S. J. & BOWTELL, D. D. L. 2017. Reversion of BRCA1/2 Germline Mutations Detected in Circulating Tumor DNA From Patients With High-Grade Serous Ovarian Cancer. *J Clin Oncol*, 35, 1274-1280.
- CIAMPI, R., KNAUF, J. A., KERLER, R., GANDHI, M., ZHU, Z., NIKIFOROVA, M. N., RABES, H. M., FAGIN, J. A. & NIKIFOROV, Y. E. 2005. Oncogenic AKAP9-BRAF fusion is a novel mechanism of MAPK pathway activation in thyroid cancer. *J Clin Invest*, 115, 94-101.
- CIBULSKIS, K., LAWRENCE, M. S., CARTER, S. L., SIVACHENKO, A., JAFFE, D., SOUGNEZ, C., GABRIEL, S., MEYERSON, M., LANDER, E. S. & GETZ, G. 2013. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*, 31, 213-9.
- CMERO, M., ONG, C. S., YUAN, K., SCHRÖDER, J., MO, K., CORCORAN, N. M., PAPPENFUSS, A. T., HOVENS, C. M., MARKOWETZ, F. & MACINTYRE, G. 2017. SVclone: inferring structural variant cancer cell fraction. *bioRxiv*.
- COHEN, Y., XING, M., MAMBO, E., GUO, Z., WU, G., TRINK, B., BELLER, U., WESTRA, W. H., LADENSON, P. W. & SIDRANSKY, D. 2003. BRAF Mutation in Papillary Thyroid Carcinoma. *JNCI: Journal of the National Cancer Institute*, 95, 625-627.
- COHEN-AUBART, F., EMILE, J. F., CARRAT, F., HELIAS-RODZEWICZ, Z., TALY, V., CHARLOTTE, F., CLUZEL, P., DONADIEU, J., IDBAIH, A., BARETE, S., AMOURA, Z. & HAROCHE, J. 2018. Phenotypes and survival in Erdheim-Chester disease: Results from a 165-patient cohort. *Am J Hematol*, 93, E114-e117.
- COONS, S. W., JOHNSON, P. C. & SHAPIRO, J. R. 1995. CYTOGENETIC AND FLOW-CYTOMETRY DNA ANALYSIS OF REGIONAL HETEROGENEITY IN A LOW-GRADE HUMAN GLIOMA. *Cancer Research*, 55, 1569-1577.
- CORTÉS-CIRIANO, I., LEE, J.-K., XI, R., JAIN, D., JUNG, Y. L., YANG, L., GORDENIN, D., KLIMCZAK, L. J., ZHANG, C.-Z., PELLMAN, D. S. & PARK,

- P. J. 2018. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *bioRxiv*, 333617.
- CRISTOFANILLI, M., TURNER, N. C., BONDARENKO, I., RO, J., IM, S. A., MASUDA, N., COLLEONI, M., DEMICHELE, A., LOI, S., VERMA, S., IWATA, H., HARBECK, N., ZHANG, K., THEALL, K. P., JIANG, Y., BARTLETT, C. H., KOEHLER, M. & SLAMON, D. 2016. Fulvestrant plus palbociclib versus fulvestrant plus placebo for treatment of hormone-receptor-positive, HER2-negative metastatic breast cancer that progressed on previous endocrine therapy (PALOMA-3): final analysis of the multicentre, double-blind, phase 3 randomised controlled trial. *Lancet Oncol*, 17, 425-439.
- CUN, Y., YANG, T.-P., ACHTER, V., LANG, U. & PEIFER, M. 2018. Copy-number analysis and inference of subclonal populations in cancer genomes using Sclust. *Nature Protocols*, 13, 1488.
- CZERNIAK, B. & DORFMAN, H. D. 2016. *Dorfman and Czerniak's bone tumors*, Philadelphia, Elsevier.
- DAGOGO-JACK, I. & SHAW, A. T. 2018. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol*, 15, 81-94.
- DAHLIN, D. C., CUPPS, R. E. & JOHNSON JR, E. W. 1970. Giant-cell tumor: A study of 195 cases. *Cancer*, 25, 1061-1070.
- DALGLIESH, G. L., FURGE, K., GREENMAN, C., CHEN, L., BIGNELL, G., BUTLER, A., DAVIES, H., EDKINS, S., HARDY, C., LATIMER, C., TEAGUE, J., ANDREWS, J., BARTHORPE, S., BEARE, D., BUCK, G., CAMPBELL, P. J., FORBES, S., JIA, M., JONES, D., KNOTT, H., KOK, C. Y., LAU, K. W., LEROY, C., LIN, M. L., MCBRIDE, D. J., MADDISON, M., MAGUIRE, S., MCLAY, K., MENZIES, A., MIRONENKO, T., MULDERRIG, L., MUDIE, L., O'MEARA, S., PLEASANCE, E., RAJASINGHAM, A., SHEPHERD, R., SMITH, R., STEBBINGS, L., STEPHENS, P., TANG, G., TARPEY, P. S., TURRELL, K., DYKEMA, K. J., KHOO, S. K., PETILLO, D., WONDERGEM, B., ANEMA, J., KAHNOSKI, R. J., TEH, B. T., STRATTON, M. R. & FUTREAL, P. A. 2010. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature*, 463, 360-3.
- DAVOLI, T., UNO, H., WOOTEN, E. C. & ELLEDGE, S. J. 2017. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science*, 355, eaaf8399.
- DAWSON, S. J., TSUI, D. W., MURTAZA, M., BIGGS, H., RUEDA, O. M., CHIN, S. F., DUNNING, M. J., GALE, D., FORSHEW, T., MAHLER-ARAUIJO, B., RAJAN, S., HUMPHRAY, S., BECQ, J., HALSALL, D., WALLIS, M., BENTLEY, D., CALDAS, C. & ROSENFELD, N. 2013. Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med*, 368, 1199-209.
- DE BRUIN, E. C., MCGRANAHAN, N., MITTER, R., SALM, M., WEDGE, D. C., YATES, L., JAMAL-HANJANI, M., SHAFI, S., MURUGAESU, N., ROWAN, A. J., GRONROOS, E., MUHAMMAD, M. A., HORSWELL, S., GERLINGER, M., VARELA, I., JONES, D., MARSHALL, J., VOET, T., VAN LOO, P., RASSL, D. M., RINTOUL, R. C., JANES, S. M., LEE, S. M., FORSTER, M., AHMAD, T., LAWRENCE, D., FALZON, M., CAPITANIO, A., HARKINS, T. T., LEE, C. C., TOM, W., TEEFE, E., CHEN, S. C., BEGUM, S., RABINOWITZ, A., PHILLIMORE, B., SPENCER-DENE, B., STAMP, G., SZALLASI, Z., MATTHEWS, N., STEWART, A., CAMPBELL, P. & SWANTON, C. 2014.

Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science*, 346, 251-6.

- DE MATTOS-ARRUDA, L., MAYOR, R., NG, C. K., WEIGELT, B., MARTINEZ-RICARTE, F., TORREJON, D., OLIVEIRA, M., ARIAS, A., RAVENTOS, C., TANG, J., GUERINI-ROCCO, E., MARTINEZ-SAEZ, E., LOIS, S., MARIN, O., DE LA CRUZ, X., PISCUOGLIO, S., TOWERS, R., VIVANCOS, A., PEG, V., RAMON Y CAJAL, S., CARLES, J., RODON, J., GONZALEZ-CAO, M., TABERNERO, J., FELIP, E., SAHUQUILLO, J., BERGER, M. F., CORTES, J., REIS-FILHO, J. S. & SEOANE, J. 2015. Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. *Nat Commun*, 6, 8839.
- DEAN, J. L., THANGAVEL, C., MCCLENDON, A. K., REED, C. A. & KNUDSEN, E. S. 2010. Therapeutic CDK4/6 inhibition in breast cancer: key mechanisms of response and failure. *Oncogene*, 29, 4018-32.
- DEGREGORI, J. 2011. Evolved tumor suppression: why are we so good at not getting cancer? *Cancer Res*, 71, 3739-44.
- DEMEULEMEESTER, J., KUMAR, P., MOLLER, E. K., NORD, S., WEDGE, D. C., PETERSON, A., MATHIESEN, R. R., FJELLDAL, R., ZAMANI ESTEKI, M., THEUNIS, K., FERNANDEZ GALLARDO, E., GRUNDSTAD, A. J., BORGES, E., BAUMBUSCH, L. O., BORRESEN-DALE, A. L., WHITE, K. P., KRISTENSEN, V. N., VAN LOO, P., VOET, T. & NAUME, B. 2016. Tracing the origin of disseminated tumor cells in breast cancer using single-cell sequencing. *Genome Biol*, 17, 250.
- DENARDO, D. G., ANDREU, P. & COUSSENS, L. M. 2010. Interactions between lymphocytes and myeloid cells regulate pro- versus anti-tumor immunity. *Cancer Metastasis Rev*, 29, 309-16.
- DENTRO, S. C., LESHCHINER, I., HAASE, K., TARABICHI, M., WINTERSINGER, J., DESHWAR, A. G., YU, K., RUBANOVA, Y., MACINTYRE, G., VAZQUEZ-GARCIA, I., KLEINHEINZ, K., LIVITZ, D. G., MALIKIC, S., DONMEZ, N., SENGUPTA, S., DEMEULEMEESTER, J., ANUR, P., JOLLY, C., CMERO, M., ROSEBROCK, D., SCHUMACHER, S., FAN, Y., FITTALL, M., DREWS, R. M., YAO, X., LEE, J., SCHLESNER, M., ZHU, H., ADAMS, D. J., GETZ, G., BOUTROS, P. C., IMIELINSKI, M., BEROUKHIM, R., SAHINALP, S. C., JI, Y., PEIFER, M., MARTINCORENA, I., MARKOWETZ, F., MUSTONEN, V., YUAN, K., GERSTUNG, M., SPELLMAN, P. T., WANG, W., MORRIS, Q. D., WEDGE, D. C. & VAN LOO, P. 2018. Portraits of genetic intra-tumour heterogeneity and subclonal selection across cancer types. *bioRxiv*, 312041.
- DENTRO, S. C., WEDGE, D. C. & VAN LOO, P. 2017. Principles of Reconstructing the Subclonal Architecture of Cancers. *Cold Spring Harb Perspect Med*, 7.
- DESHMUKH, S., KHAZAEI, S., MARCHIONE, D., HARUTYUNYAN, A., GARCIA, B. & JABADO, N. 2018. Abstract A39: Characterizing the epigenetic effects of the histone 3.3 G34W mutation in giant cell tumors of bone. *Cancer Research*, 78, A39-A39.
- DESHWAR, A. G., VEMBU, S., YUNG, C. K., JANG, G. H., STEIN, L. & MORRIS, Q. 2015. PhyloWGS: Reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. *Genome Biology*, 16, 35.

- DEY, S. S., KESTER, L., SPANJAARD, B., BIENKO, M. & VAN OUDENAARDEN, A. 2015. Integrated genome and transcriptome sequencing of the same cell. *Nat Biotechnol*, 33, 285-9.
- DEYRUP, A. T. & MONTAG, A. G. 2007. Epithelioid and Epithelial Neoplasms of Bone. *Archives of Pathology & Laboratory Medicine*, 131, 205-216.
- DING, Z., MANGINO, M., AVIV, A., CONSORTIUM, U. K., SPECTOR, T. & DURBIN, R. 2014. Estimating telomere length from whole genome sequence data. *Nucleic Acids Research*, 42, e75-e75.
- DOBIN, A., DAVIS, C. A., SCHLESINGER, F., DRENKOW, J., ZALESKI, C., JHA, S., BATUT, P., CHAISSON, M. & GINGERAS, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29, 15-21.
- DONMEZ, N., MALIKIC, S., WYATT, A. W., GLEAVE, M. E., COLLINS, C. C. & SAHINALP, S. C. 2017. Clonality Inference from Single Tumor Samples Using Low-Coverage Sequence Data. *Journal of Computational Biology*, 24, 515-523.
- DU, J., XU, L., CUI, Y., LIU, Z., SU, Y. & LI, G. 2019. Benign notochordal cell tumour: clinicopathology and molecular profiling of 13 cases. *J Clin Pathol*, 72, 66-74.
- DZAMBA, M., RAMANI, A. K., BUCZKOWICZ, P., JIANG, Y., YU, M., HAWKINS, C. & BRUDNO, M. 2017. Identification of complex genomic rearrangements in cancers using CouGaR. *Genome Res*, 27, 107-117.
- EIREW, P., STEIF, A., KHATTRA, J., HA, G., YAP, D., FARAHANI, H., GELMON, K., CHIA, S., MAR, C., WAN, A., LAKS, E., BIELE, J., SHUMANSKY, K., ROSNER, J., MCPHERSON, A., NIELSEN, C., ROTH, A. J., LEFEBVRE, C., BASHASHATI, A., DE SOUZA, C., SIU, C., ANIBA, R., BRIMHALL, J., OLOUMI, A., OSAKO, T., BRUNA, A., SANDOVAL, J. L., ALGARA, T., GREENWOOD, W., LEUNG, K., CHENG, H., XUE, H., WANG, Y., LIN, D., MUNGALL, A. J., MOORE, R., ZHAO, Y., LORETTE, J., NGUYEN, L., HUNTSMAN, D., EAVES, C. J., HANSEN, C., MARRA, M. A., CALDAS, C., SHAH, S. P. & APARICIO, S. 2015. Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature*, 518, 422-6.
- ENDESFELDER, D., BURRELL, R., KANU, N., MCGRANAHAN, N., HOWELL, M., PARKER, P. J., DOWNWARD, J., SWANTON, C. & KSCHISCHO, M. 2014. Chromosomal instability selects gene copy-number variants encoding core regulators of proliferation in ER+ breast cancer. *Cancer Res*, 74, 4853-4863.
- ENRIQUEZ-NAVAS, P. M., KAM, Y., DAS, T., HASSAN, S., SILVA, A., FOROUTAN, P., RUIZ, E., MARTINEZ, G., MINTON, S., GILLIES, R. J. & GATENBY, R. A. 2016. Exploiting evolutionary principles to prolong tumor control in preclinical models of breast cancer. *Science Translational Medicine*, 8, 327ra24.
- ERRANI, C., ZHANG, L., SUNG, Y. S., HAJDU, M., SINGER, S., MAKI, R. G., HEALEY, J. H. & ANTONESCU, C. R. 2011. A novel WWTR1-CAMTA1 gene fusion is a consistent abnormality in epithelioid hemangioendothelioma of different anatomic sites. *Genes, Chromosomes and Cancer*, 50, 644-653.
- ESSERMAN, L. J., THOMPSON, I. M., JR. & REID, B. 2013. Overdiagnosis and overtreatment in cancer: an opportunity for improvement. *JAMA*, 310, 797-8.
- ESTELLER, M., LEVINE, R., BAYLIN, S. B., ELLENSON, L. H. & HERMAN, J. G. 1998. MLH1 promoter hypermethylation is associated with the microsatellite instability phenotype in sporadic endometrial carcinomas. *Oncogene*, 17, 2413.

- EWING, A. D., HOULAHAN, K. E., HU, Y., ELLROTT, K., CALOIAN, C., YAMAGUCHI, T. N., BARE, J. C., P'NG, C., WAGGOTT, D., SABELNYKOVA, V. Y., PARTICIPANTS, I.-T. D. S. M. C. C., XI, L., DEWAL, N., FAN, Y., WANG, W., WHEELER, D., WILM, A., TING, G. H., LI, C., BERTRAND, D., NAGARAJAN, N., CHEN, Q.-R., HSU, C.-H., HU, Y., YAN, C., KIBBE, W., MEERZAMAN, D., CIBULSKIS, K., ROSENBERG, M., BERGELSON, L., KIEZUN, A., RADENBAUGH, A., SERTIER, A.-S., FERRARI, A., TONTON, L., BHUTANI, K., HANSEN, N. F., WANG, D., SONG, L., LAI, Z., LIAO, Y., SHI, W., CARBONELL-CABALLERO, J., DOPAZO, J., LAU, C. C. K., GUINNEY, J., KELLEN, M. R., NORMAN, T. C., HAUSSLER, D., FRIEND, S. H., STOLOVITZKY, G., MARGOLIN, A. A., STUART, J. M. & BOUTROS, P. C. 2015. Combining tumor genome simulation with crowdsourcing to benchmark somatic single-nucleotide-variant detection. *Nature Methods*, 12, 623.
- FAN, Y., XI, L., HUGHES, D. S., ZHANG, J., ZHANG, J., FUTREAL, P. A., WHEELER, D. A. & WANG, W. 2016. MuSE: accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling from sequencing data. *Genome Biol*, 17, 178.
- FARABEGOLI, F., SANTINI, D., CECCARELLI, C., TAFFURELLI, M., MARRANO, D. & BALDINI, N. 2001. Clone heterogeneity in diploid and aneuploid breast carcinomas as detected by FISH. *Cytometry*, 46, 50-56.
- FEINBERG, A. P., OHLSSON, R. & HENIKOFF, S. 2006. The epigenetic progenitor origin of human cancer. *Nat Rev Genet*, 7, 21-33.
- FELLENBERG, J., SÄHR, H., MANCARELLA, D., PLASS, C., LINDROTH, A., WESTHAUSER, F., LEHNER, B. & EWERBECK, V. 2019. Knock-down of oncohistone H3F3A-G34W counteracts the neoplastic phenotype of giant cell tumor of bone derived stromal cells. *Cancer letters*, 448, 61-69.
- FERRARA, P., ANDERMARCHER, E., BOSSIS, G., ACQUAVIVA, C., BROCKLY, F., JARIEL-ENCONTRE, I. & PIECHACZYK, M. 2003. The structural determinants responsible for c-Fos protein proteasomal degradation differ according to the conditions of expression. *Oncogene*, 22, 1461-74.
- FINKEL, M. P., BISKIS, B. O. & JINKINS, P. B. 1966. Virus induction of osteosarcomas in mice. *Science*, 151, 698-701.
- FINN, S. P., SMYTH, P., CAHILL, S., STRECK, C., O'REGAN, E. M., FLAVIN, R., SHERLOCK, J., HOWELLS, D., HENFREY, R., CULLEN, M., TONER, M., TIMON, C., O'LEARY, J. J. & SHEILS, O. M. 2007. Expression microarray analysis of papillary thyroid carcinoma and benign thyroid tissue: emphasis on the follicular variant and potential markers of malignancy. *Virchows Archiv : an international journal of pathology*, 450, 249-260.
- FISCHER, A., VAZQUEZ-GARCIA, I., ILLINGWORTH, C. J. & MUSTONEN, V. 2014a. High-definition reconstruction of clonal composition in cancer. *Cell Rep*, 7, 1740-52.
- FISCHER, A., VÁZQUEZ-GARCÍA, I., ILLINGWORTH, C. J. R. & MUSTONEN, V. 2014b. High-definition reconstruction of clonal composition in cancer. *Cell reports*, 7, 1740-1752.
- FITTALL, M. W. & VAN LOO, P. 2019. Translating insights into tumor evolution to clinical practice: promises and challenges. *Genome Medicine*, 11, 20.

- FITZGERALD, P. J. 1986. HOMOGENEITY AND HETEROGENEITY IN PANCREAS CANCER - PRESENCE OF PREDOMINANT AND MINOR MORPHOLOGICAL TYPES AND IMPLICATIONS. *International Journal of Pancreatology*, 1, 91-94.
- FLAVAHAN, W. A., GASKELL, E. & BERNSTEIN, B. E. 2017. Epigenetic plasticity and the hallmarks of cancer. *Science*, 357, eaal2380.
- FLETCHER, C. D. M., WORLD HEALTH ORGANIZATION. & INTERNATIONAL AGENCY FOR RESEARCH ON CANCER. 2013. *WHO classification of tumours of soft tissue and bone*, Lyon, IARC Press.
- FRANKLIN, R. E. & GOSLING, R. G. 1953. Molecular Configuration in Sodium Thymonucleate. *Nature*, 171, 740-741.
- GALON, J., COSTES, A., SANCHEZ-CABO, F., KIRILOVSKY, A., MLECNIK, B., LAGORCE-PAGÈS, C., TOSOLINI, M., CAMUS, M., BERGER, A., WIND, P., ZINZINDOHOUE, F., BRUNÉVAL, P., CUGNENC, P.-H., TRAJANOSKI, Z., FRIDMAN, W.-H. & PAGÈS, F. 2006. Type, Density, and Location of Immune Cells Within Human Colorectal Tumors Predict Clinical Outcome. *Science*, 313, 1960-1964.
- GAMBAROTTI, M., DEI TOS, A. P., VANEL, D., PICCI, P., GIBERTONI, D., KLEIN, M. J. & RIGHI, A. 2019. Osteoblastoma-like osteosarcoma: high-grade or low-grade osteosarcoma? *Histopathology*, 74, 494-503.
- GAO, R., DAVIS, A., MCDONALD, T. O., SEI, E., SHI, X., WANG, Y., TSAI, P.-C., CASASANT, A., WATERS, J., ZHANG, H., MERIC-BERNSTAM, F., MICHOR, F. & NAVIN, N. E. 2016. Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat Genet*, 48, 1119-1130.
- GARSED, D. W., MARSHALL, O. J., CORBIN, V. D., HSU, A., DI STEFANO, L., SCHRODER, J., LI, J., FENG, Z. P., KIM, B. W., KOWARSKY, M., LANSDELL, B., BROOKWELL, R., MYKLEBOST, O., MEZA-ZEPEDA, L., HOLLOWAY, A. J., PEDEUTOUR, F., CHOO, K. H., DAMORE, M. A., DEANS, A. J., PAPENFUSS, A. T. & THOMAS, D. M. 2014. The architecture and evolution of cancer neochromosomes. *Cancer Cell*, 26, 653-67.
- GASTON, C. L., GRIMER, R. J., PARRY, M., STACCHIOTTI, S., DEI TOS, A. P., GELDERBLUM, H., FERRARI, S., BALDI, G. G., JONES, R. L., CHAWLA, S., CASALI, P., LECESNE, A., BLAY, J. Y., DIJKSTRA, S. P., THOMAS, D. M. & RUTKOWSKI, P. 2016. Current status and unanswered questions on the use of Denosumab in giant cell tumor of bone. *Clin Sarcoma Res*, 6, 15.
- GATENBY, R. A., SILVA, A. S., GILLIES, R. J. & FRIEDEN, B. R. 2009. Adaptive Therapy. *Cancer Research*, 69, 4894.
- GENOVESE, G., KAHLER, A. K., HANDSAKER, R. E., LINDBERG, J., ROSE, S. A., BAKHOUM, S. F., CHAMBERT, K., MICK, E., NEALE, B. M., FROMER, M., PURCELL, S. M., SVANTESSON, O., LANDEN, M., HOGLUND, M., LEHMANN, S., GABRIEL, S. B., MORAN, J. L., LANDER, E. S., SULLIVAN, P. F., SKLAR, P., GRONBERG, H., HULTMAN, C. M. & MCCARROLL, S. A. 2014. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med*, 371, 2477-87.
- GEORGE, J., LIM, J. S., JANG, S. J., CUN, Y., OZRETIC, L., KONG, G., LEENDERS, F., LU, X., FERNANDEZ-CUESTA, L., BOSCO, G., MULLER, C., DAHMEN, I., JAHCHAN, N. S., PARK, K. S., YANG, D., KARNEZIS, A. N., VAKA, D., TORRES, A., WANG, M. S., KORBEL, J. O., MENON, R., CHUN, S. M., KIM,

D., WILKERSON, M., HAYES, N., ENGELMANN, D., PUTZER, B., BOS, M., MICHELS, S., VLASIC, I., SEIDEL, D., PINTHER, B., SCHAUB, P., BECKER, C., ALTMULLER, J., YOKOTA, J., KOHNO, T., IWAKAWA, R., TSUTA, K., NOGUCHI, M., MULEY, T., HOFFMANN, H., SCHNABEL, P. A., PETERSEN, I., CHEN, Y., SOLTERMANN, A., TISCHLER, V., CHOI, C. M., KIM, Y. H., MASSION, P. P., ZOU, Y., JOVANOVIC, D., KONTIC, M., WRIGHT, G. M., RUSSELL, P. A., SOLOMON, B., KOCH, I., LINDNER, M., MUSCARELLA, L. A., LA TORRE, A., FIELD, J. K., JAKOPOVIC, M., KNEZEVIC, J., CASTANOS-VELEZ, E., ROZ, L., PASTORINO, U., BRUSTUGUN, O. T., LUND-IVERSEN, M., THUNNISSEN, E., KOHLER, J., SCHULER, M., BOTLING, J., SANDELIN, M., SANCHEZ-CESPEDES, M., SALVESEN, H. B., ACHTER, V., LANG, U., BOGUS, M., SCHNEIDER, P. M., ZANDER, T., ANSEN, S., HALLEK, M., WOLF, J., VINGRON, M., YATABE, Y., TRAVIS, W. D., NURNBERG, P., REINHARDT, C., PERNER, S., HEUKAMP, L., BUTTNER, R., HAAS, S. A., BRAMBILLA, E., PEIFER, M., SAGE, J. & THOMAS, R. K. 2015. Comprehensive genomic profiles of small cell lung cancer. *Nature*, 524, 47-53.

GERHAUSER, C., FAVERO, F., RISCH, T., SIMON, R., FEUERBACH, L., ASSENOV, Y., HECKMANN, D., SIDIROPOULOS, N., WASZAK, S. M., HÜBSCHMANN, D., URBANUCCI, A., GIRMA, E. G., KURYSHEV, V., KLIMCZAK, L. J., SAINI, N., STÜTZ, A. M., WEICHENHAN, D., BÖTTCHER, L.-M., TOTH, R., HENDRIKSEN, J. D., KOOP, C., LUTSIK, P., MATZK, S., WARNATZ, H.-J., AMSTISLAVSKIY, V., FEUERSTEIN, C., RAEDER, B., BOGATYROVA, O., SCHMITZ, E.-M., HUBE-MAGG, C., KLUTH, M., HULAND, H., GRAEFEN, M., LAWERENZ, C., HENRY, G. H., YAMAGUCHI, T. N., MALEWSKA, A., MEINERS, J., SCHILLING, D., REISINGER, E., EILS, R., SCHLESNER, M., STRAND, D. W., BRISTOW, R. G., BOUTROS, P. C., VON KALLE, C., GORDENIN, D., SÜLTSMANN, H., BRORS, B., SAUTER, G., PLASS, C., YASPO, M.-L., KORBEL, J. O., SCHLOMM, T. & WEISCHENFELDT, J. 2018. Molecular Evolution of Early-Onset Prostate Cancer Identifies Molecular Risk Markers and Clinical Trajectories. *Cancer Cell*, 34, 996-1011.e8.

GERLINGER, M., ROWAN, A. J., HORSWELL, S., LARKIN, J., ENDESFELDER, D., GRONROOS, E., MARTINEZ, P., MATTHEWS, N., STEWART, A., TARPEY, P., VARELA, I., PHILLIMORE, B., BEGUM, S., MCDONALD, N. Q., BUTLER, A., JONES, D., RAINE, K., LATIMER, C., SANTOS, C. R., NOHADANI, M., EKLUND, A. C., SPENCER-DENE, B., CLARK, G., PICKERING, L., STAMP, G., GORE, M., SZALLASI, Z., DOWNWARD, J., FUTREAL, P. A. & SWANTON, C. 2012. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med*, 366, 883-92.

GERSTUNG, M., JOLLY, C., LESHCHINER, I., DENTRO, S. C., GONZALEZ, S., MITCHELL, T. J., RUBANOVA, Y., ANUR, P., ROSEBROCK, D., YU, K., TARABICHI, M., DESHWAR, A., WINTERSINGER, J., KLEINHEINZ, K., VAZQUEZ-GARCIA, I., HAASE, K., SENGUPTA, S., MACINTYRE, G., MALIKIC, S., DONMEZ, N., LIVITZ, D. G., CMERO, M., DEMEULEMEESTER, J., SCHUMACHER, S., FAN, Y., YAO, X., LEE, J., SCHLESNER, M., BOUTROS, P. C., BOWTELL, D. D., ZHU, H., GETZ, G.,

- IMIELINSKI, M., BEROUKHIM, R., SAHINALP, S. C., JI, Y., PEIFER, M., MARKOWETZ, F., MUSTONEN, V., YUAN, K., WANG, W., MORRIS, Q. D., SPELLMAN, P. T., WEDGE, D. C., VAN LOO, P., & 2018. The evolutionary history of 2,658 cancers. *bioRxiv*.
- GESSI, M., GIELEN, G. H., HAMMES, J., DORNER, E., MUHLEN, A. Z., WAHA, A. & PIETSCH, T. 2013. H3.3 G34R mutations in pediatric primitive neuroectodermal tumors of central nervous system (CNS-PNET) and pediatric glioblastomas: possible diagnostic and therapeutic implications? *J Neurooncol*, 112, 67-72.
- GHAVI-HELM, Y., JANKOWSKI, A., MEIERS, S., VIALES, R. R., KORBEL, J. O. & FURLONG, E. E. M. 2019. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nature Genetics*, 51, 1272-1282.
- GIACONA, M. B., RUBEN, G. C., ICZKOWSKI, K. A., ROOS, T. B., PORTER, D. M. & SORENSON, G. D. 1998. Cell-free DNA in human blood plasma: length measurements in patients with pancreatic cancer and healthy controls. *Pancreas*, 17, 89-97.
- GIANNICO, G., HOLT, G. E., HOMLAR, K. C., JOHNSON, J., PINNT, J. & BRIDGE, J. A. 2009. Osteoblastoma characterized by a three-way translocation: report of a case and review of the literature. *Cancer Genet Cytogenet*, 195, 168-71.
- GIBSON, C. J. & STEENSMA, D. P. 2018. New Insights from Studies of Clonal Hematopoiesis. *Clinical Cancer Research*, clincanres.3044.
- GOVIND, S. K., ZIA, A., HENNINGS-YEOMANS, P. H., WATSON, J. D., FRASER, M., ANGHEL, C., WYATT, A. W., VAN DER KWAST, T., COLLINS, C. C., MCPHERSON, J. D., BRISTOW, R. G. & BOUTROS, P. C. 2014. Shatterproof: operational detection and quantification of chromothripsis. *BMC Bioinformatics*, 15.
- GRIVENNIKOV, S. I., GRETEN, F. R. & KARIN, M. 2010. Immunity, inflammation, and cancer. *Cell*, 140, 883-99.
- GROSSET, C., CHEN, C. Y., XU, N., SONENBERG, N., JACQUEMIN-SABLON, H. & SHYU, A. B. 2000. A mechanism for translationally coupled mRNA turnover: interaction between the poly(A) tail and a c-fos RNA coding determinant via a protein complex. *Cell*, 103, 29-40.
- GUNDEM, G., VAN LOO, P., KREMEYER, B., ALEXANDROV, L. B., TUBIO, J. M., PAPAEMMANUIL, E., BREWER, D. S., KALLIO, H. M., HOGNAS, G., ANNALA, M., KIVINUMMI, K., GOODY, V., LATIMER, C., O'MEARA, S., DAWSON, K. J., ISAACS, W., EMMERT-BUCK, M. R., NYKTER, M., FOSTER, C., KOTE-JARAI, Z., EASTON, D., WHITAKER, H. C., GROUP, I. P. U., NEAL, D. E., COOPER, C. S., EELES, R. A., VISAKORPI, T., CAMPBELL, P. J., MCDERMOTT, U., WEDGE, D. C. & BOVA, G. S. 2015. The evolutionary history of lethal metastatic prostate cancer. *Nature*, 520, 353-7.
- HAKIM, D. N., PELLY, T., KULENDRAN, M. & CARIS, J. A. 2015. Benign tumours of the bone: A review. *Journal of bone oncology*, 4, 37-41.
- HANAHAHAN, D. & COUSSENS, LISA M. 2012. Accessories to the Crime: Functions of Cells Recruited to the Tumor Microenvironment. *Cancer Cell*, 21, 309-322.
- HANAHAHAN, D. & WEINBERG, R. A. 2011. Hallmarks of cancer: the next generation. *Cell*, 144, 646-74.
- HARUTYUNYAN, A. S., KRUG, B., CHEN, H., PAPILLON-CAVANAGH, S., ZEINIEH, M., DE JAY, N., DESHMUKH, S., CHEN, C. C. L., BELLE, J.,

- MIKAEL, L. G., MARCHIONE, D. M., LI, R., NIKBAKHT, H., HU, B., CAGNONE, G., CHEUNG, W. A., MOHAMMADNIA, A., BECHET, D., FAURY, D., MCCONECHY, M. K., PATHANIA, M., JAIN, S. U., ELLEZAM, B., WEIL, A. G., MONTPETIT, A., SALOMONI, P., PASTINEN, T., LU, C., LEWIS, P. W., GARCIA, B. A., KLEINMAN, C. L., JABADO, N. & MAJEWSKI, J. 2019. H3K27M induces defective chromatin spread of PRC2-mediated repressive H3K27me2/me3 and is essential for glioma tumorigenesis. *Nat Commun*, 10, 1262.
- HEIDE, T., ZAPATA, L., WILLIAMS, M. J., WERNER, B., CARAVAGNA, G., BARNES, C. P., GRAHAM, T. A. & SOTTORIVA, A. 2018. Reply to 'Neutral tumor evolution?'. *Nature Genetics*, 50, 1633-1637.
- HENAO DIAZ, E., YACHNIN, J., GRONBERG, H. & LINDBERG, J. 2016. The In Vitro Stability of Circulating Tumour DNA. *PLoS One*, 11, e0168153.
- HIRSCH, F. R., OTTESEN, G., PODENPHANT, J. & OLSEN, J. 1983. TUMOR HETEROGENEITY IN LUNG-CANCER BASED ON LIGHT MICROSCOPIC FEATURES - A RETROSPECTIVE STUDY OF A CONSECUTIVE SERIES OF 200 PATIENTS, TREATED SURGICALLY. *Virchows Archiv a-Pathological Anatomy and Histopathology*, 402, 147-153.
- HORN, S., FIGL, A., RACHAKONDA, P. S., FISCHER, C., SUCKER, A., GAST, A., KADEL, S., MOLL, I., NAGORE, E., HEMMINKI, K., SCHADENDORF, D. & KUMAR, R. 2013. TERT promoter mutations in familial and sporadic melanoma. *Science*, 339, 959-61.
- HUANG, L., XU, J., WOOD, D. J. & ZHENG, M. H. 2000. Gene expression of osteoprotegerin ligand, osteoprotegerin, and receptor activator of NF-kappaB in giant cell tumor of bone: possible involvement in tumor cell-induced osteoclast-like cell formation. *Am J Pathol*, 156, 761-7.
- HUANG, S. C., ZHANG, L., SUNG, Y. S., CHEN, C. L., KRAUSZ, T., DICKSON, B. C., KAO, Y. C., AGARAM, N. P., FLETCHER, C. D. & ANTONESCU, C. R. 2015. Frequent FOS Gene Rearrangements in Epithelioid Hemangioma: A Molecular Study of 58 Cases With Morphologic Reappraisal. *Am J Surg Pathol*, 39, 1313-21.
- HUNG, Y. P., FLETCHER, C. D. & HORNICK, J. L. 2017. FOSB is a Useful Diagnostic Marker for Pseudomyogenic Hemangioendothelioma. *Am J Surg Pathol*, 41, 596-606.
- HUTTER, R. V. P., WORCESTER JR., J. N., FRANCIS, K. C., FOOTE JR., F. W. & STEWART, F. W. 1962. Benign and malignant giant cell tumors of bone. A clinicopathological analysis of the natural history of the disease. *Cancer*, 15, 653-690.
- HYMAN, D. M., PUZANOV, I., SUBBIAH, V., FARIS, J. E., CHAU, I., BLAY, J.-Y., WOLF, J., RAJE, N. S., DIAMOND, E. L., HOLLEBECQUE, A., GERVAIS, R., ELEZ-FERNANDEZ, M. E., ITALIANO, A., HOFHEINZ, R.-D., HIDALGO, M., CHAN, E., SCHULER, M., LASSERRE, S. F., MAKRUTZKI, M., SIRZEN, F., VERONESE, M. L., TABERNERO, J. & BASELGA, J. 2015. Vemurafenib in Multiple Nonmelanoma Cancers with BRAF V600 Mutations. *New England Journal of Medicine*, 373, 726-736.
- IDE, Y. H., TSUKAMOTO, Y., ITO, T., WATANABE, T., NAKAGAWA, N., HANEDA, T., NAGAI, M., YAMANISHI, K. & HIROTA, S. 2015. Penile pseudomyogenic hemangioendothelioma/epithelioid sarcoma-like

- hemangioendothelioma with a novel pattern of SERPINE1-FOSB fusion detected by RT-PCR--report of a case. *Pathol Res Pract*, 211, 415-20.
- IGNATIADIS, M. & SOTIRIOU, C. 2008. Understanding the molecular basis of histologic grade. *Pathobiology*, 75, 104-111.
- ITKIN, B., STRAMINSKY, S., DE RONATO, G., LEWI, D., MARANTZ, A. & BARDACH, A. 2018. Prognosis of metastatic giant cell tumor of bone in the pre-denosumab era. A systematic review and a meta-analysis. *Jpn J Clin Oncol*, 48, 640-652.
- JAFFE, H. L. 1953. Giant-cell tumour (osteoclastoma) of bone: its pathologic delimitation and the inherent clinical implications. *Annals of the Royal College of Surgeons of England*, 13, 343-355.
- JAMAL-HANJANI, M., WILSON, G. A., MCGRANAHAN, N., BIRKBAK, N. J., WATKINS, T. B. K., VEERIAH, S., SHAFI, S., JOHNSON, D. H., MITTER, R., ROSENTHAL, R., SALM, M., HORSWELL, S., ESCUDERO, M., MATTHEWS, N., ROWAN, A., CHAMBERS, T., MOORE, D. A., TURAJLIC, S., XU, H., LEE, S.-M., FORSTER, M. D., AHMAD, T., HILEY, C. T., ABBOSH, C., FALZON, M., BORG, E., MARAFIOTI, T., LAWRENCE, D., HAYWARD, M., KOLVEKAR, S., PANAGIOTOPOULOS, N., JANES, S. M., THAKRAR, R., AHMED, A., BLACKHALL, F., SUMMERS, Y., SHAH, R., JOSEPH, L., QUINN, A. M., CROSBIE, P. A., NAIDU, B., MIDDLETON, G., LANGMAN, G., TROTTER, S., NICOLSON, M., REMMEN, H., KERR, K., CHETTY, M., GOMERSALL, L., FENNELL, D. A., NAKAS, A., RATHINAM, S., ANAND, G., KHAN, S., RUSSELL, P., EZHIL, V., ISMAIL, B., IRVINSELLERS, M., PRAKASH, V., LESTER, J. F., KORNASZEWSKA, M., ATTANOOS, R., ADAMS, H., DAVIES, H., DENTRO, S., TANIÈRE, P., O'SULLIVAN, B., LOWE, H. L., HARTLEY, J. A., ILES, N., BELL, H., NGAI, Y., SHAW, J. A., HERRERO, J., SZALLASI, Z., SCHWARZ, R. F., STEWART, A., QUEZADA, S. A., LE QUESNE, J., VAN LOO, P., DIVE, C., HACKSHAW, A. & SWANTON, C. 2017. Tracking the Evolution of Non-Small-Cell Lung Cancer. *New England Journal of Medicine*, 376, 2109-2121.
- JESELSON, R., BUCHWALTER, G., DE ANGELIS, C., BROWN, M. & SCHIFF, R. 2015. ESR1 mutations-a mechanism for acquired endocrine resistance in breast cancer. *Nat Rev Clin Oncol*, 12, 573-83.
- JEWELL, J. H. & BUSH, L. F. 1964. "Benign" Giant-Cell Tumor of Bone with a Solitary Pulmonary Metastasis: A CASE REPORT. *JBJS*, 46, 848-852.
- JOHANNESSEN, C. M., BOEHM, J. S., KIM, S. Y., THOMAS, S. R., WARDWELL, L., JOHNSON, L. A., EMERY, C. M., STRANSKY, N., COGDILL, A. P., BARRETINA, J., CAPONIGRO, G., HIERONYMUS, H., MURRAY, R. R., SALEHI-ASHTIANI, K., HILL, D. E., VIDAL, M., ZHAO, J. J., YANG, X., ALKAN, O., KIM, S., HARRIS, J. L., WILSON, C. J., MYER, V. E., FINAN, P. M., ROOT, D. E., ROBERTS, T. M., GOLUB, T., FLAHERTY, K. T., DUMMER, R., WEBER, B. L., SELLERS, W. R., SCHLEGEL, R., WARGO, J. A., HAHN, W. C. & GARRAWAY, L. A. 2010. COT drives resistance to RAF inhibition through MAP kinase pathway reactivation. *Nature*, 468, 968-72.
- JOLLY, C. & VAN LOO, P. 2018. Timing somatic events in the evolution of cancer. *Genome Biol*, 19, 95.
- JONES, D., RAINE, K. M., DAVIES, H., TARPEY, P. S., BUTLER, A. P., TEAGUE, J. W., NIK-ZAINAL, S. & CAMPBELL, P. J. 2016. cgpCaVEManWrapper:

- Simple Execution of CaVEMan in Order to Detect Somatic Single Nucleotide Variants in NGS Data. *Curr Protoc Bioinformatics*, 56, 15 10 1-15 10 18.
- JOOSS, K. U., FUNK, M. & MÜLLER, R. 1994. An autonomous N-terminal transactivation domain in Fos protein plays a crucial role in transformation. *The EMBO Journal*, 13, 1467-1475.
- JUNG, S. H., SHIN, S., KIM, M. S., BAEK, I. P., LEE, J. Y., LEE, S. H., KIM, T. M., LEE, S. H. & CHUNG, Y. J. 2016. Genetic Progression of High Grade Prostatic Intraepithelial Neoplasia to Prostate Cancer. *Eur Urol*, 69, 823-30.
- KARLSSON, J., VALIND, A., HOLMQUIST MENGELBIER, L., BREDIN, S., CORNMARK, L., JANSSON, C., WALI, A., STAAF, J., VIKLUND, B., ØRA, I., BÖRJESSON, A., BACKMAN, T., BRAEKEVELDT, N., SANDSTEDT, B., PAL, N., ISAKSSON, A., LACKNER, B. G., JONSON, T., BEXELL, D. & GISSELSSON, D. 2018. Four evolutionary trajectories underlie genetic intratumoral variation in childhood cancer. *Nature Genetics*, 50, 944-950.
- KATO KANEKO, M., LIU, X., OKI, H., OGASAWARA, S., NAKAMURA, T., SAIDOH, N., TSUJIMOTO, Y., MATSUYAMA, Y., URUNO, A., SUGAWARA, M., TSUCHIYA, T., YAMAKAWA, M., YAMAMOTO, M., TAKAGI, M. & KATO, Y. 2014. Isocitrate dehydrogenase mutation is frequently observed in giant cell tumor of bone. *Cancer Sci*, 105, 744-8.
- KHAN, A. & ZHANG, X. 2016. dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res*, 44, D164-71.
- KHAN, K. H., CUNNINGHAM, D., WERNER, B., VLACHOGIANNIS, G., SPITERI, I., HEIDE, T., MATEOS, J. F., VATSIU, A., LAMPIS, A., DAMAVANDI, M. D., LOTE, H., HUNTINGFORD, I. S., HEDAYAT, S., CHAU, I., TUNARIU, N., MENTRASTI, G., TREVISANI, F., RAO, S., ANANDAPPA, G., WATKINS, D., STARLING, N., THOMAS, J., PECKITT, C., KHAN, N., RUGGE, M., BEGUM, R., HEZELOVA, B., BRYANT, A., JONES, T., PROSZEK, P., FASSAN, M., HAHNE, J. C., HUBANK, M., BRACONI, C., SOTTORIVA, A. & VALERI, N. 2018. Longitudinal Liquid Biopsy and Mathematical Modeling of Clonal Evolution Forecast Time to Treatment Failure in the PROSPECT-C Phase II Colorectal Cancer Clinical Trial. *Cancer Discov*, 8, 1270-1285.
- KIM, C., GAO, R., SEI, E., BRANDT, R., HARTMAN, J., HATSCHEK, T., CROSETTO, N., FOUKAKIS, T. & NAVIN, N. E. 2018. Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. *Cell*, 173, 879-893 e13.
- KIM, D., PERTEA, G., TRAPNELL, C., PIMENTEL, H., KELLEY, R. & SALZBERG, S. L. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*, 14.
- KLEINHEINZ, K., BLUDAU, I., HUEBSCHMANN, D., HEINOLD, M., KENSCH, P., GU, Z., LOPEZ, C., HUMMEL, M., KLAPPER, W., MOELLER, P., VATER, I., WAGENER, R., BRORS, B., SIEBERT, R., EILS, R. & SCHLESNER, M. 2017. ACEseq - allele specific copy number estimation from whole genome sequencing. *bioRxiv*.
- KLEMM, S. L., SHIPONY, Z. & GREENLEAF, W. J. 2019. Chromatin accessibility and the regulatory epigenome. *Nature Reviews Genetics*, 20, 207-220.
- KNAUF, J. A., ELISEI, R., MOCHLY-ROSEN, D., LIRON, T., CHEN, X. N., GONSKY, R., KORENBERG, J. R. & FAGIN, J. A. 1999. Involvement of protein kinase Cepsilon (PKCepsilon) in thyroid cell death. A truncated chimeric

- PKCepsilon cloned from a thyroid cancer cell line protects thyroid cells from apoptosis. *J Biol Chem*, 274, 23414-25.
- KNEZEVIĆ, S. R., MCFADDEN, D. E., TAO, W., LIM, J. F. & SORENSEN, P. H. 1998. A novel ETV6-NTRK3 gene fusion in congenital fibrosarcoma. *Nat Genet*, 18, 184-7.
- KOELSCHE, C., SCHRIMPF, D., THARUN, L., ROTH, E., STURM, D., JONES, D. T. W., RENKER, E. K., SILL, M., BAUDE, A., SAHM, F., CAPPER, D., BEWERUNGE-HUDLER, M., HARTMANN, W., KULOZIK, A. E., PETERSEN, I., FLUCKE, U., SCHREUDER, H. W. B., BUTTNER, R., WEBER, M. A., SCHIRMACHER, P., PLASS, C., PFISTER, S. M., VON DEIMLING, A. & MECHTERSHEIMER, G. 2017. Histone 3.3 hotspot mutations in conventional osteosarcomas: a comprehensive clinical and molecular characterization of six H3F3A mutated cases. *Clin Sarcoma Res*, 7, 9.
- KOLANOWSKA, M., WÓJCICKA, A., KUBIAK, A., ŚWIERNIAK, M., KOTLAREK, M., MACIĄG, M., GAJ, P., KOPERSKI, Ł., GÓRNICKA, B. & JAŹDŹEWSKI, K. 2017. Functional analysis of a novel, thyroglobulin-embedded microRNA gene deregulated in papillary thyroid carcinoma. *Scientific Reports*, 7, 9942.
- KORBEL, J. O. & CAMPBELL, P. J. 2013. Criteria for inference of chromothripsis in cancer genomes. *Cell*, 152.
- KOVAC, M., BLATTMANN, C., RIBI, S., SMIDA, J., MUELLER, N. S., ENGERT, F., CASTRO-GINER, F., WEISCHENFELDT, J., KOVACOVA, M., KRIEG, A., ANDREOU, D., TUNN, P. U., DURR, H. R., RECHL, H., SCHASER, K. D., MELCHER, I., BURDACH, S., KULOZIK, A., SPECHT, K., HEINIMANN, K., FULDA, S., BIELACK, S., JUNDT, G., TOMLINSON, I., KORBEL, J. O., NATHRATH, M. & BAUMHOER, D. 2015. Exome sequencing of osteosarcoma reveals mutation signatures reminiscent of BRCA deficiency. *Nat Commun*, 6, 8940.
- KREBS, M. G., METCALF, R. L., CARTER, L., BRADY, G., BLACKHALL, F. H. & DIVE, C. 2014. Molecular analysis of circulating tumour cells-biology and biomarkers. *Nat Rev Clin Oncol*, 11, 129-44.
- KRUG, B., DE JAY, N., HARUTYUNYAN, A. S., DESHMUKH, S., MARCHIONE, D. M., GUILHAMON, P., BERTRAND, K. C., MIKAEL, L. G., MCCONECHY, M. K., CHEN, C. C. L., KHAZAEI, S., KONCAR, R. F., AGNIHOTRI, S., FAURY, D., ELLEZAM, B., WEIL, A. G., URSINI-SIEGEL, J., DE CARVALHO, D. D., DIRKS, P. B., LEWIS, P. W., SALOMONI, P., LUPIEN, M., ARROWSMITH, C., LASKO, P. F., GARCIA, B. A., KLEINMAN, C. L., JABADO, N. & MACK, S. C. 2019. Pervasive H3K27 Acetylation Leads to ERV Expression and a Therapeutic Vulnerability in H3K27M Gliomas. *Cancer Cell*, 35, 782-797.e8.
- KRUGER, S., THORNS, C., BOHLE, A. & FELLER, A. C. 2003. Prognostic significance of a grading system considering tumor heterogeneity in muscle-invasive urothelial carcinoma of the urinary bladder. *International urology and nephrology*, 35, 169-73.
- KUMAR, V. 2015. *Robbins and Cotran pathologic basis of disease*, Philadelphia, Philadelphia : Elsevier/Saunders.
- LAWRENCE, M. S., STOJANOV, P., POLAK, P., KRYUKOV, G. V., CIBULSKIS, K., SIVACHENKO, A., CARTER, S. L., STEWART, C., MERMEL, C. H., ROBERTS, S. A., KIEZUN, A., HAMMERMAN, P. S., MCKENNA, A., DRIER,

- Y., ZOU, L., RAMOS, A. H., PUGH, T. J., STRANSKY, N., HELMAN, E., KIM, J., SOUGNEZ, C., AMBROGIO, L., NICKERSON, E., SHEFLER, E., CORTES, M. L., AUCLAIR, D., SAKSENA, G., VOET, D., NOBLE, M., DICARA, D., LIN, P., LICHTENSTEIN, L., HEIMAN, D. I., FENNELL, T., IMIELINSKI, M., HERNANDEZ, B., HODIS, E., BACA, S., DULAK, A. M., LOHR, J., LANDAU, D.-A., WU, C. J., MELENDEZ-ZAJGLA, J., HIDALGO-MIRANDA, A., KOREN, A., MCCARROLL, S. A., MORA, J., LEE, R. S., CROMPTON, B., ONOFRIO, R., PARKIN, M., WINCKLER, W., ARDLIE, K., GABRIEL, S. B., ROBERTS, C. W. M., BIEGEL, J. A., STEGMAIER, K., BASS, A. J., GARRAWAY, L. A., MEYERSON, M., GOLUB, T. R., GORDENIN, D. A., SUNYAEV, S., LANDER, E. S. & GETZ, G. 2013. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*, 499, 214-218.
- LAWRENCE, T. S., ROSENBERG, S. A., HELLMAN, S., DEVITA, V. T., LAWRENCE, T. S. & ROSENBERG, S. A. 2019. *DeVita, Hellman, and Rosenberg's cancer : principles & practice of oncology / [edited by] Vincent T. DeVita, Jr., Theodore S. Lawrence, Steven A. Rosenberg*, Philadelphia, Philadelphia, PA : Wolters Kluwer.
- LEE, J. C., LIANG, C. W. & FLETCHER, C. D. 2017a. Giant cell tumor of soft tissue is genetically distinct from its bone counterpart. *Mod Pathol*, 30, 728-733.
- LEE, J. J., PARK, S., PARK, H., KIM, S., LEE, J., LEE, J., YOUK, J., YI, K., AN, Y., PARK, I. K., KANG, C. H., CHUNG, D. H., KIM, T. M., JEON, Y. K., HONG, D., PARK, P. J., JU, Y. S. & KIM, Y. T. 2019. Tracing Oncogene Rearrangements in the Mutational History of Lung Adenocarcinoma. *Cell*, 177, 1842-1857.e21.
- LEE, J. K., LOUZADA, S., AN, Y., KIM, S. Y., KIM, S., YOUK, J., PARK, S., KOO, S. H., KEAM, B., JEON, Y. K., KU, J. L., YANG, F., KIM, T. M. & JU, Y. S. 2017b. Complex chromosomal rearrangements by single catastrophic pathogenesis in NUT midline carcinoma. *Ann Oncol*, 28, 890-897.
- LESHCHINER, I., LIVITZ, D., GAINOR, J. F., ROSEBROCK, D., SPIRO, O., MARTINEZ, A., MROZ, E., LIN, J. J., STEWART, C., KIM, J., ELAGINA, L., MINO-KENUDSON, M., ROONEY, M., IGNATIUS OU, S.-H., WU, C. J., ROCCO, J. W., ENGELMAN, J. A., SHAW, A. T. & GETZ, G. 2018. Comprehensive analysis of tumour initiation, spatial and temporal progression under multiple lines of treatment. *bioRxiv*, 508127.
- LEVAN, G. & MITELMAN, F. 1975. Clustering of aberrations to specific chromosomes in human neoplasms. *Hereditas*, 79, 156-60.
- LEWIS, P. W., MULLER, M. M., KOLETSKY, M. S., CORDERO, F., LIN, S., BANASZYNSKI, L. A., GARCIA, B. A., MUIR, T. W., BECHER, O. J. & ALLIS, C. D. 2013. Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science*, 340, 857-61.
- LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25, 1754-60.
- LI, Y., ROBERTS, N. D., WALA, J. A., SHAPIRA, O., SCHUMACHER, S. E., KUMAR, K., KHURANA, E., WASZAK, S., KORBEL, J. O., HABER, J. E., IMIELINSKI, M., WEISCHENFELDT, J., BEROUKHIM, R. & CAMPBELL, P. J. 2020. Patterns of somatic structural variation in human cancer genomes. *Nature*, 578, 112-121.
- LI, Y. & TOLLEFSBOL, T. O. 2011. DNA methylation detection: bisulfite genomic sequencing analysis. *Methods in molecular biology (Clifton, N.J.)*, 791, 11-21.

- LI, Y., ZHOU, S., SCHWARTZ, D. C. & MA, J. 2016. Allele-Specific Quantification of Structural Variations in Cancer Genomes. *Cell Syst*, 3, 21-34.
- LIM, J., PARK, J. H., BAUDE, A., YOO, Y., LEE, Y. K., SCHMIDT, C. R., PARK, J. B., FELLEBERG, J., ZUSTIN, J., HALLER, F., KRÜCKEN, I., KANG, H. G., PARK, Y. J., PLASS, C. & LINDROTH, A. M. 2017. The histone variant H3.3 G34W substitution in giant cell tumor of the bone link chromatin and RNA processing. *Scientific Reports*, 7, 13459.
- LINDER, D. & GARTLER, S. M. 1965. Glucose-6-phosphate dehydrogenase mosaicism: utilization as a cell marker in the study of leiomyomas. *Science*, 150, 67-9.
- LIU, P., EREZ, A., NAGAMANI, S. C., DHAR, S. U., KOLODZIEJSKA, K. E., DHARMADHIKARI, A. V., COOPER, M. L., WISZNIEWSKA, J., ZHANG, F., WITHERS, M. A., BACINO, C. A., CAMPOS-ACEVEDO, L. D., DELGADO, M. R., FREEDENBERG, D., GARNICA, A., GREBE, T. A., HERNANDEZ-ALMAGUER, D., IMMKEN, L., LALANI, S. R., MCLEAN, S. D., NORTHRUP, H., SCAGLIA, F., STRATHEARN, L., TRAPANE, P., KANG, S. H., PATEL, A., CHEUNG, S. W., HASTINGS, P. J., STANKIEWICZ, P., LUPSKI, J. R. & BI, W. 2011. Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements. *Cell*, 146, 889-903.
- LONG, G. V., STROYAKOVSKIY, D., GOGAS, H., LEVCHENKO, E., DE BRAUD, F., LARKIN, J., GARBE, C., JOUARY, T., HAUSCHILD, A., GROB, J. J., CHIARION SILENI, V., LEBBE, C., MANDALA, M., MILLWARD, M., ARANCE, A., BONDARENKO, I., HAANEN, J. B., HANSSON, J., UTIKAL, J., FERRARESI, V., KOVALENKO, N., MOHR, P., PROBACHAI, V., SCHADENDORF, D., NATHAN, P., ROBERT, C., RIBAS, A., DEMARINI, D. J., IRANI, J. G., CASEY, M., OUELLET, D., MARTIN, A. M., LE, N., PATEL, K. & FLAHERTY, K. 2014. Combined BRAF and MEK inhibition versus BRAF inhibition alone in melanoma. *N Engl J Med*, 371, 1877-88.
- LORD, C. J. & ASHWORTH, A. 2016. BRCAness revisited. *Nat Rev Cancer*, 16, 110-20.
- LU, C., JAIN, S. U., HOELPER, D., BECHET, D., MOLDEN, R. C., RAN, L., MURPHY, D., VENNETI, S., HAMEED, M., PAWEL, B. R., WUNDER, J. S., DICKSON, B. C., LUNDGREN, S. M., JANI, K. S., DE JAY, N., PAPILLON-CAVANAGH, S., ANDRULIS, I. L., SAWYER, S. L., GRYNSPAN, D., TURCOTTE, R. E., NADAF, J., FAHIMINIYAH, S., MUIR, T. W., MAJEWSKI, J., THOMPSON, C. B., CHI, P., GARCIA, B. A., ALLIS, C. D., JABADO, N. & LEWIS, P. W. 2016. Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. *Science*, 352, 844-9.
- LUENGO-ALONSO, G., MELLADO-ROMERO, M., SHEMESH, S., RAMOS-PASCUA, L. & PRETELL-MAZZINI, J. 2019. Denosumab treatment for giant-cell tumor of bone: a systematic review of the literature. *Arch Orthop Trauma Surg*.
- LUKE, J., VON BAER, A., SCHREIBER, J., LUBBEHUSEN, C., BREINING, T., MELLERT, K., MARIENFELD, R., SCHULTHEISS, M., MOLLER, P. & BARTH, T. F. E. 2017. H3F3A mutation in giant cell tumour of the bone is detected by immunohistochemistry using a monoclonal antibody against the G34W mutated site of the histone H3.3 variant. *Histopathology*, 71, 125-133.

- LUO, Y., TANG, F., WANG, Y., ZHOU, Y., MIN, L., ZHANG, W., SHI, R., DUAN, H. & TU, C. 2018. Safety and efficacy of denosumab in the treatment of pulmonary metastatic giant cell tumor of bone. *Cancer Manag Res*, 10, 1901-1906.
- MACAULAY, I. C., HAERTY, W., KUMAR, P., LI, Y. I., HU, T. X., TENG, M. J., GOOLAM, M., SAURAT, N., COUPLAND, P., SHIRLEY, L. M., SMITH, M., VAN DER AA, N., BANERJEE, R., ELLIS, P. D., QUAIL, M. A., SWERDLOW, H. P., ZERNICKA-GOETZ, M., LIVESEY, F. J., PONTING, C. P. & VOET, T. 2015. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat Methods*, 12, 519-22.
- MACAULAY, I. C., PONTING, C. P. & VOET, T. 2017. Single-Cell Multiomics: Multiple Measurements from Single Cells. *Trends Genet*, 33, 155-168.
- MACINTYRE, G., GORANOVA, T. E., DE SILVA, D., ENNIS, D., PISKORZ, A. M., ELDRIDGE, M., SIE, D., LEWSLEY, L. A., HANIF, A., WILSON, C., DOWSON, S., GLASSPOOL, R. M., LOCKLEY, M., BROCKBANK, E., MONTES, A., WALTHER, A., SUNDAR, S., EDMONDSON, R., HALL, G. D., CLAMP, A., GOURLEY, C., HALL, M., FOTOPOULOU, C., GABRA, H., PAUL, J., SUPERNAT, A., MILLAN, D., HOYLE, A., BRYSON, G., NOURSE, C., MINCARELLI, L., SANCHEZ, L. N., YLSTRA, B., JIMENEZ-LINAN, M., MOORE, L., HOFMANN, O., MARKOWETZ, F., MCNEISH, I. A. & BRENTON, J. D. 2018. Copy number signatures and mutational processes in ovarian carcinoma. *Nat Genet*, 50, 1262-1270.
- MAHER, C. A. & WILSON, R. K. 2012. Chromothripsis and human disease: piecing together the shattering process. *Cell*, 148, 29-32.
- MALEY, C. C., GALIPEAU, P. C., FINLEY, J. C., WONGSURAWAT, V. J., LI, X. H., SANCHEZ, C. A., PAULSON, T. G., BLOUNT, P. L., RISQUES, R. A., RABINOVITCH, P. S. & REID, B. J. 2006. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nature Genetics*, 38, 468-473.
- MARKOWETZ, F. 2016. A saltationist theory of cancer evolution. *Nat Genet*, 48, 1102-3.
- MARTINCORENA, I., FOWLER, J. C., WABIK, A., LAWSON, A. R. J., ABASCAL, F., HALL, M. W. J., CAGAN, A., MURAI, K., MAHBUBANI, K., STRATTON, M. R., FITZGERALD, R. C., HANDFORD, P. A., CAMPBELL, P. J., SAEB-PARSY, K. & JONES, P. H. 2018. Somatic mutant clones colonize the human esophagus with age. *Science*, 362, 911-917.
- MARTINCORENA, I., RAINE, K. M., GERSTUNG, M., DAWSON, K. J., HAASE, K., VAN LOO, P., DAVIES, H., STRATTON, M. R. & CAMPBELL, P. J. 2017. Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell*, 171, 1029-1041.e21.
- MARTINCORENA, I., ROSHAN, A., GERSTUNG, M., ELLIS, P., VAN LOO, P., MCLAREN, S., WEDGE, D. C., FULLAM, A., ALEXANDROV, L. B., TUBIO, J. M., STEBBINGS, L., MENZIES, A., WIDAA, S., STRATTON, M. R., JONES, P. H. & CAMPBELL, P. J. 2015. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science*, 348, 880-6.
- MATSUBAYASHI, S., NAKASHIMA, M., KUMAGAI, K., EGASHIRA, M., NARUKE, Y., KONDO, H., HAYASHI, T. & SHINDO, H. 2009. Immunohistochemical analyses of β -catenin and cyclin D1 expression in giant cell

- tumor of bone (GCTB): A possible role of Wnt pathway in GCTB tumorigenesis. *Pathology - Research and Practice*, 205, 626-633.
- MCCLINTOCK, B. 1938. The Production of Homozygous Deficient Tissues with Mutant Characteristics by Means of the Aberrant Mitotic Behavior of Ring-Shaped Chromosomes. *Genetics*, 23, 315-376.
- MCGRANAHAN, N., FURNESS, A. J., ROSENTHAL, R., RAMSKOV, S., LINGAA, R., SAINI, S. K., JAMAL-HANJANI, M., WILSON, G. A., BIRKBAK, N. J., HILEY, C. T., WATKINS, T. B., SHAFI, S., MURUGAESU, N., MITTER, R., AKARCA, A. U., LINARES, J., MARAFIOTI, T., HENRY, J. Y., VAN ALLEN, E. M., MIAO, D., SCHILLING, B., SCHADENDORF, D., GARRAWAY, L. A., MAKAROV, V., RIZVI, N. A., SNYDER, A., HELLMANN, M. D., MERGHOUB, T., WOLCHOK, J. D., SHUKLA, S. A., WU, C. J., PEGGS, K. S., CHAN, T. A., HADRUP, S. R., QUEZADA, S. A. & SWANTON, C. 2016. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science*, 351, 1463-9.
- MCGRANAHAN, N., ROSENTHAL, R., HILEY, C. T., ROWAN, A. J., WATKINS, T. B. K., WILSON, G. A., BIRKBAK, N. J., VEERIAH, S., VAN LOO, P., HERRERO, J., SWANTON, C., SWANTON, C., JAMAL-HANJANI, M., VEERIAH, S., SHAFI, S., CZYZEWSKA-KHAN, J., JOHNSON, D., LAYCOCK, J., BOSSHARD-CARTER, L., ROSENTHAL, R., GORMAN, P., HYND, R. E., WILSON, G., BIRKBAK, N. J., WATKINS, T. B. K., MCGRANAHAN, N., HORSWELL, S., MITTER, R., ESCUDERO, M., STEWART, A., VAN LOO, P., ROWAN, A., XU, H., TURAJLIC, S., HILEY, C., ABBOSH, C., GOLDMAN, J., STONE, R. K., DENNER, T., MATTHEWS, N., ELGAR, G., WARD, S., COSTA, M., BEGUM, S., PHILLIMORE, B., CHAMBERS, T., NYE, E., GRACA, S., AL BAKIR, M., JOSHI, K., FURNESS, A., BEN AISSA, A., WONG, Y. N. S., GEORGIU, A., QUEZADA, S., HARTLEY, J. A., LOWE, H. L., HERRERO, J., LAWRENCE, D., HAYWARD, M., PANAGIOTOPOULOS, N., KOLVEKAR, S., FALZON, M., BORG, E., MARAFIOTI, T., SIMEON, C., HECTOR, G., SMITH, A., ARANDA, M., NOVELLI, M., OUKRIF, D., JANES, S. M., THAKRAR, R., FORSTER, M., AHMAD, T., LEE, S. M., PAPADATOS-PASTOS, D., CARNELL, D., MENDES, R., GEORGE, J., NAVANI, N., AHMED, A., TAYLOR, M., CHOUDHARY, J., SUMMERS, Y., CALIFANO, R., TAYLOR, P., SHAH, R., KRYSIAK, P., RAMMOHAN, K., FONTAINE, E., BOOTON, R., EIVSON, M., CROSBIE, P., MOSS, S., IDRIES, F., JOSEPH, L., BISHOP, P., CHATURVED, A., QUINN, A. M., et al. Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell*, 171, 1259-1271.e11.
- MCKENNA, A., HANNA, M., BANKS, E., SIVACHENKO, A., CIBULSKIS, K., KERNYTSKY, A., GARIMELLA, K., ALTSHULER, D., GABRIEL, S., DALY, M. & DEPRISTO, M. A. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*, 20, 1297-303.
- MCPHERSON, A., HORMOZDIARI, F., ZAYED, A., GIULIANI, R., HA, G., SUN, M. G., GRIFFITH, M., HERAVI MOUSSAVI, A., SENZ, J., MELNYK, N., PACHECO, M., MARRA, M. A., HIRST, M., NIELSEN, T. O., SAHINALP, S. C., HUNTSMAN, D. & SHAH, S. P. 2011. deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. *PLoS Comput Biol*, 7, e1001138.

- MEDVEDEV, P., FIUME, M., DZAMBA, M., SMITH, T. & BRUDNO, M. 2010. Detecting copy number variation with mated short reads. *Genome Res*, 20, 1613-22.
- MERLO, A., HERMAN, J. G., MAO, L., LEE, D. J., GABRIELSON, E., BURGER, P. C., BAYLIN, S. B. & SIDRANSKY, D. 1995. 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. *Nature Medicine*, 1, 686-692.
- MERTENS, F., ANTONESCU, C. R., HOHENBERGER, P., LADANYI, M., MODENA, P., D'INCALCI, M., CASALI, P. G., AGLIETTA, M. & ALVEGARD, T. 2009. Translocation-related sarcomas. *Semin Oncol*, 36, 312-23.
- METZGER, E., WILLMANN, D., MCMILLAN, J., FORNE, I., METZGER, P., GERHARDT, S., PETROLL, K., VON MAESSENHAUSEN, A., URBAN, S., SCHOTT, A. K., ESPEJO, A., EBERLIN, A., WOHLWEND, D., SCHULE, K. M., SCHLEICHER, M., PERNER, S., BEDFORD, M. T., JUNG, M., DENGJEL, J., FLAIG, R., IMHOF, A., EINSLE, O. & SCHULE, R. 2016. Assembly of methylated KDM1A and CHD1 drives androgen receptor-dependent transcription and translocation. *Nat Struct Mol Biol*, 23, 132-9.
- MEYERSON, M., GABRIEL, S. & GETZ, G. 2010. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet*, 11, 685-96.
- MILLER, A. D., CURRAN, T. & VERMA, I. M. 1984. c-fos protein can induce cellular transformation: a novel mechanism of activation of a cellular oncogene. *Cell*, 36, 51-60.
- MILO, I., BEDORA-FAURE, M., GARCIA, Z., THIBAUT, R., PÉRIÉ, L., SHAKHAR, G., DERIANO, L. & BOUSSO, P. 2018. The immune system profoundly restricts intratumor genetic heterogeneity. *Science immunology*, 3, eaat1435.
- MITCHELL, T. J., TURAJLIC, S., ROWAN, A., NICOL, D., FARMERY, J. H. R., O'BRIEN, T., MARTINCORENA, I., TARPEY, P., ANGELOPOULOS, N., YATES, L. R., BUTLER, A. P., RAINE, K., STEWART, G. D., CHALLACOMBE, B., FERNANDO, A., LOPEZ, J. I., HAZELL, S., CHANDRA, A., CHOWDHURY, S., RUDMAN, S., SOULTATI, A., STAMP, G., FOTIADIS, N., PICKERING, L., AU, L., SPAIN, L., LYNCH, J., STARES, M., TEAGUE, J., MAURA, F., WEDGE, D. C., HORSWELL, S., CHAMBERS, T., LITCHFIELD, K., XU, H., STEWART, A., ELAIDI, R., OUDARD, S., MCGRANAHAN, N., CSABAI, I., GORE, M., FUTREAL, P. A., LARKIN, J., LYNCH, A. G., SZALLASI, Z., SWANTON, C., CAMPBELL, P. J. & CONSORTIUM, T. R. R. 2018. Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer: TRACERx Renal. *Cell*, 173, 611-623 e17.
- MONCUNILL, V., GONZALEZ, S., BEA, S., ANDRIEUX, L. O., SALAVERRIA, I., ROYO, C., MARTINEZ, L., PUIGGROS, M., SEGURA-WANG, M., STUTZ, A. M., NAVARRO, A., ROYO, R., GELPI, J. L., GUT, I. G., LOPEZ-OTIN, C., OROZCO, M., KORBEL, J. O., CAMPO, E., PUENTE, X. S. & TORRENTS, D. 2014. Comprehensive characterization of complex structural variations in cancer by directly comparing genome sequence reads. *Nat Biotechnol*, 32, 1106-12.
- MORA, J., CHEUNG, N. K. V. & GERALD, W. L. 2001. Genetic heterogeneity and clonal evolution in neuroblastoma. *British Journal of Cancer*, 85, 182-189.
- MORICEAU, G., HUGO, W., HONG, A., SHI, H., KONG, X., YU, C. C., KOYA, R. C., SAMATAR, A. A., KHANLOU, N., BRAUN, J., RUCHALSKI, K., SEIFERT, H., LARKIN, J., DAHLMAN, K. B., JOHNSON, D. B., ALGAZI, A., SOSMAN,

- J. A., RIBAS, A. & LO, R. S. 2015. Tunable-combinatorial mechanisms of acquired resistance limit the efficacy of BRAF/MEK cotargeting but result in melanoma drug addiction. *Cancer Cell*, 27, 240-56.
- MUHEREMU, A. & NIU, X. 2014. Pulmonary metastasis of giant cell tumor of bones. *World J Surg Oncol*, 12, 261.
- MUKHERJEE, S. 2011. *The emperor of all maladies : a biography of cancer*, London, Fourth Estate.
- MULLER, F. L., AQUILANTI, E. A. & DEPINHO, R. A. 2015. Collateral Lethality: A new therapeutic strategy in oncology. *Trends Cancer*, 1, 161-173.
- MULLER, F. L., COLLA, S., AQUILANTI, E., MANZO, V. E., GENOVESE, G., LEE, J., EISENSON, D., NARURKAR, R., DENG, P., NEZI, L., LEE, M. A., HU, B., HU, J., SAHIN, E., ONG, D., FLETCHER-SANANIKONE, E., HO, D., KWONG, L., BRENNAN, C., WANG, Y. A., CHIN, L. & DEPINHO, R. A. 2012. Passenger deletions generate therapeutic vulnerabilities in cancer. *Nature*, 488, 337-42.
- MURTAZA, M., DAWSON, S. J., POGREBNIAK, K., RUEDA, O. M., PROVENZANO, E., GRANT, J., CHIN, S. F., TSUI, D. W., MARASS, F., GALE, D., ALI, H. R., SHAH, P., CONTENTE-CUOMO, T., FARAHANI, H., SHUMANSKY, K., KINGSBURY, Z., HUMPHRAY, S., BENTLEY, D., SHAH, S. P., WALLIS, M., ROSENFELD, N. & CALDAS, C. 2015. Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. *Nat Commun*, 6, 8760.
- NACEV, B. A., FENG, L., BAGERT, J. D., LEMIESZ, A. E., GAO, J., SOSHNEV, A. A., KUNDRA, R., SCHULTZ, N., MUIR, T. W. & ALLIS, C. D. 2019. The expanding landscape of 'oncohistone' mutations in human cancers. *Nature*, 567, 473-478.
- NAVIN, N. & HICKS, J. 2011. Future medical applications of single-cell sequencing in cancer. *Genome Med*, 3, 31.
- NAVIN, N., KENDALL, J., TROGE, J., ANDREWS, P., RODGERS, L., MCINDOO, J., COOK, K., STEPANSKY, A., LEVY, D., ESPOSITO, D., MUTHUSWAMY, L., KRASNITZ, A., MCCOMBIE, W. R., HICKS, J. & WIGLER, M. 2011. Tumour evolution inferred by single-cell sequencing. *Nature*, 472, 90-4.
- NAVIN, N., KRASNITZ, A., RODGERS, L., COOK, K., METH, J., KENDALL, J., RIGGS, M., EBERLING, Y., TROGE, J., GRUBOR, V., LEVY, D., LUNDIN, P., MANER, S., ZETTERBERG, A., HICKS, J. & WIGLER, M. 2010. Inferring tumor progression from genomic heterogeneity. *Genome Research*, 20, 68-80.
- NAVIN, N. E. 2014. Cancer genomics: one cell at a time. *Genome Biol*, 15, 452.
- NAVIN, N. E. 2015. The first five years of single-cell cancer genomics and beyond. *Genome Research*, 25, 1499-1507.
- NAZARIAN, R., SHI, H., WANG, Q., KONG, X., KOYA, R. C., LEE, H., CHEN, Z., LEE, M. K., ATTAR, N., SAZEGAR, H., CHODON, T., NELSON, S. F., MCARTHUR, G., SOSMAN, J. A., RIBAS, A. & LO, R. S. 2010. Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature*, 468, 973-7.
- NEWMAN, A. M., BRATMAN, S. V., TO, J., WYNNE, J. F., ECLOV, N. C., MODLIN, L. A., LIU, C. L., NEAL, J. W., WAKELEE, H. A., MERRITT, R. E., SHRAGER, J. B., LOO, B. W., JR., ALIZADEH, A. A. & DIEHN, M. 2014. An ultrasensitive

method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med*, 20, 548-54.

- NIK-ZAINAL, S., ALEXANDROV, L. B., WEDGE, D. C., VAN LOO, P., GREENMAN, C. D., RAINE, K., JONES, D., HINTON, J., MARSHALL, J., STEBBINGS, L. A., MENZIES, A., MARTIN, S., LEUNG, K., CHEN, L., LEROY, C., RAMAKRISHNA, M., RANCE, R., LAU, K. W., MUDIE, L. J., VARELA, I., MCBRIDE, D. J., BIGNELL, G. R., COOKE, S. L., SHLIEN, A., GAMBLE, J., WHITMORE, I., MADDISON, M., TARPEY, P. S., DAVIES, H. R., PAPAEMMANUIL, E., STEPHENS, P. J., MCLAREN, S., BUTLER, A. P., TEAGUE, J. W., JONSSON, G., GARBER, J. E., SILVER, D., MIRON, P., FATIMA, A., BOYAULT, S., LANGEROD, A., TUTT, A., MARTENS, J. W., APARICIO, S. A., BORG, A., SALOMON, A. V., THOMAS, G., BORRESENDALE, A. L., RICHARDSON, A. L., NEUBERGER, M. S., FUTREAL, P. A., CAMPBELL, P. J., STRATTON, M. R. & BREAST CANCER WORKING GROUP OF THE INTERNATIONAL CANCER GENOME, C. 2012a. Mutational processes molding the genomes of 21 breast cancers. *Cell*, 149, 979-93.
- NIK-ZAINAL, S., DAVIES, H., STAAF, J., RAMAKRISHNA, M., GLODZIK, D., ZOU, X., MARTINCORENA, I., ALEXANDROV, L. B., MARTIN, S., WEDGE, D. C., VAN LOO, P., JU, Y. S., SMID, M., BRINKMAN, A. B., MORGANELLA, S., AURE, M. R., LINGJAERDE, O. C., LANGEROD, A., RINGNER, M., AHN, S. M., BOYAULT, S., BROCK, J. E., BROEKS, A., BUTLER, A., DESMEDT, C., DIRIX, L., DRONOV, S., FATIMA, A., FOEKENS, J. A., GERSTUNG, M., HOOIJER, G. K., JANG, S. J., JONES, D. R., KIM, H. Y., KING, T. A., KRISHNAMURTHY, S., LEE, H. J., LEE, J. Y., LI, Y., MCLAREN, S., MENZIES, A., MUSTONEN, V., O'MEARA, S., PAUORTE, I., PIVOT, X., PURDIE, C. A., RAINE, K., RAMAKRISHNAN, K., RODRIGUEZ-GONZALEZ, F. G., ROMIEU, G., SIEUWERTS, A. M., SIMPSON, P. T., SHEPHERD, R., STEBBINGS, L., STEFANSSON, O. A., TEAGUE, J., TOMMASI, S., TREILLEUX, I., VAN DEN EYNDEN, G. G., VERMEULEN, P., VINCENT-SALOMON, A., YATES, L., CALDAS, C., VAN'T VEER, L., TUTT, A., KNAPPSKOG, S., TAN, B. K., JONKERS, J., BORG, A., UENO, N. T., SOTIRIOU, C., VIARI, A., FUTREAL, P. A., CAMPBELL, P. J., SPAN, P. N., VAN LAERE, S., LAKHANI, S. R., EYFJORD, J. E., THOMPSON, A. M., BIRNEY, E., STUNNENBERG, H. G., VAN DE VIJVER, M. J., MARTENS, J. W., BORRESENDALE, A. L., RICHARDSON, A. L., KONG, G., THOMAS, G. & STRATTON, M. R. 2016. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*, 534, 47-54.
- NIK-ZAINAL, S., VAN LOO, P., WEDGE, D. C., ALEXANDROV, L. B., GREENMAN, C. D., LAU, K. W., RAINE, K., JONES, D., MARSHALL, J., RAMAKRISHNA, M., SHLIEN, A., COOKE, S. L., HINTON, J., MENZIES, A., STEBBINGS, L. A., LEROY, C., JIA, M., RANCE, R., MUDIE, L. J., GAMBLE, S. J., STEPHENS, P. J., MCLAREN, S., TARPEY, P. S., PAPAEMMANUIL, E., DAVIES, H. R., VARELA, I., MCBRIDE, D. J., BIGNELL, G. R., LEUNG, K., BUTLER, A. P., TEAGUE, J. W., MARTIN, S., JONSSON, G., MARIANI, O., BOYAULT, S., MIRON, P., FATIMA, A., LANGEROD, A., APARICIO, S. A., TUTT, A., SIEUWERTS, A. M., BORG, A., THOMAS, G., SALOMON, A. V.,

- RICHARDSON, A. L., BORRESEN-DALE, A. L., FUTREAL, P. A., STRATTON, M. R., CAMPBELL, P. J. & BREAST CANCER WORKING GROUP OF THE INTERNATIONAL CANCER GENOME, C. 2012b. The life history of 21 breast cancers. *Cell*, 149, 994-1007.
- NIK-ZAINAL, S., WEDGE, D. C., ALEXANDROV, L. B., PETLJAK, M., BUTLER, A. P., BOLLI, N., DAVIES, H. R., KNAPPSKOG, S., MARTIN, S., PAPAEMMANUIL, E., RAMAKRISHNA, M., SHLIEN, A., SIMONIC, I., XUE, Y., TYLER-SMITH, C., CAMPBELL, P. J. & STRATTON, M. R. 2014. Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nat Genet*, 46, 487-91.
- NILSSON, M., DOMANSKI, H. A., MERTENS, F. & MANDAHL, N. 2004. Molecular cytogenetic characterization of recurrent translocation breakpoints in bizarre parosteal osteochondromatous proliferation (Nora's lesion). *Hum Pathol*, 35, 1063-9.
- NORD, K. H., NILSSON, J., ARBAJIAN, E., VULT VON STEYERN, F., BROSJO, O., CLETON-JANSEN, A. M., SZUHAI, K. & HOGENDOORN, P. C. 2013. Recurrent chromosome 22 deletions in osteoblastoma affect inhibitors of the Wnt/beta-catenin signaling pathway. *PLoS One*, 8, e80725.
- NOWELL, P. & HUNGERFORD, D. 1960. A Minute Chromosome in Human Chronic Granulocytic Leukaemia. *Science*, 132, 1497.
- NOWELL, P. C. 1976. The clonal evolution of tumor cell populations. *Science*, 194, 23-8.
- O'HAGAN, R. C., CHANG, S., MASER, R. S., MOHAN, R., ARTANDI, S. E., CHIN, L. & DEPINHO, R. A. 2002. Telomere dysfunction provokes regional amplification and deletion in cancer genomes. *Cancer Cell*, 2, 149-155.
- O'LEARY, B., CUTTS, R. J., LIU, Y., HREBIEN, S., HUANG, X., FENWICK, K., ANDRE, F., LOIBL, S., LOI, S., GARCIA-MURILLAS, I., CRISTOFANILLI, M., HUANG BARTLETT, C. & TURNER, N. C. 2018a. The Genetic Landscape and Clonal Evolution of Breast Cancer Resistance to Palbociclib plus Fulvestrant in the PALOMA-3 Trial. *Cancer Discov*, 8, 1390-1403.
- O'LEARY, B., HREBIEN, S., MORDEN, J. P., BEANEY, M., FRIBBENS, C., HUANG, X., LIU, Y., BARTLETT, C. H., KOEHLER, M., CRISTOFANILLI, M., GARCIA-MURILLAS, I., BLISS, J. M. & TURNER, N. C. 2018b. Early circulating tumor DNA dynamics and clonal selection with palbociclib and fulvestrant for breast cancer. *Nat Commun*, 9, 896.
- OESPER, L., RITZ, A., AERNI, S. J., DREBIN, R. & RAPHAEL, B. J. 2012. Reconstructing cancer genomes from paired-end sequencing data. *BMC Bioinformatics*, 13 Suppl 6, S10.
- OEY, H., DANIELS, M., RELAN, V., CHEE, T. M., DAVIDSON, M. R., YANG, I. A., ELLIS, J. J., FONG, K. M., KRAUSE, L. & BOWMAN, R. V. 2019. Whole-genome sequencing of human malignant mesothelioma tumours and cell lines. *Carcinogenesis*, 40, 724-734.
- OGURA, K., HOSODA, F., NAKAMURA, H., HAMA, N., TOTOKI, Y., YOSHIDA, A., OHASHI, S., ROKUTAN, H., TAKAI, E., YACHIDA, S., KAWAI, A., TANAKA, S. & SHIBATA, T. 2017. Highly recurrent H3F3A mutations with additional epigenetic regulator alterations in giant cell tumor of bone. *Genes Chromosomes Cancer*, 56, 711-718.

- OKUBO, T., SAITO, T., MITOMI, H., TAKAGI, T., TORIGOE, T., SUEHARA, Y., KANEKO, K. & YAO, T. 2013. p53 mutations may be involved in malignant transformation of giant cell tumor of bone through interaction with GPX1. *Virchows Arch*, 463, 67-77.
- OLINS, D. E. & OLINS, A. L. 2003. Chromatin history: our view from the bridge. *Nature Reviews Molecular Cell Biology*, 4, 809.
- ONS. 2015. *How has life expectancy changed over time?* [Online]. Available: <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/articles/howhaslifeexpectancychangedovertime/2015-09-09> [Accessed 2019/09/20].
- ONS. 2017. *Causes of death of 100 years* [Online]. Available: <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/causesofdeathover100years/2017-09-18> [Accessed 2019/09/19].
- OROSZ, Z. & ATHANASOU, N. A. 2017. Giant Cell-Containing Tumors of Bone. *Surg Pathol Clin*, 10, 553-573.
- OXNARD, G. R., THRESS, K. S., ALDEN, R. S., LAWRENCE, R., PAWELETZ, C. P., CANTARINI, M., YANG, J. C., BARRETT, J. C. & JANNE, P. A. 2016. Association Between Plasma Genotyping and Outcomes of Treatment With Osimertinib (AZD9291) in Advanced Non-Small-Cell Lung Cancer. *J Clin Oncol*, 34, 3375-82.
- PANAGOPOULOS, I., MERTENS, F., LOFVENBERG, R. & MANDAHL, N. 2008. Fusion of the COL1A1 and USP6 genes in a benign bone tumor. *Cancer Genet Cytogenet*, 180, 70-3.
- PANEBIANCO, F., KELLY, L. M., LIU, P., ZHONG, S., DACIC, S., WANG, X., SINGHI, A. D., DHIR, R., CHIOSEA, S. I., KUAN, S.-F., BHARGAVA, R., DABBS, D., TRIVEDI, S., GANDHI, M., DIAZ, R., WALD, A. I., CARTY, S. E., FERRIS, R. L., LEE, A. V., NIKIFOROVA, M. N. & NIKIFOROV, Y. E. 2017. *THADA* fusion is a mechanism of IGF2BP3 activation and IGF1R signaling in thyroid cancer. *Proceedings of the National Academy of Sciences*, 114, 2307.
- PANTOU, D., RIZOU, H., TSAROUHA, H., POULI, A., PAPANASTASIOU, K., STAMATELLOU, M., TRANGAS, T., PANDIS, N. & BARDI, G. 2005. Cytogenetic manifestations of multiple myeloma heterogeneity. *Genes Chromosomes & Cancer*, 42, 44-57.
- PATCH, A. M., CHRISTIE, E. L., ETEMADMOGHADAM, D., GARSEED, D. W., GEORGE, J., FEREDAY, S., NONES, K., COWIN, P., ALSOP, K., BAILEY, P. J., KASSAHN, K. S., NEWELL, F., QUINN, M. C., KAZAKOFF, S., QUEK, K., WILHELM-BENARTZI, C., CURRY, E., LEONG, H. S., AUSTRALIAN OVARIAN CANCER STUDY, G., HAMILTON, A., MILESHKIN, L., AU-YEUNG, G., KENNEDY, C., HUNG, J., CHIEW, Y. E., HARNETT, P., FRIEDLANDER, M., QUINN, M., PYMAN, J., CORDNER, S., O'BRIEN, P., LEDITSCHKE, J., YOUNG, G., STRACHAN, K., WARING, P., AZAR, W., MITCHELL, C., TRAFICANTE, N., HENDLEY, J., THORNE, H., SHACKLETON, M., MILLER, D. K., ARNAU, G. M., TOTHILL, R. W., HOLLOWAY, T. P., SEMPLE, T., HARLIWONG, I., NOURSE, C., NOURBAKHS, E., MANNING, S., IDRISOGLU, S., BRUXNER, T. J., CHRIST, A. N., POUDEL, B., HOLMES, O., ANDERSON, M., LEONARD, C.,

- LONIE, A., HALL, N., WOOD, S., TAYLOR, D. F., XU, Q., FINK, J. L., WADDELL, N., DRAPKIN, R., STRONACH, E., GABRA, H., BROWN, R., JEWELL, A., NAGARAJ, S. H., MARKHAM, E., WILSON, P. J., ELLUL, J., MCNALLY, O., DOYLE, M. A., VEDURURU, R., STEWART, C., LENGYEL, E., PEARSON, J. V., WADDELL, N., DEFAZIO, A., GRIMMOND, S. M. & BOWTELL, D. D. 2015. Whole-genome characterization of chemoresistant ovarian cancer. *Nature*, 521, 489-94.
- PCAWG 2020. Pan-cancer analysis of whole genomes. *Nature*, 578, 82-93.
- PEIFER, M., HERTWIG, F., ROELS, F., DREIDAX, D., GARTLGRUBER, M., MENON, R., KRAMER, A., RONCAIOLI, J. L., SAND, F., HEUCKMANN, J. M., IKRAM, F., SCHMIDT, R., ACKERMANN, S., ENGESSER, A., KAHLERT, Y., VOGEL, W., ALTMULLER, J., NURNBERG, P., THIERRY-MIEG, J., THIERRY-MIEG, D., MARIAPPAN, A., HEYNCK, S., MARIOTTI, E., HENRICH, K. O., GLOECKNER, C., BOSCO, G., LEUSCHNER, I., SCHWEIGER, M. R., SAVELYEVA, L., WATKINS, S. C., SHAO, C., BELL, E., HOFER, T., ACHTER, V., LANG, U., THEISSEN, J., VOLLAND, R., SAADATI, M., EGGERT, A., DE WILDE, B., BERTHOLD, F., PENG, Z., ZHAO, C., SHI, L., ORTMANN, M., BUTTNER, R., PERNER, S., HERO, B., SCHRAMM, A., SCHULTE, J. H., HERRMANN, C., O'SULLIVAN, R. J., WESTERMANN, F., THOMAS, R. K. & FISCHER, M. 2015. Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature*, 526, 700-4.
- PETIT, M. M., SWARTS, S., BRIDGE, J. A. & VAN DE VEN, W. J. 1998. Expression of reciprocal fusion transcripts of the HMGIC and LPP genes in parosteal lipoma. *Cancer Genet Cytogenet*, 106, 18-23.
- PETLJAK, M., ALEXANDROV, L. B., BRAMMELD, J. S., PRICE, S., WEDGE, D. C., GROSSMANN, S., DAWSON, K. J., JU, Y. S., IORIO, F., TUBIO, J. M. C., KOH, C. C., GEORGAKOPOULOS-SOARES, I., RODRIGUEZ-MARTIN, B., OTLU, B., O'MEARA, S., BUTLER, A. P., MENZIES, A., BHOSLE, S. G., RAINE, K., JONES, D. R., TEAGUE, J. W., BEAL, K., LATIMER, C., O'NEILL, L., ZAMORA, J., ANDERSON, E., PATEL, N., MADDISON, M., NG, B. L., GRAHAM, J., GARNETT, M. J., MCDERMOTT, U., NIK-ZAINAL, S., CAMPBELL, P. J. & STRATTON, M. R. 2019. Characterizing Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis. *Cell*, 176, 1282-1294.e20.
- PRESNEAU, N., BAUMHOER, D., BEHJATI, S., PILLAY, N., TARPEY, P., CAMPBELL, P. J., JUNDT, G., HAMOUDI, R., WEDGE, D. C., LOO, P. V., HASSAN, A. B., KHATRI, B., YE, H., TIRABOSCO, R., AMARY, M. F. & FLANAGAN, A. M. 2015. Diagnostic value of H3F3A mutations in giant cell tumour of bone compared to osteoclast-rich mimics. *J Pathol Clin Res*, 1, 113-23.
- PREUD'HOMME, J. L. & SELIGMANN, M. 1972. Surface bound immunoglobulins as a cell marker in human lymphoproliferative diseases. *Blood*, 40, 777-94.
- PRIESTLEY, P., BABER, J., LOLKEMA, M. P., STEEGHS, N., BRUIJN, E. D., DUYVESTYEN, K., HAIDARI, S., HOECK, A. V., ONSTENK, W., ROEPMAN, P., SHALE, C., VODA, M., BLOEMENDAL, H. J., TJANHEIJNEN, V. C. G., VAN HERPEN, C. M. L., LABOTS, M., WITTEVEEN, P. O., SMIT, E. F., SLEIJFER, S., VOEST, E. E. & CUPPEN, E. 2018. Pan-cancer whole genome analyses of metastatic solid tumors. *bioRxiv*, 415133.

- PRIESTLEY, P., BABER, J., LOLKEMA, M. P., STEEGHS, N., DE BRUIJN, E., SHALE, C., DUYVESTYEN, K., HAIDARI, S., VAN HOECK, A., ONSTENK, W., ROEPMAN, P., VODA, M., BLOEMENDAL, H. J., TJAN-HEIJNEN, V. C. G., VAN HERPEN, C. M. L., LABOTS, M., WITTEVEEN, P. O., SMIT, E. F., SLEIJFER, S., VOEST, E. E. & CUPPEN, E. 2019. Pan-cancer whole genome analyses of metastatic solid tumors. *bioRxiv*, 415133.
- PRINS, M. J. D., RUURDA, J. P., VAN DIEST, P. J., VAN HILLEGERSBERG, R. & TEN KATE, F. J. W. 2013. The significance of the HER-2 status in esophageal adenocarcinoma for survival: an immunohistochemical and an in situ hybridization study†. *Annals of Oncology*, 24, 1290-1297.
- QIAN, B. Z. & POLLARD, J. W. 2010. Macrophage diversity enhances tumor progression and metastasis. *Cell*, 141, 39-51.
- RAUSCH, T., ZICHNER, T., SCHLATT, A., STÜTZ, A. M., BENES, V. & KORBEL, J. O. 2012. Delly: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*, 28.
- RIGHI, A., MANCINI, I., GAMBAROTTI, M., PICCI, P., GAMBERI, G., MARRACCINI, C., DEI TOS, A. P., SIMI, L., PINZANI, P. & FRANCHI, A. 2017. Histone 3.3 mutations in giant cell tumor and giant cell-rich sarcomas of bone. *Hum Pathol*, 68, 128-135.
- RIMMER, A., PHAN, H., MATHIESON, I., IQBAL, Z., TWIGG, S. R. F., WILKIE, A. O. M., MCVEAN, G. & LUNTER, G. 2014. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat Genet*, 46, 912-918.
- ROCK, M. G., PRITCHARD, D. J. & UNNI, K. K. 1984. Metastases from histologically benign giant-cell tumor of bone. *J Bone Joint Surg Am*, 66, 269-74.
- RODRIGUEZ-MARTIN, B., ALVAREZ, E. G., BAEZ-ORTEGA, A., DEMEULEMEESTER, J., JU, Y. S., ZAMORA, J., DETERING, H., LI, Y., CONTINO, G., DENTRO, S. C., BRUZOS, A. L., DUESO-BARROSO, A., ARDELJAN, D., TOJO, M., ROBERTS, N. D., BLANCO, M. G., EDWARDS, P. A. W., WEISCHENFELDT, J., SANTAMARINA, M., PUIGGROS, M., CHONG, Z., CHEN, K., LEE, E. A., WALA, J. A., RAINE, K., BUTLER, A., WASZAK, S. M., NAVARRO, F. C. P., SCHUMACHER, S. E., MONLONG, J., MAURA, F., BOLLI, N., BOURQUE, G., GERSTEIN, M., PARK, P. J., BERROUKHIM, R., TORRENTS, D., KORBEL, J. O., MARTINCORENA, I., FITZGERALD, R. C., VAN LOO, P., KAZAZIAN, H. H., BURNS, K. H., CAMPBELL, P. J., TUBIO, J. M. C., & 2017. Pan-cancer analysis of whole genomes reveals driver rearrangements promoted by LINE-1 retrotransposition in human tumours. *bioRxiv*.
- ROKA, S., FIEGL, M., ZOJER, N., FILIPITS, M., SCHUSTER, R., STEINER, B., JAKESZ, R., HUBER, H. & DRACH, J. 1998. Aneuploidy of chromosome 8 as detected by interphase fluorescence in situ hybridization is a recurrent finding in primary and metastatic breast cancer. *Breast Cancer Research and Treatment*, 48, 125-133.
- ROSARIO, M., KIM, H. S., YUN, J. Y. & HAN, I. 2017. Surveillance for lung metastasis from giant cell tumor of bone. *J Surg Oncol*, 116, 907-913.
- ROSENTHAL, R., CADIEUX, E. L., SALGADO, R., BAKIR, M. A., MOORE, D. A., HILEY, C. T., LUND, T., TANIC, M., READING, J. L., JOSHI, K., HENRY, J. Y., GHORANI, E., WILSON, G. A., BIRKBAK, N. J., JAMAL-HANJANI, M.,

- VEERIAH, S., SZALLASI, Z., LOI, S., HELLMANN, M. D., FEBER, A., CHAIN, B., HERRERO, J., QUEZADA, S. A., DEMEULEMEESTER, J., VAN LOO, P., BECK, S., MCGRANAHAN, N. & SWANTON, C. 2019. Neoantigen-directed immune escape in lung cancer evolution. *Nature*, 567, 479-485.
- ROSS, J. S., WANG, K., CHMIELECKI, J., GAY, L., JOHNSON, A., CHUDNOVSKY, J., YELENSKY, R., LIPSON, D., ALI, S. M., ELVIN, J. A., VERGILIO, J. A., ROELS, S., MILLER, V. A., NAKAMURA, B. N., GRAY, A., WONG, M. K. & STEPHENS, P. J. 2016. The distribution of BRAF gene fusions in solid tumors and response to targeted therapy. *International Journal of Cancer*, 138, 881-890.
- ROTH, A., KHATTRA, J., YAP, D., WAN, A., LAKS, E., BIELE, J., HA, G., APARICIO, S., BOUCHARD-CÔTÉ, A. & SHAH, S. P. 2014. PyClone: statistical inference of clonal population structure in cancer. *Nature Methods*, 11, 396.
- ROUS, P. 1910. A TRANSMISSIBLE AVIAN NEOPLASM. (SARCOMA OF THE COMMON FOWL.). *J Exp Med*, 12, 696-705.
- ROUX, S., AMAZIT, L., MEDURI, G., GUIOCHON-MANTEL, A., MILGROM, E. & MARIETTE, X. 2002. RANK (receptor activator of nuclear factor kappa B) and RANK ligand are expressed in giant cell tumors of bone. *Am J Clin Pathol*, 117, 210-6.
- ROWLEY, J. D. 1973. A New Consistent Chromosomal Abnormality in Chronic Myelogenous Leukaemia identified by Quinacrine Fluorescence and Giemsa Staining. *Nature*, 243, 290-293.
- RUDOLPH, K. L., MILLARD, M., BOSENBERG, M. W. & DEPINHO, R. A. 2001. Telomere dysfunction and evolution of intestinal carcinoma in mice and humans. *Nature Genetics*, 28, 155-159.
- RUTKOWSKI, P., FERRARI, S., GRIMER, R. J., STALLEY, P. D., DIJKSTRA, S. P., PIENKOWSKI, A., VAZ, G., WUNDER, J. S., SEEGER, L. L., FENG, A., ROBERTS, Z. J. & BACH, B. A. 2015. Surgical downstaging in an open-label phase II trial of denosumab in patients with giant cell tumor of bone. *Ann Surg Oncol*, 22, 2860-8.
- RYSER, M. D., YU, M., GRADY, W., SIEGMUND, K. & SHIBATA, D. 2018. Epigenetic Heterogeneity in Human Colorectal Tumors Reveals Preferential Conservation And Evidence of Immune Surveillance. *Sci Rep*, 8, 17292.
- SAADA, E., PEOC'H, M., DECOUVELAERE, A. V., COLLARD, O., PEYRON, A. C. & PEDEUTOUR, F. 2011. CCND1 and MET genomic amplification during malignant transformation of a giant cell tumor of bone. *J Clin Oncol*, 29, e86-9.
- SAITO, T., MITOMI, H., IZUMI, H., SUEHARA, Y., OKUBO, T., TORIGOE, T., TAKAGI, T., KANEKO, K., SATO, K., MATSUMOTO, T. & YAO, T. 2011a. A case of secondary malignant giant-cell tumor of bone with p53 mutation after long-term follow-up. *Hum Pathol*, 42, 727-33.
- SAITO, T., MITOMI, H., SUEHARA, Y., OKUBO, T., TORIGOE, T., TAKAGI, T., KANEKO, K. & YAO, T. 2011b. A case of de novo secondary malignant giant-cell tumor of bone with loss of heterozygosity of p53 gene that transformed within a short-term follow-up. *Pathol Res Pract*, 207, 664-9.
- SANERKIN, N. G. 1980. Malignancy, aggressiveness, and recurrence in giant cell tumor of bone. *Cancer*, 46, 1641-1649.
- SAUTER, G., MOCH, H., GASSER, T. C., MIHATSCH, M. J. & WALDMAN, F. M. 1995. HETEROGENEITY OF CHROMOSOME-17 AND ERBB-2 GENE

COPY NUMBER IN PRIMARY AND METASTATIC BLADDER-CANCER.
Cytometry, 21, 40-46.

- SCHWARTZENTRUBER, J., KORSHUNOV, A., LIU, X. Y., JONES, D. T., PFAFF, E., JACOB, K., STURM, D., FONTEBASSO, A. M., QUANG, D. A., TONJES, M., HOVESTADT, V., ALBRECHT, S., KOOL, M., NANTEL, A., KONERMANN, C., LINDROTH, A., JAGER, N., RAUSCH, T., RYZHOVA, M., KORBEL, J. O., HIELSCHER, T., HAUSER, P., GARAMI, M., KLEKNER, A., BOGNAR, L., EBINGER, M., SCHUHMANN, M. U., SCHEURLEN, W., PEKRUN, A., FRUHWALD, M. C., ROGGENDORF, W., KRAMM, C., DURKEN, M., ATKINSON, J., LEPAGE, P., MONTPETIT, A., ZAKRZEWSKA, M., ZAKRZEWSKI, K., LIBERSKI, P. P., DONG, Z., SIEGEL, P., KULOZIK, A. E., ZAPATKA, M., GUHA, A., MALKIN, D., FELSBERG, J., REIFENBERGER, G., VON DEIMLING, A., ICHIMURA, K., COLLINS, V. P., WITT, H., MILDE, T., WITT, O., ZHANG, C., CASTELOBRANCO, P., LICHTER, P., FAURY, D., TABORI, U., PLASS, C., MAJEWSKI, J., PFISTER, S. M. & JABADO, N. 2012. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature*, 482, 226-31.
- SENGUPTA, S., WANG, J., LEE, J., MULLER, P., GULUKOTA, K., BANERJEE, A. & JI, Y. 2015. Bayclone: Bayesian nonparametric inference of tumor subclones using NGS data. *Pac Symp Biocomput*, 467-78.
- SHAH, K. N., BHATT, R., ROTOW, J., ROHRBERG, J., OLIVAS, V., WANG, V. E., HEMMATI, G., MARTINS, M. M., MAYNARD, A., KUHN, J., GALEAS, J., DONNELLA, H. J., KAUSHIK, S., KU, A., DUMONT, S., KRINGS, G., HARINGSMA, H. J., ROBILLARD, L., SIMMONS, A. D., HARDING, T. C., MCCORMICK, F., GOGA, A., BLAKELY, C. M., BIVONA, T. G. & BANDYOPADHYAY, S. 2019. Aurora kinase A drives the evolution of resistance to third-generation EGFR inhibitors in lung cancer. *Nature Medicine*, 25, 111-118.
- SHARMA, A., CAO, E. Y., KUMAR, V., ZHANG, X., LEONG, H. S., WONG, A. M. L., RAMAKRISHNAN, N., HAKIMULLAH, M., TEO, H. M. V., CHONG, F. T., CHIA, S., THANGAVELU, M. T., KWANG, X. L., GUPTA, R., CLARK, J. R., PERIYASAMY, G., IYER, N. G. & DASGUPTA, R. 2018. Longitudinal single-cell RNA sequencing of patient-derived primary cells reveals drug-induced infidelity in stem cell hierarchy. *Nature Communications*, 9, 4931.
- SNYDER, A., MAKAROV, V., MERGHOUB, T., YUAN, J., ZARETSKY, J. M., DESRICHARD, A., WALSH, L. A., POSTOW, M. A., WONG, P., HO, T. S., HOLLMANN, T. J., BRUGGEMAN, C., KANNAN, K., LI, Y., ELIPENAHILI, C., LIU, C., HARBISON, C. T., WANG, L., RIBAS, A., WOLCHOK, J. D. & CHAN, T. A. 2014. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med*, 371, 2189-2199.
- SONG, W., VAN DEN BERG, E., KWEE, T. C., JUTTE, P. C., CLETON-JANSEN, A. M., BOVEE, J. & SUURMEIJER, A. J. 2018. Low-grade central fibroblastic osteosarcoma may be differentiated from its mimicker desmoplastic fibroma by genetic analysis. *Clin Sarcoma Res*, 8, 16.
- SOTTORIVA, A., KANG, H., MA, Z., GRAHAM, T. A., SALOMON, M. P., ZHAO, J., MARJORAM, P., SIEGMUND, K., PRESS, M. F., SHIBATA, D. & CURTIS,

- C. 2015. A Big Bang model of human colorectal tumor growth. *Nat Genet*, 47, 209-16.
- SPINA, V., BRUSCAGGIN, A., CUCCARO, A., MARTINI, M., DI TRANI, M., FORESTIERI, G., MANZONI, M., CONDOLUCI, A., ARRIBAS, A., TERZIDI-BERGAMO, L., LOCATELLI, S. L., CUPELLI, E., CERIANI, L., MOCCIA, A. A., STATHIS, A., NASSI, L., DEAMBROGI, C., DIOP, F., GUIDETTI, F., COCOMAZZI, A., ANNUNZIATA, S., RUFINI, V., GIORDANO, A., NERI, A., BOLDORINI, R., GERBER, B., BERTONI, F., GHIELMINI, M., STUSSI, G., SANTORO, A., CAVALLI, F., ZUCCA, E., LAROCCA, L. M., GAIDANO, G., HOHAUS, S., CARLO-STELLA, C. & ROSSI, D. 2018. Circulating tumor DNA reveals genetics, clonal evolution, and residual disease in classical Hodgkin lymphoma. *Blood*, 131, 2413-2425.
- STEELE, C. D., TARABICHI, M., OUKRIF, D., WEBSTER, A. P., YE, H., FITTALL, M., LOMBARD, P., MARTINCORENA, I., TARPEY, P. S., COLLORD, G., HAASE, K., STRAUSS, S. J., BERISHA, F., VAIKKINEN, H., DHAMI, P., JANSEN, M., BEHJATI, S., AMARY, M. F., TIRABOSCO, R., FEBER, A., CAMPBELL, P. J., ALEXANDROV, L. B., VAN LOO, P., FLANAGAN, A. M. & PILLAY, N. 2019. Undifferentiated Sarcomas Develop through Distinct Evolutionary Pathways. *Cancer Cell*, 35, 441-456.e8.
- STEHELIN, D., VARMUS, H. E., BISHOP, J. M. & VOGT, P. K. 1976. DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature*, 260, 170-3.
- STEPHENS, P. J., GREENMAN, C. D., FU, B., YANG, F., BIGNELL, G. R., MUDIE, L. J., PLEASANCE, E. D., LAU, K. W., BEARE, D., STEBBINGS, L. A., MCLAREN, S., LIN, M. L., MCBRIDE, D. J., VARELA, I., NIK-ZAINAL, S., LEROY, C., JIA, M., MENZIES, A., BUTLER, A. P., TEAGUE, J. W., QUAIL, M. A., BURTON, J., SWERDLOW, H., CARTER, N. P., MORSBERGER, L. A., IACOBUZIO-DONAHUE, C., FOLLOWS, G. A., GREEN, A. R., FLANAGAN, A. M., STRATTON, M. R., FUTREAL, P. A. & CAMPBELL, P. J. 2011. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*, 144, 27-40.
- STORLAZZI, C. T., WOZNIAK, A., PANAGOPOULOS, I., SCIOT, R., MANDAHN, N., MERTENS, F. & DEBIEC-RYCHTER, M. 2006. Rearrangement of the COL12A1 and COL4A5 genes in subungual exostosis: molecular cytogenetic delineation of the tumor-specific translocation t(X;6)(q13-14;q22). *Int J Cancer*, 118, 1972-6.
- STRACHAN, T. & READ, A. P. 2010. *Human molecular genetics*, New York, Garland Science.
- STURM, D., WITT, H., HOVESTADT, V., KHUONG-QUANG, D. A., JONES, D. T., KONERMANN, C., PFAFF, E., TONJES, M., SILL, M., BENDER, S., KOOL, M., ZAPATKA, M., BECKER, N., ZUCKNICK, M., HIELSCHER, T., LIU, X. Y., FONTEBASSO, A. M., RYZHOVA, M., ALBRECHT, S., JACOB, K., WOLTER, M., EBINGER, M., SCHUHMANN, M. U., VAN METER, T., FRUHWALD, M. C., HAUCH, H., PEKRUN, A., RADLWIMMER, B., NIEHUES, T., VON KOMOROWSKI, G., DURKEN, M., KULOZIK, A. E., MADDEN, J., DONSON, A., FOREMAN, N. K., DRISSI, R., FOULADI, M., SCHEURLLEN, W., VON DEIMLING, A., MONORANU, C., ROGGENDORF, W., HEROLD-MENDE, C., UNTERBERG, A., KRAMM, C. M., FELSBURG,

- J., HARTMANN, C., WIESTLER, B., WICK, W., MILDE, T., WITT, O., LINDROTH, A. M., SCHWARTZENTRUBER, J., FAURY, D., FLEMING, A., ZAKRZEWSKA, M., LIBERSKI, P. P., ZAKRZEWSKI, K., HAUSER, P., GARAMI, M., KLEKNER, A., BOGNAR, L., MORRISSY, S., CAVALLI, F., TAYLOR, M. D., VAN SLUIS, P., KOSTER, J., VERSTEEG, R., VOLCKMANN, R., MIKKELSEN, T., ALDAPE, K., REIFENBERGER, G., COLLINS, V. P., MAJEWSKI, J., KORSHUNOV, A., LICHTER, P., PLASS, C., JABADO, N. & PFISTER, S. M. 2012. Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell*, 22, 425-37.
- SUBRAMANIAN, A., TAMAYO, P., MOOTHA, V. K., MUKHERJEE, S., EBERT, B. L., GILLETTE, M. A., PAULOVICH, A., POMEROY, S. L., GOLUB, T. R., LANDER, E. S. & MESIROV, J. P. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102, 15545-15550.
- SUGITA, S., HIRANO, H., KIKUCHI, N., KUBO, T., ASANUMA, H., AOYAMA, T., EMORI, M. & HASEGAWA, T. 2016. Diagnostic utility of FOSB immunohistochemistry in pseudomyogenic hemangioendothelioma and its histological mimics. *Diagn Pathol*, 11, 75.
- SULLIVAN, L. M., FOLPE, A. L., PAWEL, B. R., JUDKINS, A. R. & BIEGEL, J. A. 2013. Epithelioid sarcoma is associated with a high percentage of SMARCB1 deletions. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, 26, 385-392.
- SUTHERLAND, J. A., COOK, A., BANNISTER, A. J. & KOUZARIDES, T. 1992. Conserved motifs in Fos and Jun define a new class of activation domain. *Genes Dev*, 6, 1810-9.
- SWANTON, C., BURRELL, R. A. & FUTREAL, P. A. 2011. Breast cancer genome heterogeneity: a challenge to personalised medicine? *Breast cancer research : BCR*, 13, 104-104.
- SWANTON, C., MCGRANAHAN, N., STARRETT, G. J. & HARRIS, R. S. 2015. APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. *Cancer Discov*, 5, 704-12.
- SZIKRISZT, B., PÓTI, Á., PIPEK, O., KRZYSTANEK, M., KANU, N., MOLNÁR, J., RIBLI, D., SZELTNER, Z., TUSNÁDY, G. E., CSABAI, I., SZALLASI, Z., SWANTON, C. & SZÜTS, D. 2016. A comprehensive survey of the mutagenic impact of common cancer cytotoxics. 17.
- SZUHAI, K., IJSZENGA, M., KNIJNENBURG, J., DIJKSTRA, S., DE SCHEPPER, A., TANKE, H. J. & HOGENDOORN, P. C. 2007. Does parosteal liposarcoma differ from other atypical lipomatous tumors/well-differentiated liposarcomas? A molecular cytogenetic study using combined multicolor COBRA-FISH karyotyping and array-based comparative genomic hybridization. *Cancer Genet Cytogenet*, 176, 115-20.
- TAKEDA, D. Y., SPISÁK, S., SEO, J.-H., BELL, C., O'CONNOR, E., KORTHAUER, K., RIBLI, D., CSABAI, I., SOLYMOSI, N., SZÁLLÁSI, Z., STILLMAN, D. R., CEJAS, P., QIU, X., LONG, H. W., TISZA, V., NUZZO, P. V., ROHANIZADEGAN, M., POMERANTZ, M. M., HAHN, W. C. & FREEDMAN, M. L. 2018. A Somatic Acquired Enhancer of the Androgen

- Receptor Is a Noncoding Driver in Advanced Prostate Cancer. *Cell*, 174, 422-432.e13.
- TARABICHI, M., MARTINCORENA, I., GERSTUNG, M., LEROI, A. M., MARKOWETZ, F., DENTRO, S. C., LESHCHINER, I., GERSTUNG, M., JOLLY, C., HAASE, K., TARABICHI, M., WINTERSINGER, J., DESHWAR, A. G., YU, K., GONZALEZ, S., RUBANOVA, Y., MACINTYRE, G., ADAMS, D. J., ANUR, P., BEROUKHIM, R., BOUTROS, P. C., BOWTELL, D. D., CAMPBELL, P. J., CAO, S., CHRISTIE, E. L., CMERO, M., CUN, Y., DAWSON, K. J., DEMEULEMEESTER, J., DONMEZ, N., DREWS, R. M., EILS, R., FAN, Y., FITTALL, M., GARSEED, D. W., GETZ, G., HA, G., IMIELINSKI, M., JERMAN, L., JI, Y., KLEINHEINZ, K., LEE, J., LEE-SIX, H., LIVITZ, D. G., MALIKIC, S., MARKOWETZ, F., MARTINCORENA, I., MITCHELL, T. J., MUSTONEN, V., OESPER, L., PEIFER, M., PETO, M., RAPHAEL, B. J., ROSEBROCK, D., SAHINALP, S. C., SALCEDO, A., SCHLESNER, M., SCHUMACHER, S., SENGUPTA, S., SHI, R., SHIN, S. J., STEIN, L. D., VÁZQUEZ-GARCÍA, I., VEMBU, S., WHEELER, D. A., YANG, T.-P., YAO, X., YUAN, K., ZHU, H., WANG, W., MORRIS, Q. D., SPELLMAN, P. T., WEDGE, D. C., VAN LOO, P., SPELLMAN, P. T., MORRIS, Q. D., LINGJÆRDE, O. C., WEDGE, D. C., VAN LOO, P., THE, P. E. & HETEROGENEITY WORKING, G. 2018. Neutral tumor evolution? *Nature Genetics*, 50, 1630-1633.
- TARPEY, P. S., BEHJATI, S., COOKE, S. L., VAN LOO, P., WEDGE, D. C., PILLAY, N., MARSHALL, J., O'MEARA, S., DAVIES, H., NIK-ZAINAL, S., BEARE, D., BUTLER, A., GAMBLE, J., HARDY, C., HINTON, J., JIA, M. M., JAYAKUMAR, A., JONES, D., LATIMER, C., MADDISON, M., MARTIN, S., MCLAREN, S., MENZIES, A., MUDIE, L., RAINE, K., TEAGUE, J. W., TUBIO, J. M., HALAI, D., TIRABOSCO, R., AMARY, F., CAMPBELL, P. J., STRATTON, M. R., FLANAGAN, A. M. & FUTREAL, P. A. 2013. Frequent mutation of the major cartilage collagen gene COL2A1 in chondrosarcoma. *Nat Genet*, 45, 923-6.
- TARPEY, P. S., BEHJATI, S., YOUNG, M. D., MARTINCORENA, I., ALEXANDROV, L. B., FARNDON, S. J., GUZZO, C., HARDY, C., LATIMER, C., BUTLER, A. P., TEAGUE, J. W., SHLIEN, A., FUTREAL, P. A., SHAH, S., BASHASHATI, A., JAMSHIDI, F., NIELSEN, T. O., HUNTSMAN, D., BAUMHOER, D., BRANDNER, S., WUNDER, J., DICKSON, B., COGSWELL, P., SOMMER, J., PHILLIPS, J. J., AMARY, M. F., TIRABOSCO, R., PILLAY, N., YIP, S., STRATTON, M. R., FLANAGAN, A. M. & CAMPBELL, P. J. 2017. The driver landscape of sporadic chordoma. *Nat Commun*, 8, 890.
- TAYLOR, B. J., NIK-ZAINAL, S., WU, Y. L., STEBBINGS, L. A., RAINE, K., CAMPBELL, P. J., RADA, C., STRATTON, M. R. & NEUBERGER, M. S. 2013. DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. *Elife*, 2, e00534.
- TAYLOR, R. M., KASHIMA, T. G., FERGUSON, D. J., SZUHAI, K., HOGENDOORN, P. C. & ATHANASOU, N. A. 2012. Analysis of stromal cells in osteofibrous dysplasia and adamantinoma of long bones. *Mod Pathol*, 25, 56-64.
- TESCHENDORFF, A. E., BREEZE, C. E., ZHENG, S. C. & BECK, S. 2017. A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies. *BMC Bioinformatics*, 18, 105.

- THIERRY, A. R., MOULIERE, F., GONGORA, C., OLLIER, J., ROBERT, B., YCHOU, M., DEL RIO, M. & MOLINA, F. 2010. Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts. *Nucleic Acids Res*, 38, 6159-75.
- THOMAS, D., HENSHAW, R., SKUBITZ, K., CHAWLA, S., STADDON, A., BLAY, J. Y., ROUDIER, M., SMITH, J., YE, Z., SOHN, W., DANSEY, R. & JUN, S. 2010. Denosumab in patients with giant-cell tumour of bone: an open-label, phase 2 study. *Lancet Oncol*, 11, 275-80.
- THWAY, K., JONES, R. L., NOUJAIM, J. & FISHER, C. 2016. Epithelioid Sarcoma: Diagnostic Features and Genetics. *Adv Anat Pathol*, 23, 41-9.
- TIGHE, J. E., DAGA, A. & CALABI, F. 1993. Translocation breakpoints are clustered on both chromosome 8 and chromosome 21 in the t(8;21) of acute myeloid leukemia. *Blood*, 81, 592-6.
- TSUKAMOTO, S., RIGHI, A., VANEL, D., HONOKI, K., DONATI, D. M. & ERRANI, C. 2017. Development of high-grade osteosarcoma in a patient with recurrent giant cell tumor of the ischium while receiving treatment with denosumab. *Jpn J Clin Oncol*, 47, 1090-1096.
- TSUKAMOTO, Y., FUTANI, H., KIHARA, T., WATANABE, T., KUMANISHI, S., MATSUO, S., HIROTA, S., UEDA, T., YAMAMOTO, H. & YOSHIYA, S. 2018. An extremely rare case of primary malignancy in giant cell tumor of bone, arising in the right femur and harboring H3F3A mutation. *Pathol Res Pract*, 214, 1504-1509.
- TURAJLIC, S., XU, H., LITCHFIELD, K., ROWAN, A., CHAMBERS, T., LOPEZ, J. I., NICOL, D., O'BRIEN, T., LARKIN, J., HORSWELL, S., STARES, M., AU, L., JAMAL-HANJANI, M., CHALLACOMBE, B., CHANDRA, A., HAZELL, S., EICHLER-JONSSON, C., SOULTATI, A., CHOWDHURY, S., RUDMAN, S., LYNCH, J., FERNANDO, A., STAMP, G., NYE, E., JABBAR, F., SPAIN, L., LALL, S., GUARCH, R., FALZON, M., PROCTOR, I., PICKERING, L., GORE, M., WATKINS, T. B. K., WARD, S., STEWART, A., DINATALE, R., BECERRA, M. F., REZNIK, E., HSIEH, J. J., RICHMOND, T. A., MAYHEW, G. F., HILL, S. M., MCNALLY, C. D., JONES, C., ROSENBAUM, H., STANISLAW, S., BURGESS, D. L., ALEXANDER, N. R., SWANTON, C., PEACE & CONSORTIUM, T. R. R. 2018a. Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal. *Cell*, 173, 581-594 e12.
- TURAJLIC, S., XU, H., LITCHFIELD, K., ROWAN, A., HORSWELL, S., CHAMBERS, T., O'BRIEN, T., LOPEZ, J. I., WATKINS, T. B. K., NICOL, D., STARES, M., CHALLACOMBE, B., HAZELL, S., CHANDRA, A., MITCHELL, T. J., AU, L., EICHLER-JONSSON, C., JABBAR, F., SOULTATI, A., CHOWDHURY, S., RUDMAN, S., LYNCH, J., FERNANDO, A., STAMP, G., NYE, E., STEWART, A., XING, W., SMITH, J. C., ESCUDERO, M., HUFFMAN, A., MATTHEWS, N., ELGAR, G., PHILLIMORE, B., COSTA, M., BEGUM, S., WARD, S., SALM, M., BOEING, S., FISHER, R., SPAIN, L., NAVAS, C., GRONROOS, E., HOBOR, S., SHARMA, S., AURANGZEB, I., LALL, S., POLSON, A., VARIA, M., HORSFIELD, C., FOTIADIS, N., PICKERING, L., SCHWARZ, R. F., SILVA, B., HERRERO, J., LUSCOMBE, N. M., JAMAL-HANJANI, M., ROSENTHAL, R., BIRKBAK, N. J., WILSON, G. A., PIPEK, O., RIBLI, D., KRZYSTANEK, M., CSABAI, I., SZALLASI, Z., GORE, M., MCGRANAHAN, N., VAN LOO, P., CAMPBELL, P., LARKIN, J.,

- SWANTON, C. & CONSORTIUM, T. R. R. 2018b. Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal. *Cell*, 173, 595-610 e11.
- TURNBULL, C. 2018. Introducing whole-genome sequencing into routine cancer care: the Genomics England 100 000 Genomes Project. *Annals of Oncology*, 29, 784-787.
- VALENZUELA, D. M. & GROFFEN, J. 1986. Four human carcinoma cell lines with novel mutations in position 12 of c-K-ras oncogene. *Nucleic Acids Res*, 14, 843-52.
- VAN BEVEREN, C., VAN STRAATEN, F., CURRAN, T., MULLER, R. & VERMA, I. M. 1983. Analysis of FBJ-MuSV provirus and c-fos (mouse) gene reveals that viral and cellular fos gene products have different carboxy termini. *Cell*, 32, 1241-55.
- VAN DEN EYNDEN, J., JIMÉNEZ-SÁNCHEZ, A., MILLER, M. L. & LARSSON, E. 2019. Lack of detectable neoantigen depletion signals in the untreated cancer genome. *Nature Genetics*, 51, 1741-1748.
- VAN DER HEIJDEN, L., DIJKSTRA, P. D. S., BLAY, J. Y. & GELDERBLOM, H. 2017. Giant cell tumour of bone in the denosumab era. *Eur J Cancer*, 77, 75-83.
- VAN DER POEL, H. G., OOSTERHOF, G. O. N., SCHAAFSMA, H. E., DEBRUYNE, F. M. J. & SCHALKEN, J. A. 1997. Intratumoral nuclear morphologic heterogeneity in prostate cancer. *Urology*, 49, 652-657.
- VAN IJZENDOORN, D. G., DE JONG, D., ROMAGOSA, C., PICCI, P., BENASSI, M. S., GAMBAROTTI, M., DAUGAARD, S., VAN DE SANDE, M., SZUHAI, K. & BOVEE, J. V. 2015. Fusion events lead to truncation of FOS in epithelioid hemangioma of bone. *Genes Chromosomes Cancer*, 54, 565-74.
- VAN IJZENDOORN, D. G. P. & BOVEE, J. 2017. Vascular Tumors of Bone: The Evolvement of a Classification Based on Molecular Developments. *Surg Pathol Clin*, 10, 621-635.
- VAN IJZENDOORN, D. G. P., FORGHANY, Z., LIEBELT, F., VERTEGAAL, A. C., JOCHEMSEN, A. G., BOVEE, J., SZUHAI, K. & BAKER, D. A. 2017. Functional Analyses of a Human Vascular Tumor FOS Variant Identify a Novel Degradation Mechanism and a link to Tumorigenesis. *J Biol Chem*, 292, 21282-21290.
- VAN LOO, P., NORDGARD, S. H., LINGJAERDE, O. C., RUSSNES, H. G., RYE, I. H., SUN, W., WEIGMAN, V. J., MARYNEN, P., ZETTERBERG, A., NAUME, B., PEROU, C. M., BORRESEN-DALE, A. L. & KRISTENSEN, V. N. 2010. Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci U S A*, 107, 16910-5.
- VAN LOO, P. & VOET, T. 2014. Single cell analysis of cancer genomes. *Curr Opin Genet Dev*, 24, 82-91.
- VAN STRAATEN, F., MULLER, R., CURRAN, T., VAN BEVEREN, C. & VERMA, I. M. 1983. Complete nucleotide sequence of a human c-onc gene: deduced amino acid sequence of the human c-fos protein. *Proc Natl Acad Sci U S A*, 80, 3183-7.
- VERELST, S. J., HANS, J., HANSELMANN, R. G. & WIRBEL, R. J. 2004. Genetic instability in primary leiomyosarcoma of bone. *Hum Pathol*, 35, 1404-12.
- VERSCHOOR, A. J., BOVEE, J., MASTBOOM, M. J. L., SANDER DIJKSTRA, P. D., VAN DE SANDE, M. A. J. & GELDERBLOM, H. 2018. Incidence and

- demographics of giant cell tumor of bone in The Netherlands: First nationwide Pathology Registry Study. *Acta Orthop*, 89, 570-574.
- VINAGRE, J., ALMEIDA, A., PÓPULO, H., BATISTA, R., LYRA, J., PINTO, V., COELHO, R., CELESTINO, R., PRAZERES, H., LIMA, L., MELO, M., ROCHA, A. G. D., PRETO, A., CASTRO, P., CASTRO, L., PARDAL, F., LOPES, J. M., SANTOS, L. L., REIS, R. M., CAMESELLE-TEIJEIRO, J., SOBRINHO-SIMÕES, M., LIMA, J., MÁXIMO, V. & SOARES, P. 2013. Frequency of TERT promoter mutations in human cancers. *Nature Communications*, 4, 2185.
- VISWANATHAN, S. R., HA, G., HOFF, A. M., WALA, J. A., CARROT-ZHANG, J., WHELAN, C. W., HARADHVALA, N. J., FREEMAN, S. S., REED, S. C., RHOADES, J., POLAK, P., CIPICCHIO, M., WANKOWICZ, S. A., WONG, A., KAMATH, T., ZHANG, Z., GYDUSH, G. J., ROTEM, D., LOVE, J. C., GETZ, G., GABRIEL, S., ZHANG, C.-Z., DEHM, S. M., NELSON, P. S., VAN ALLEN, E. M., CHOUDHURY, A. D., ADALSTEINSSON, V. A., BEROUKHIM, R., TAPLIN, M.-E. & MEYERSON, M. 2018. Structural Alterations Driving Castration-Resistant Prostate Cancer Revealed by Linked-Read Genome Sequencing. *Cell*, 174, 433-447.e19.
- VOET, T., KUMAR, P., VAN LOO, P., COOKE, S. L., MARSHALL, J., LIN, M. L., ZAMANI ESTEKI, M., VAN DER AA, N., MATEIU, L., MCBRIDE, D. J., BIGNELL, G. R., MCLAREN, S., TEAGUE, J., BUTLER, A., RAINE, K., STEBBINGS, L. A., QUAIL, M. A., D'HOOGHE, T., MOREAU, Y., FUTREAL, P. A., STRATTON, M. R., VERMEESCH, J. R. & CAMPBELL, P. J. 2013. Single-cell paired-end genome sequencing reveals structural variation per cell cycle. *Nucleic Acids Res*, 41, 6119-38.
- VOON, H. P. J., UDUGAMA, M., LIN, W., HII, L., LAW, R. H. P., STEER, D. L., DAS, P. P., MANN, J. R. & WONG, L. H. 2018. Inhibition of a K9/K36 demethylase by an H3.3 point mutation found in paediatric glioblastoma. *Nature Communications*, 9, 3142.
- WAGLE, N., VAN ALLEN, E. M., TREACY, D. J., FREDERICK, D. T., COOPER, Z. A., TAYLOR-WEINER, A., ROSENBERG, M., GOETZ, E. M., SULLIVAN, R. J., FARLOW, D. N., FRIEDRICH, D. C., ANDERKA, K., PERRIN, D., JOHANNESSEN, C. M., MCKENNA, A., CIBULSKIS, K., KRYUKOV, G., HODIS, E., LAWRENCE, D. P., FISHER, S., GETZ, G., GABRIEL, S. B., CARTER, S. L., FLAHERTY, K. T., WARGO, J. A. & GARRAWAY, L. A. 2014. MAP kinase pathway alterations in BRAF-mutant melanoma patients with acquired resistance to combined RAF/MEK inhibition. *Cancer Discov*, 4, 61-8.
- WALA, J. A., BANDOPADHAYAY, P., GREENWALD, N. F., O'ROURKE, R., SHARPE, T., STEWART, C., SCHUMACHER, S., LI, Y., WEISCHENFELDT, J., YAO, X., NUSBAUM, C., CAMPBELL, P., GETZ, G., MEYERSON, M., ZHANG, C. Z., IMIELINSKI, M. & BEROUKHIM, R. 2018. SvABA: genome-wide detection of structural variants and indels by local assembly. *Genome Res*, 28, 581-591.
- WALTHER, C., TAYEBWA, J., LILLJEBJORN, H., MAGNUSSON, L., NILSSON, J., VON STEYERN, F. V., ORA, I., DOMANSKI, H. A., FIORETOS, T., NORD, K. H., FLETCHER, C. D. & MERTENS, F. 2014. A novel SERPINE1-FOSB fusion gene results in transcriptional up-regulation of FOSB in pseudomyogenic haemangioendothelioma. *J Pathol*, 232, 534-40.

- WANG, B., CHEN, W., XIE, X., TU, J., HUANG, G., ZOU, C., YIN, J., WEN, L. & SHEN, J. 2017. Development and validation of a prognostic index to predict pulmonary metastasis of giant cell tumor of bone. *Oncotarget*, 8, 108054-108063.
- WANG, L., MOTOI, T., KHANIN, R., OLSHEN, A., MERTENS, F., BRIDGE, J., DAL CIN, P., ANTONESCU, C. R., SINGER, S., HAMEED, M., BOVEE, J. V., HOGENDOORN, P. C., SOCCI, N. & LADANYI, M. 2012. Identification of a novel, recurrent HEY1-NCOA2 fusion in mesenchymal chondrosarcoma based on a genome-wide screen of exon-level expression data. *Genes Chromosomes Cancer*, 51, 127-39.
- WANG, Y. & NAVIN, N. E. 2015. Advances and applications of single-cell sequencing technologies. *Mol Cell*, 58, 598-609.
- WANG, Y., WATERS, J., LEUNG, M. L., UNRUH, A., ROH, W., SHI, X., CHEN, K., SCHEET, P., VATTATHIL, S., LIANG, H., MULTANI, A., ZHANG, H., ZHAO, R., MICHOR, F., MERIC-BERNSTAM, F. & NAVIN, N. E. 2014. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature*.
- WATSON, J. D. & CRICK, F. H. C. 1953. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, 171, 737-738.
- WATSON, P. A., ARORA, V. K. & SAWYERS, C. L. 2015. Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nat Rev Cancer*, 15, 701-11.
- WEDGE, D. C., GUNDEM, G., MITCHELL, T., WOODCOCK, D. J., MARTINCORENA, I., GHORI, M., ZAMORA, J., BUTLER, A., WHITAKER, H., KOTE-JARAI, Z., ALEXANDROV, L. B., VAN LOO, P., MASSIE, C. E., DENTRO, S., WARREN, A. Y., VERRILL, C., BERNEY, D. M., DENNIS, N., MERSON, S., HAWKINS, S., HOWAT, W., LU, Y.-J., LAMBERT, A., KAY, J., KREMEYER, B., KARASZI, K., LUXTON, H., CAMACHO, N., MARSDEN, L., EDWARDS, S., MATTHEWS, L., BO, V., LEONGAMORNERT, D., MCLAREN, S., NG, A., YU, Y., ZHANG, H., DADAEV, T., THOMAS, S., EASTON, D. F., AHMED, M., BANCROFT, E., FISHER, C., LIVNI, N., NICOL, D., TAVARÉ, S., GILL, P., GREENMAN, C., KHOO, V., VAN AS, N., KUMAR, P., OGDEN, C., CAHILL, D., THOMPSON, A., MAYER, E., ROWE, E., DUDDERIDGE, T., GNANAPRAGASAM, V., SHAH, N. C., RAINE, K., JONES, D., MENZIES, A., STEBBINGS, L., TEAGUE, J., HAZELL, S., CORBISHLEY, C., DE BONO, J., ATTARD, G., ISAACS, W., VISAKORPI, T., FRASER, M., BOUTROS, P. C., BRISTOW, R. G., WORKMAN, P., SANDER, C., HAMDY, F. C., FUTREAL, A., MCDERMOTT, U., AL-LAZIKANI, B., LYNCH, A. G., BOVA, G. S., FOSTER, C. S., BREWER, D. S., NEAL, D. E., COOPER, C. S. & EELES, R. A. 2018. Sequencing of prostate cancers identifies new cancer genes, routes of progression and drug targets. *Nature Genetics*, 50, 682-692.
- WEEKES, D., KASHIMA, T. G., ZANDUETA, C., PERURENA, N., THOMAS, D. P., SUNTERS, A., VUILLIER, C., BOZEC, A., EL-EMIR, E., MILETICH, I., PATINO-GARCIA, A., LECANDA, F. & GRIGORIADIS, A. E. 2016. Regulation of osteosarcoma cell lung metastasis by the c-Fos/AP-1 target FGFR1. *Oncogene*, 35, 2948.
- WEIGELT, B., COMINO-MENDEZ, I., DE BRUIJN, I., TIAN, L., MEISEL, J. L., GARCIA-MURILLAS, I., FRIBBENS, C., CUTTS, R., MARTELOTTO, L. G., NG, C. K. Y., LIM, R. S., SELENICA, P., PISCUOGLIO, S., AGHAJANIAN,

- C., NORTON, L., MURALI, R., HYMAN, D. M., BORSU, L., ARCILA, M. E., KONNER, J., REIS-FILHO, J. S., GREENBERG, R. A., ROBSON, M. E. & TURNER, N. C. 2017. Diverse BRCA1 and BRCA2 Reversion Mutations in Circulating Cell-Free DNA of Therapy-Resistant Breast or Ovarian Cancer. *Clin Cancer Res*, 23, 6708-6720.
- WEINBERG, R. A. 2014. *The biology of cancer / Robert A. Weinberg*, New York : London, New York : London : Garland Science.
- WEINREB, C., OESPER, L. & RAPHAEL, B. J. 2014. Open adjacencies and k-breaks: detecting simultaneous rearrangements in cancer genomes. *BMC Genomics*, 15, S4.
- WEISCHENFELDT, J., DUBASH, T., DRAINAS, A. P., MARDIN, B. R., CHEN, Y., STÜTZ, A. M., WASZAK, S. M., BOSCO, G., HALVORSEN, A. R., RAEDER, B., EFTHYMIPOULOS, T., ERKEK, S., SIEGL, C., BRENNER, H., BRUSTUGUN, O. T., DIETER, S. M., NORTHCOTT, P. A., PETERSEN, I., PFISTER, S. M., SCHNEIDER, M., SOLBERG, S. K., THUNISSEN, E., WEICHERT, W., ZICHER, T., THOMAS, R., PEIFER, M., HELLAND, A., BALL, C. R., JECHLINGER, M., SOTILLO, R., GLIMM, H. & KORBEL, J. O. 2016. Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nature Genetics*, 49, 65.
- WEISS, L. M., WARNKE, R. A., SKLAR, J. & CLEARLY, M. L. 1987. Molecular analysis of the t(14;18) chromosomal translocation in malignant lymphomas. *N Engl J Med*, 317, 1185-9.
- WELCH, J. S., LEY, T. J., LINK, D. C., MILLER, C. A., LARSON, D. E., KOBOLDT, D. C., WARTMAN, L. D., LAMPRECHT, T. L., LIU, F., XIA, J., KANDOTH, C., FULTON, R. S., MCLELLAN, M. D., DOOLING, D. J., WALLIS, J. W., CHEN, K., HARRIS, C. C., SCHMIDT, H. K., KALICKI-VEIZER, J. M., LU, C., ZHANG, Q., LIN, L., O'LAUGHLIN, M. D., MCMICHAEL, J. F., DELEHAUNTY, K. D., FULTON, L. A., MAGRINI, V. J., MCGRATH, S. D., DEMETER, R. T., VICKERY, T. L., HUNDAL, J., COOK, L. L., SWIFT, G. W., REED, J. P., ALLDREDGE, P. A., WYLIE, T. N., WALKER, J. R., WATSON, M. A., HEATH, S. E., SHANNON, W. D., VARGHESE, N., NAGARAJAN, R., PAYTON, J. E., BATY, J. D., KULKARNI, S., KLCO, J. M., TOMASSON, M. H., WESTERVELT, P., WALTER, M. J., GRAUBERT, T. A., DIPERSIO, J. F., DING, L., MARDIS, E. R. & WILSON, R. K. 2012. The origin and evolution of mutations in acute myeloid leukemia. *Cell*, 150, 264-78.
- WILKINS, M. H. F., STOKES, A. R. & WILSON, H. R. 1953. Molecular Structure of Nucleic Acids: Molecular Structure of Deoxyribose Nucleic Acids. *Nature*, 171, 738-740.
- WILLIAMS, M. J., WERNER, B., BARNES, C. P., GRAHAM, T. A. & SOTTORIVA, A. 2016. Identification of neutral tumor evolution across cancer types. *Nat Genet*, 48, 238-44.
- WU, G., BRONISCHER, A., MCEACHRON, T. A., LU, C., PAUGH, B. S., BECKSFORT, J., QU, C., DING, L., HUETHER, R., PARKER, M., ZHANG, J., GAJJAR, A., DYER, M. A., MULLIGHAN, C. G., GILBERTSON, R. J., MARDIS, E. R., WILSON, R. K., DOWNING, J. R., ELLISON, D. W., ZHANG, J. & BAKER, S. J. 2012. Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nat Genet*, 44, 251-3.

- WU, S., POWERS, S., ZHU, W. & HANNUN, Y. A. 2016. Substantial contribution of extrinsic risk factors to cancer development. *Nature*, 529, 43-47.
- WYLIE, A. A., SCHOEPFER, J., JAHNKE, W., COWAN-JACOB, S. W., LOO, A., FURET, P., MARZINZIK, A. L., PELLE, X., DONOVAN, J., ZHU, W., BUONAMICI, S., HASSAN, A. Q., LOMBARDO, F., IYER, V., PALMER, M., BERELLINI, G., DODD, S., THOHAN, S., BITTER, H., BRANFORD, S., ROSS, D. M., HUGHES, T. P., PETRUZZELLI, L., VANASSE, K. G., WARMUTH, M., HOFMANN, F., KEEN, N. J. & SELLERS, W. R. 2017. The allosteric inhibitor ABL001 enables dual targeting of BCR-ABL1. *Nature*, 543, 733.
- XIE, M., LU, C., WANG, J., MCLELLAN, M. D., JOHNSON, K. J., WENDL, M. C., MCMICHAEL, J. F., SCHMIDT, H. K., YELLAPANTULA, V., MILLER, C. A., OZENBERGER, B. A., WELCH, J. S., LINK, D. C., WALTER, M. J., MARDIS, E. R., DIPERSIO, J. F., CHEN, F., WILSON, R. K., LEY, T. J. & DING, L. 2014. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med*, 20, 1472-8.
- YACHIDA, S., JONES, S., BOZIC, I., ANTAL, T., LEARY, R., FU, B., KAMIYAMA, M., HRUBAN, R. H., ESHLEMAN, J. R., NOWAK, M. A., VELCULESCU, V. E., KINZLER, K. W., VOGELSTEIN, B. & IACOBUZIO-DONAHUE, C. A. 2010. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature*, 467, 1114-1117.
- YALCINKAYA, U., DOGANAVSARGIL, B., SEZAK, M., KECECI, B., ARGIN, M., BASDEMIR, G. & OZTOP, F. 2014. Clinical and morphological characteristics of osteoid osteoma and osteoblastoma: a retrospective single-center analysis of 204 patients. *Ann Diagn Pathol*, 18, 319-25.
- YAMAMOTO, H., IWASAKI, T., YAMADA, Y., MATSUMOTO, Y., OTSUKA, H., YOSHIMOTO, M., KOHASHI, K., TAGUCHI, K., YOKOYAMA, R., NAKASHIMA, Y. & ODA, Y. 2018. Diagnostic utility of histone H3.3 G34W, G34R, and G34V mutant-specific antibodies for giant cell tumors of bone. *Hum Pathol*, 73, 41-50.
- YAMAZAKI, F., NAKATANI, F., ASANO, N., WAKAI, S., SEKIMIZU, M., MITANI, S., KUBO, T., KAWAI, A., ICHIKAWA, H. & YOSHIDA, A. 2019. Novel NTRK3 Fusions in Fibrosarcomas of Adults. *Am J Surg Pathol*, 43, 523-530.
- YATES, L. R., KNAPPSKOG, S., WEDGE, D., FARMERY, J. H. R., GONZALEZ, S., MARTINCORENA, I., ALEXANDROV, L. B., VAN LOO, P., HAUGLAND, H. K., LILLENG, P. K., GUNDEM, G., GERSTUNG, M., PAPPAMEMMANUIL, E., GAZINSKA, P., BHOSLE, S. G., JONES, D., RAINE, K., MUDIE, L., LATIMER, C., SAWYER, E., DESMEDT, C., SOTIRIOU, C., STRATTON, M. R., SIEUWERTS, A. M., LYNCH, A. G., MARTENS, J. W., RICHARDSON, A. L., TUTT, A., LØNNING, P. E. & CAMPBELL, P. J. 2017. Genomic Evolution of Breast Cancer Metastasis and Relapse. *Cancer Cell*, 32, 169-184.e7.
- YE, K., SCHULZ, M. H., LONG, Q., APWEILER, R. & NING, Z. 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*, 25, 2865-71.
- YOKOYAMA, A., KAKIUCHI, N., YOSHIZATO, T., NANNYA, Y., SUZUKI, H., TAKEUCHI, Y., SHIOZAWA, Y., SATO, Y., AOKI, K., KIM, S. K., FUJII, Y., YOSHIDA, K., KATAOKA, K., NAKAGAWA, M. M., INOUE, Y., HIRANO, T., SHIRAISHI, Y., CHIBA, K., TANAKA, H., SANADA, M., NISHIKAWA,

- Y., AMANUMA, Y., OHASHI, S., AOYAMA, I., HORIMATSU, T., MIYAMOTO, S. I., TSUNODA, S., SAKAI, Y., NARAHARA, M., BROWN, J. B., SATO, Y., SAWADA, G., MIMORI, K., MINAMIGUCHI, S., HAGA, H., SENO, H., MIYANO, S., MAKISHIMA, H., MUTO, M. & OGAWA, S. 2019. Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature*, 565, 312-317.
- YOSHIDA, K.-I., NAKANO, Y., HONDA-KITAHARA, M., WAKAI, S., MOTOI, T., OGURA, K., SANO, N., SHIBATA, T., OKUMA, T., IWATA, S., KAWAI, A., ICHIMURA, K. & YOSHIDA, A. 2019. Absence of H3F3A mutation in a subset of malignant giant cell tumor of bone. *Modern Pathology*.
- YUAN, K., MACINTYRE, G., LIU, W. & MARKOWETZ, F. 2018. Ccube: A fast and robust method for estimating cancer cell fractions. *bioRxiv*, 484402.
- ZAHN, H., STEIF, A., LAKS, E., EIREW, P., VANINSBERGHE, M., SHAH, S. P., APARICIO, S. & HANSEN, C. L. 2017. Scalable whole-genome single-cell library preparation without preamplification. *Nat Methods*, 14, 167-173.
- ZEHIR, A., BENAYED, R., SHAH, R. H., SYED, A., MIDDHA, S., KIM, H. R., SRINIVASAN, P., GAO, J., CHAKRAVARTY, D., DEVLIN, S. M., HELLMANN, M. D., BARRON, D. A., SCHRAM, A. M., HAMEED, M., DOGAN, S., ROSS, D. S., HECHTMAN, J. F., DELAIR, D. F., YAO, J., MANDELKER, D. L., CHENG, D. T., CHANDRAMOHAN, R., MOHANTY, A. S., PTASHKIN, R. N., JAYAKUMARAN, G., PRASAD, M., SYED, M. H., REMA, A. B., LIU, Z. Y., NAFA, K., BORSU, L., SADOWSKA, J., CASANOVA, J., BACARES, R., KIECKA, I. J., RAZUMOVA, A., SON, J. B., STEWART, L., BALDI, T., MULLANEY, K. A., AL-AHMADIE, H., VAKIANI, E., ABESHOUSE, A. A., PENSON, A. V., JONSSON, P., CAMACHO, N., CHANG, M. T., WON, H. H., GROSS, B. E., KUNDRA, R., HEINS, Z. J., CHEN, H. W., PHILLIPS, S., ZHANG, H., WANG, J., OCHOA, A., WILLS, J., EUBANK, M., THOMAS, S. B., GARDOS, S. M., REALES, D. N., GALLE, J., DURANY, R., CAMBRIA, R., ABIDA, W., CERCEK, A., FELDMAN, D. R., GOUNDER, M. M., HAKIMI, A. A., HARDING, J. J., IYER, G., JANJIGIAN, Y. Y., JORDAN, E. J., KELLY, C. M., LOWERY, M. A., MORRIS, L. G. T., OMURO, A. M., RAJ, N., RAZAVI, P., SHOUSHARI, A. N., SHUKLA, N., SOUMERAI, T. E., VARGHESE, A. M., YAEGER, R., COLEMAN, J., BOCHNER, B., RIELY, G. J., SALTZ, L. B., SCHER, H. I., SABBATINI, P. J., ROBSON, M. E., KLIMSTRA, D. S., TAYLOR, B. S., BASELGA, J., SCHULTZ, N., HYMAN, D. M., ARCILA, M. E., SOLIT, D. B., LADANYI, M. & BERGER, M. F. 2017. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med*, 23, 703-713.
- ZHANG, A. W., MCPHERSON, A., MILNE, K., KROEGER, D. R., HAMILTON, P. T., MIRANDA, A., FUNNELL, T., LITTLE, N., DE SOUZA, C. P. E., LAAN, S., LEDOUX, S., COCHRANE, D. R., LIM, J. L. P., YANG, W., ROTH, A., SMITH, M. A., HO, J., TSE, K., ZENG, T., SHLAFMAN, I., MAYO, M. R., MOORE, R., FAILMEZGER, H., HEINDL, A., WANG, Y. K., BASHASHATI, A., GREWAL, D. S., BROWN, S. D., LAI, D., WAN, A. N. C., NIELSEN, C. B., HUEBNER, C., TESSIER-CLOUTIER, B., ANGLÉSIO, M. S., BOUCHARD-COTE, A., YUAN, Y., WASSERMAN, W. W., GILKS, C. B., KARNEZIS, A. N., APARICIO, S., MCALPINE, J. N., HUNTSMAN, D. G., HOLT, R. A.,

- NELSON, B. H. & SHAH, S. P. 2018a. Interfaces of Malignant and Immunologic Clonal Dynamics in Ovarian Cancer. *Cell*, 173, 1755-1769.e22.
- ZHANG, A. W., MCPHERSON, A., MILNE, K., KROEGER, D. R., HAMILTON, P. T., MIRANDA, A., FUNNELL, T., LITTLE, N., DE SOUZA, C. P. E., LAAN, S., LEDOUX, S., COCHRANE, D. R., LIM, J. L. P., YANG, W., ROTH, A., SMITH, M. A., HO, J., TSE, K., ZENG, T., SHLAFMAN, I., MAYO, M. R., MOORE, R., FAILMEZGER, H., HEINDL, A., WANG, Y. K., BASHASHATI, A., GREWAL, D. S., BROWN, S. D., LAI, D., WAN, A. N. C., NIELSEN, C. B., HUEBNER, C., TESSIER-CLOUTIER, B., ANGLÉSIO, M. S., BOUCHARDCÔTÉ, A., YUAN, Y., WASSERMAN, W. W., GILKS, C. B., KARNEZIS, A. N., APARICIO, S., MCALPINE, J. N., HUNTSMAN, D. G., HOLT, R. A., NELSON, B. H. & SHAH, S. P. 2018b. Interfaces of Malignant and Immunologic Clonal Dynamics in Ovarian Cancer. *Cell*, 173, 1755-1769.e22.
- ZHANG, C.-Z., SPEKTOR, A., CORNILS, H., FRANCIS, J. M., JACKSON, E. K., LIU, S., MEYERSON, M. & PELLMAN, D. 2015. Chromothripsis from DNA damage in micronuclei. *Nature*, 522, 179-184.
- ZHANG, J., CUNNINGHAM, J. J., BROWN, J. S. & GATENBY, R. A. 2017. Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. *Nature Communications*, 8, 1816.
- ZHANG, Y. & ROSENBERG, A. E. 2017. Bone-Forming Tumors. *Surg Pathol Clin*, 10, 513-535.
- ZHENG, G. X., TERRY, J. M., BELGRADER, P., RYVKIN, P., BENT, Z. W., WILSON, R., ZIRALDO, S. B., WHEELER, T. D., MCDERMOTT, G. P., ZHU, J., GREGORY, M. T., SHUGA, J., MONTESCLAROS, L., UNDERWOOD, J. G., MASQUELIER, D. A., NISHIMURA, S. Y., SCHNALL-LEVIN, M., WYATT, P. W., HINDSON, C. M., BHARADWAJ, R., WONG, A., NESS, K. D., BEPPU, L. W., DEEG, H. J., MCFARLAND, C., LOEB, K. R., VALENTE, W. J., ERICSON, N. G., STEVENS, E. A., RADICH, J. P., MIKKELSEN, T. S., HINDSON, B. J. & BIELAS, J. H. 2017. Massively parallel digital transcriptional profiling of single cells. *Nat Commun*, 8, 14049.
- ZHU, G., BENAYED, R., HO, C., MULLANEY, K., SUKHADIA, P., RIOS, K., BERRY, R., RUBIN, B. P., NAFA, K., WANG, L., KLIMSTRA, D. S., LADANYI, M. & HAMEED, M. R. 2019. Diagnosis of known sarcoma fusions and novel fusion partners by targeted RNA sequencing with identification of a recurrent ACTB-FOSB fusion in pseudomyogenic hemangioendothelioma. *Modern Pathology*, 32, 609-620.