

Characterisation of the genomes of brain metastases

Danny Ulahannan

Nuffield Department of Clinical Medicine

Green Templeton College

University of Oxford

Trinity Term 2015

A thesis submitted for the Degree of Doctor of Philosophy

Characterisation of the genomes of brain metastases

Danny Ulahannan, Green Templeton College, Oxford

D.Phil, Trinity Term, 2015

Abstract

Brain metastases are generally associated with a poor prognosis. The number of metastases, location and control of extracranial disease typically influences the management of patients. Treatment options are currently limited and are dependent on the clinical context. Interventions include surgery and stereotactic radiosurgery with few metastases, with whole brain radiotherapy serving a role in patients with several metastases. In recent years, there have been increased efforts directed towards early phase trials to address the role of targeted biological agents towards pathogenic drivers namely, *EGFR*, *HER2* and *BRAF*. Due to the location and appropriateness of the clinical situation biopsies are impractical in these patients.

In this DPhil, I have characterised the genomic landscape of a cohort of brain metastases of 4 colorectal, 7 ductal breast and 7 lung adenocarcinomas. Analysis of genomes involved: characterisation of mutation burdens, spectra and signatures, profiling copy number aberrations, driver gene discovery and reconstructing the clonal architecture of the genomes. Unpaired primary tumours from The Cancer Genome Atlas, and published reports served as a comparative tool to evaluate the degree of heterogeneity compared to primary tumours.

The brain metastases generally resembled primary tumours with regards to mutation burdens, spectra and signatures. Drivers of primary tumours including those targeted by current biological agents were generally, though not universally located in major clonal groups in the metastases.

Increased chromosomal instability was observed in the metastases. Although my sample size was modest the brain metastases had an increased number of copy number alterations

in comparison to primary tumours. Furthermore, my findings suggest that chromothripsis may occur more frequently in brain metastases. Elucidating mechanisms involved in chromosomal instability in the future will provide further insight into tumour evolution and potential targets for new therapeutic interventions.

Statement of originality

All laboratory procedures were conducted by myself unless otherwise stated. The bioinformatic analysis was undertaken by myself, initially under the guidance of Dr Michal Kovac. I am independent at analysing sequencing and SNP array data and can generate appropriate bespoke scripts for data analysis. Throughout my DPhil I have undertaken courses in computer programming to develop these skills. A robust pipeline to filter and generate confident somatic calls for the work outlined in this DPhil was generated by myself independently.

Funding

I was funded by Hospital Corporation of America (HCA) for four years as a research fellow during this DPhil.

Word Count

49,732

Publications

I am the first and corresponding author for the article below.

Technical and implementation issues in using next generation sequencing of cancers in clinical practice. D. Ulahannan, M Kovac, P J Mulholland, I Tomlinson. Br J Cancer, 2013, (109), 827-835.

Abbreviations

AES	Amino-terminal Enhancer of Split
ANGPLT4	Angiopoietin-like 4
apkc	Atypical Protein Kinase
BBB	Blood Brain Barrier
CCBM	Colorectal Cancer Brain Metastases
CCLM	Colorectal Cancer Liver Metastases
CGH	Comparative Genomic Hybridisation
CNA	Copy Number Alterations
crb	Crumbs Protein
CFP	Cyan Fluorescent Protein
DLG	Discs Large
DTC	Disseminating Tumour Cells
dsDNA	Double-stranded DNA
EGF	Epidermal Growth Factor)
EGFR	Epidermal Growth Factor Receptor
EPCAM	Epithelial Cell Adhesion Molecule
EMT	Epithelial mesenchymal transition
ECM	Extracellular Matrix
FDR	False Discovery Rate
FFPE	Formalin Fixed and Paraffin Embedded
GFP	Green Fluorescent Protein
HK2	Hexokinase 2
HGF	Hepatocyte Growth Factor
IGFR	Insulin-like Growth Factors
IGV	Integrative Genome Viewer
LGL	Lethal Giant Larvae
LOH	Loss of Heterozygosity
MMP	Matrix Metalloproteinases
MET	Mesenchymal-Epithelial Transition
mRNA	Messenger RNA
miRNAs	Micro RNAs
MSI	Microsatellite Instability
MMR	Mismatch Repair
NCAM	Neural Cell Adhesion Molecule
NGS	Next Generation Sequencing
NSCLC	Non-small Cell Lung Cancer
OD	Optical Density
patj	Pals1-associated Tight Junction Protein
PTHrP	Parathyroid-hormone-related-peptide
PGM	Personal Genome Machine
PDGF	Platelet Derived Growth Factor
PCP	Polarity Complex Proteins
pals1	Protein Associated with Lin-7 1
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
siRNA	Small Interfering RNA
SCLC	Small-cell Lung Cancer
TCGA	The Cancer Genome Atlas
TNM	Tumour, Nodes, Metastasis
VAF	Variant Allele Frequency
VCF	Variant Call Format
VEGFR	Vascular Endothelial Growth Factor

WBRT	Whole Brain Radiotherapy
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
ZEB	Zinc-finger E-box-binding Transcription Factors

Table of Contents

Abstract i

Statement of originality iii

Funding iii

Publications iv

Abbreviations v

List of Tables xiv

List of Figures xvi

Chapter One 2

Introduction and aims of thesis 2

1.1 Metastasis 2

1.2 Overview of the metastatic cascade 3

1.3 Organisation of epithelial cells 4

1.4 Local environment – a barrier to invasion 5

1.5 Epithelial-mesenchymal transition 5

1.5.1 Disassembly of cell junctions 6

1.5.2 Loss of apical-basal polarity 6

1.5.4 Regulation of EMT at the RNA level 8

1.6 Tumour microenvironment 9

1.7 Mode of cell motility 10

1.8 Intravasation and dissemination of tumour cells 11

1.8.1 Mode of dissemination 11

1.8.2 Intravasation 12

1.8.3 Anoikis 12

1.9 Evasion of immune system 13

1.10 Arrest and extravasation at target site 13

1.11 Proliferation and Colonisation 14

1.12 Inefficiency of the metastatic process 15

1.13 Chronological models defining the metastatic process 15

1.13.1 The linear progression model 16

1.13.2 The parallel progression model 16

1.14 Clinical evidence supporting linear/parallel progression models 17

1.14.1 The relationship between tumour size and molecularly advanced cancers 17

1.15 Genetic evidence supporting parallel/linear progression models 18

1.15.1 Characterisation of disseminated tumour cells 18

1.15.2 Concordance of driver mutations 19

1.15.3 Convergent evolution 19

1.16 Tissue tropism of metastases 20

1.16.1 Seed and soil hypothesis 20

1.16.2 Factors influencing organ specific colonisation 21

1.17 Tumour latency 23

1.18 Brain Metastasis 24

1.18.1 Demographics of brain metastases 24

1.18.2 Clinical presentation of brain metastasis 25
1.18.3 Prognostic factors associated with brain metastasis 25
1.18.4 Therapies for brain metastasis 25
1.18.5 Targeted therapy for brain metastases 26
1.18.6 Exclusion of patients with brain metastasis from clinical trials 27
1.19 Blood Brain Barrier 27
1.20 Understanding the molecular mechanisms underpinning brain metastasis 28
1.20.1 The role of host stromal cells in the brain 28
1.20.2 The role angiogenesis in the formation of brain metastasis 29
1.20.3 Comparative analysis of gene expression in humans 30
1.20.4 Animal models of brain metastasis 30
1.20.5 Limitations of mouse models 32
1.21 Next generation sequencing 32
1.22 Complexity of cancer genomes 34
1.23 Practical considerations 35
1.24 NGS analysis of cancer genomes 35
1.25 NGS analysis of metastases 36
1.27 Hypotheses 39
1.28 Objectives 39
1.29 Outline of thesis 40

Chapter 2 41

Materials and Methods 41

2.1 Sample description of tumour specimens 41
2.1.1 Sequenced brain metastases 41
2.1.2 Single nucleotide polymorphism arrays of brain metastases 41
2.1.3 Liver metastases 41
2.2 TCGA data 42
2.2.1 Somatic mutation data from primary tumours 42
2.2.2 SNP array data from primary tumours 42
2.3 Ethical approval 42

Laboratory procedures 43

2.4 DNA extraction 43
2.4.1 DNA extraction of fresh frozen brain metastases 43
2.4.2 DNA extraction of buffy coat lymphocytes 43
2.4.3 Phenol-chloroform extraction of fixed paraffin embedded tumours 44
2.4.4 Phenol-Chloroform extraction protocol 45
2.4.5 Harris haematoxylin and eosin staining 46
2.5 Removal of RNA from DNA extractions 47
2.6 Assessment of DNA quality and quantification 48
2.6.1 Agarose gel electrophoresis 48
2.6.2 Spectrophotometry 49
2.6.3 Quantification of DNA using PicoGreen assay 49
2.6.4 Quantification of DNA using Qubit assay 50
2.7 Technical validation of identified somatic variants 51
2.7.1 Amplification of genomic targets of DNA by PCR 51
reaction 51
2.7.2 Partial digestion of primer sequences 53

2.7.3 Ligation of barcode adaptors 53
2.7.4 Purification of unamplified library 53
2.7.5 Amplification of library 53
2.7.6 Purification of amplification clean up 54
2.7.7 Bioanalyzer 54
2.7.7 Sequencing of libraries 55
2.8 Single nucleotide polymorphism arrays 55
2.8.1 Summary of tumours processed using SNP arrays 55
2.8.2 Amplification of DNA 56
2.8.3 Fragmentation and precipitation of DNA 56
2.8.4 Resuspension of precipitated DNA 57
2.8.5 Hybridization on to multi-use beadchip 57
2.8.6 Single base extension and staining of beadchips 58
2.8.7 Imaging beadchips 58
2.9 Analysis of FFPE brain metastases using SNP arrays 59
2.9.1 Illumina FFPE qc protocol 59
2.9.2 Illumina restoration protocol for FFPE arrays 60
Analysis of sequencing data 61
2.10 Next generation sequencing 61
2.11 Analysis of Illumina whole genome sequencing data 62
2.11.1 Summary of sequencing characteristics 62
2.11.2 Mapping of sequencing data 62
2.11.3 Variant calling and filtering of somatic mutations 64
2.11.4 Annotation and further filtering of somatic variants 65
2.11.5 Analysis of trio dataset 66
2.12 Analysis of Complete Genomics data 66
2.13 Multi-region analysis of primary colorectal tumours 68
2.14 Downstream Analysis of sequencing data 68
2.14.1 Analysis of mutation spectra 68
2.14.2 Analysis of mutation signatures 68
2.14.3 OncoSNP-SEQ 69
2.14.4 Analysis of clonal architecture 69
2.14.5 IntOGen analysis of somatic mutations 70
2.14.6 Mutsig analysis of somatic mutations 70
2.14.7 Analysis of structural variants 70
2.14.8 Identification of regions of kataegis 71
2.14.9 Identification of microsatellite instability 72
2.14.10 GISTIC analysis 73
2.14.11 STAC analysis 73
2.14.12 Interrogation of non-coding regions 74
2.14.12.a Transcription factor binding sites 74
2.14.12.b MicroRNA database 74
2.15 Analysis of technical validation 74
2.16 Bioinformatic analysis of SNP arrays 74
2.16.1 Method to obtain comparable SNP densities across platforms 74
2.16.2 OncoSNP analysis 75
2.16.3 Consensus clustering of SNP arrays 76
2.17 Statistical analysis 76

Chapter 3 ..... **79**
Analysis of colorectal cancer brain metastasis genomes **79**
3.1 Introduction 79
3.2 Methods 80
3.2.1 Sequencing of CCBM genomes 80
3.2.2 Downstream analysis of WGS data 80
3.3 Comparison Datasets 81
3.3.1 Somatic mutations from TCGA primary colorectal tumours 81
3.3.2 Copy number information from primary colorectal tumours 81
3.3.3 Colorectal cancer metastases to the liver 81
3.4 Technical validation of sequencing data 82
3.5 Patient Demographics 82
Results **85**
3.6 Mutation burdens of CCBM 85
3.7 Screen for microsatellite instability and *POLE* mutations 87
3.8 Analysis of mutation spectra 88
3.9 Analysis of mutation signatures 89
3.10 Identification of putative drivers of CCBM 91
3.11 Screen for drivers of primary colorectal cancer 92
3.12 Clonal architecture of CCBM 94
3.13 Evaluation of copy number alterations 97
3.14 Identification of recurrent focal copy number aberrations 98
3.16 Structural Variants 100
3.17 Comparison of colorectal brain metastases with other data sets. 101
3.17.1 Comparison of mutation burdens with primary tumours 102
3.17.2 Mutation burdens of CCLM 102
3.17.3 Comparison of mutation spectra 104
3.17.4 Evaluation of copy number alterations 105
3.17.5 Identification of recurrent focal copy number aberrations 108
3.17.6 Identification of chromosomal aberrations 110
3.18 Technical validation of sequencing data 111
3.19 Discussion 113
3.19.1 Mutation burden and spectra/signatures 113
3.19.2 Clonal architecture of the colorectal cancer brain metastasis 113
3.19.3 Profiling copy number alterations in colorectal metastases 114
3.19.4 Screen for chromosomal aberrations 115
Chapter 4 ..... **116**
Characterisation of a paired colorectal primary tumour and brain metastasis **116**
4.1 Introduction 116
4.2 Methods 117
4.2.1 WGS of a matched CCBM and primary tumour 117
4.3 Comparison Datasets 117
4.3.1 Analysis of multi-region sampling from primary colorectal adenocarcinomas 117
4.4 Case Description 118
Results **118**
4.5 Mutation burdens 118

4.5.1 Mutation burdens-assessment of original dataset 118
4.5.2 Mutation burden-assessment following downsampling 119
4.6 Evaluation of variant allele frequency 122
4.7 Characterisation of mutation signatures 123
4.8 Evaluation of copy number alterations 126
4.9 Screen for drivers of primary colorectal cancer 127
4.10 Characterisation of clonal architecture 128
4.11 Screen for structural variants 131
4.12 Technical validation of sequencing results 132
4.13 Comparative analysis of heterogeneity using multi-region analysis of primary colorectal adenocarcinomas 133
4.14 Discussion 137
4.14.1 Comparison of sequencing studies of primary and metastatic tumours 137
4.14.2 Parallel and linear progression models 138
Chapter 5 140
Analysis of lung adenocarcinoma brain metastases genomes 140
5.1 Introduction 140
5.2 Methods 141
5.2.1 WGS of LABM genomes 141
5.3 Comparison Datasets 141
5.3.1 TCGA primary lung adenocarcinoma tumours 141
5.4 Patient Demographics 141
Results 143
5.5 Mutation burden 143
5.5.1 Mutation burdens of LABM 143
5.6 Analysis of mutation spectra 146
5.7 Analysis of mutation signatures 147
5.8 Identification of putative drivers of LABM 149
5.9 Screen for primary lung adenocarcinoma driver mutations 150
5.10 Clonal architecture of LABM 153
5.11 Copy number aberrations in the LABM 158
5.11.1 Evaluation of copy number alterations in LABM 158
5.11.2 Identification of recurrent focal copy number aberrations 159
5.11.3 Concordance of copy number information from SNP array and sequencing data. 160
5.12 Structural Variation 160
5.13 Technical validation of sequencing data 162
5.14 Comparison of LABM with primary tumours 163
5.14.1 Comparison of mutation burdens with primary lung adenocarcinoma tumours 163
5.14.2 Comparison of mutation spectra with primary lung adenocarcinoma tumours 163
5.14.3 Comparison of copy number alterations in primary lung adenocarcinoma tumours and LABM 164
5.15 Discussion 166
5.15.1 Mutation burdens and spectra 166
5.15.2 Driver mutations and clonal architecture 167
5.15.3 Evaluation of copy number alterations 167

Chapter 6 169
Analysis of Breast Ductal Carcinoma Brain Metastases genomes 169
6.1 Introduction 169
6.2 Methods 170
6.2.1 WGS of BDCBM genomes 170
6.3 Comparison Datasets 170
6.3.1 Somatic mutations from TCGA primary breast tumours 170
6.4 Patient Demographics 170
Results 172
6.5 Mutation Burdens 172
6.5.1 Mutation burdens of BDCBM genomes 172
6.6 Analysis of mutation spectra 174
6.7 Analysis of mutation signatures 175
6.8 Screen for germline mutations in BDCBM 176
6.9 Identification of putative drivers of breast brain metastases 177
6.10 Screen for primary driver mutations in BDCBM 177
6.11 Clonal architecture of BDCBM 179
6.12 Evaluation of copy number alterations in BDCBM 183
6.13 Identification of recurrent focal copy number aberrations 184
6.14 Concordance of copy number information from SNP array and sequencing data 185
6.15 Structural Variation 186
6.15.1 chromothripsis/chromoplexy 186
6.15.2 Identification of regions of kataegis 187
6.16 Technical validation of sequencing data 188
6.17 Comparison of BDCBM with primary tumours 189
6.17.1 Comparison of mutation burdens with published findings 189
6.17.2 Analysis of mutation spectra 190
6.17.3 Evaluation of copy number alterations in BDCBM 191
6.18 Discussion 191
6.18.1 Mutation burdens and signatures 191
6.18.2 Screen for chromosomal aberrations 192
6.18.3 Structural variants 193
Chapter 7 194
Amalgamated analysis of brain metastases genomes 194
7.1 Introduction 194
7.2 Methods 194
7.2.1 Sequencing of brain metastases 194
7.2.2 Evaluation of non-protein coding regions from sequencing data 194
7.2.3 Copy number analysis of brain metastases 195
7.2.3.1 GISTIC and STAC analysis of brain metastases 195
7.2.3.2 Analysis of FFPE brain metastases using SNP arrays 195
7.2.3.3 Cluster analysis of copy number aberrations 195
Results 196
Table 7.1 | Summary of characteristics for each brain metastasis. **A time interval of 0 refers to brain metastases manifesting at the initial presentation of the cancer. ... 196**
7.3.1 Mutation Burdens 197

7.3.2 Mutation spectra of brain metastases 199
7.3.3 Mutation signatures 200
7.3.4 Clonal structures of brain metastases 202
7.3.5 Copy number aberrations 203
7.3.6 Chronological relationships 204
7.4 Interrogation of non-protein coding regions 204
7.5 Driver gene discovery in brain metastasis 204
7.6 Clonal location of *DST* 207
7.7 Copy number analysis of brain metastases as a cohort 210
7.7.1 GISTIC and STAC analysis of brain metastases 210
7.8 SNP array analysis of FFPE brain metastases 211
7.9 Consensus clustering of brain metastases 214
7.10 Discussion and overall conclusions 215
7.10.1 Mutation Burdens 216
7.10.2 Mutation spectra/signatures 217
7.10.3 Driver mutations in brain metastases 217
7.10.4 Chromosomal aberrations 218
References 221

List of Tables

Table 1.1 Examples of miRNAs associated with transcription factors regulating EMT.....9

Table 1.2 Overview of the genes that have been implicated with brain metastases.....33

Table 1.3 Summary of recent NGS projects characterising metastasis.....38

Table 2.1 Summary of platforms used for SNP arrays.....56

Table 2.2 Incubation steps in single base extension.....58

Table 2.3 Sequential steps in the staining process.....58

Table 2.4 Summary of SNP array platforms and density of probes.....75

Table 2.5 Summary of sequenced tumours and overview of analysis.....77

Table 2.6 Summary of tumours analysed for copy number alterations.....78

Table 3.1 Summary of clinical information for patients who had their brain metastasis sequenced...82

Table 3.2 Summary of clinical information for patients who had their liver metastases resected.....83

Table 3.3 Primary driver mutations in colorectal cancer in the CCBM.....92

Table 3.4 Number of clonal populations in CCBM.....93

Table 3.5 Summary of copy number aberrations in individual CCBM.....97

Table 3.6 Summary of focal copy number aberrations in CCBM and CCLM.....108

Table 3.7 Summary of chromothripsis-like changes in CCBM and CCLM.....109

Table 3.8 Summary of technical validation of CCBM and CCLM.....124

Table 4.1 Variant allele frequencies in the primary (P1) and brain metastasis (C1MET).....121

Table 4.2 Summary of colorectal driver mutations identified in P1 and C1MET.....128

Table 4.3 Summary of validated mutations in C1MET and P1.....131

Table 4.4 Spatial heterogeneity in MR1.....135

Table 4.5 Spatial heterogeneity in MR2.....135

Table 5.1 Summary of clinical information for patients who had their lung adenocarcinoma brain metastasis sequenced.....141

Table 5.2 Summary of identified somatic mutations in driver genes primary lung adenocarcinoma in LABM.....151

Table 5.3 Summary of clonal populations in the LABM.....152

Table 5.4 Summary of copy number aberrations in individual LABM.....157

Table 5.5 Summary of mutations screened in technical validation.....161

Table 6.1 Summary of clinical information for patients who had their breast brain metastases sequenced.....171

Table 6.2 Summary of identified somatic mutations in primary driver genes in the BDCBM.....178

Table 6.3 Summary of clonal populations in breast brain metastases.....179

Table 6.3 Summary of copy number aberrations characteristics in BDCBM.....184

Table 6.4 Summary of somatic variants validated in BDCBM.....189

Table 7.1 Summary of characteristics for each brain metastasis.....197

Table 7.2 *DST* was identified as putative driver for brain metastases.....207

Table 7.3 Summary of chromosomal aberrations in FFPE derived brain metastases.....214

List of Figures

Figure 1.1 The sequential steps involved in the metastatic cascade....3
Figure 1.2 Organisation of cell junction and polarity complex proteins....7
Figure 1.3 Target genes regulated by EMT core transcription factors....8
Figure 1.4 Model for the development of osteolytic metastasis....23
Figure 2.1 DNA extraction using phenol-chloroform and Qiagen DNA extraction protocols....45
Figure 2.2 Haematoxylin and eosin staining of sections of primary colorectal cancer....47
Figure 2.3 Workflow summarising steps in library preparation for Ion Torrent....52
Figure 2.4 Electropherogram demonstrating quality of DNA library....55
Figure 2.5 qPCR trace using Illumina FFPE qc kit....60
Figure 2.6 Outline of processing steps involved in cancer genome analysis....63
Figure 2.7 Example of kataegis identified in breast brain metastasis....72
Figure 3.1 Mutation burdens in the CCBM in the whole genome....84
Figure 3.2 Number of mutations in exome in CCBM....85
Figure 3.3 Distribution of types of protein-coding mutations in CBBM....86
Figure 3.4 Proportion of base substitutions in CCBM in the exome....87
Figure 3.5 Mutation signatures detected in CCBM....90
Figure 3.6 Clonal architecture of C1MET....94
Figure 3.7 Clonal architecture of C2MET....94
Figure 3.8 Clonal architecture of C3MET....95
Figure 3.9 Clonal architecture of C5MET....95
Figure 3.10 Focal deletions in *RBFOX1*....98
Figure 3.11 Focal deletions in MACROD2....98
Figure 3.12 Focal deletion in *MACROD2* in C1MET detected by OncoSNP-SEQ....99
Figure 3.13 Chromothripsis-like alterations in C5MET....100
Figure 3.14 Mutation burdens in CCLM in the whole genome....102
Figure 3.15 Exonic mutation burdens in CCLM in whole genome....103
Figure 3.16 Mutation spectra in primary and metastatic colorectal cancers....104
Figure 3.17 Copy number variation in colorectal primary tumours and metastases....106
Figure 3.18 Focal deletion detected in the CCLM....108

Figure 3.19 Chromothripsis-like alterations in CCBM and CCLM.....110

Figure 4.1 Mutation burdens in C1MET and P1 in original and downsampled analysis (whole genome).....119

Figure 4.2 Distribution of mutations in the exome in C1MET and P1 in original and downsampled analysis.....120

Figure 4.3 Characterisation of mutation signatures in C1MET and P1.....123

Figure 4.4 Relative contribution of mutation signatures in C1MET and P1.....124

Figure 4.5 Regions of gains and deletions in C1MET and P1.....126

Figure 4.6 Clonal architecture represented in P1.....129

Figure 4.7 Clonal architecture represented in C1MET.....129

Figure 4.8 Mutation burdens in multi-region analysis in primary colorectal tumour MR1.....133

Figure 4.9 Mutation burdens in multi-region analysis in the primary colorectal tumour MR2.....134

Figure 5.1 Mutation burdens in LABM in the whole genome.....142

Figure 5.2 Number of mutations in exome in LABM.....143

Figure 5.3 Distribution of types of protein-coding mutations in LABM.....144

Figure 5.4 Distribution of the 6 classes of base substitutions in the LABM.....146

Figure 5.5 Mutation signatures in LABM.....147

Figure 5.6 Clonal architecture in brain metastasis L1MET.....153

Figure 5.7 Clonal architecture in brain metastasis L2MET.....153

Figure 5.8 Clonal architecture in brain metastasis L3MET.....154

Figure 5.9 Clonal architecture in brain metastasis L4MET.....154

Figure 5.10 Clonal architecture in brain metastasis L6MET.....155

Figure 5.11 Clonal architecture in brain metastasis L7MET.....155

Figure 5.12 Clonal architecture in brain metastasis L8MET.....156

Figure 5.13 Example of focal amplification in DLGAP1 in L3MET.....158

Figure 5.14 Focal amplification in DLGAP1 in L3MET detected in OncoSNP-SEQ.....159

Figure 5.15 Chromothripsis-like alteration L6MET.....160

Figure 5.16 Mutation spectra in primary and metastatic lung adenocarcinomas.....163

Figure 5.17 Copy number variations in lung adenocarcinomas.....164

Figure 6.1 Mutation burdens in BDCBM in the whole genome.....172

Figure 6.2 Number of mutations in exome in BDCBM.....173

Figure 6.3 Distribution of types of protein-coding mutations in the BDCBM.....173

Figure 6.4 Distribution of the 6 classes of base substitutions in the BDCBM across the genome....174
Figure 6.5 Mutation signatures in BDCBM in sequence context....175
Figure 6.6 Clonal architecture in B1MET....180
Figure 6.7 Clonal architecture in B2MET....180
Figure 6.8 Clonal architecture in B3MET....181
Figure 6.9 Clonal architecture in B4MET....181
Figure 6.10 Clonal architecture in B5MET....182
Figure 6.11 Clonal architecture in B6MET....182
Figure 6.12 Clonal architecture in B7MET....183
Figure 6.13 Examples of focal amplifications in HER2 in B4MET and B5MET....185
Figure 6.14 Example of focal amplifications in *HER2* in B5MET identified with OncoSNPSEQ....186
Figure 6.15 Examples of chromothripsis-like aberrations B1MET and B5MET....187
Figure 6.16 Regions of kataegis identified in chromosome 14 and 15 in B6MET....188
Figure 6.17 Mutation spectra in primary and metastatic breast adenocarcinomas....190
Figure 7.1 SNV mutation burdens in the brain metastases....198
Figure 7.2 Mutations in the exome in the brain metastases....199
Figure 7.3 Mutations in protein-coding regions represented as a proportion of the mutation burden in the exome....200
Figure 7.4 Mutation signatures in brain metastases....202
Figure 7.5 Clonal architecture of C1MET....208
Figure 7.6 Clonal architecture of C3MET....209
Figure 7.7 Clonal architecture in L2MET....209
Figure 7.8 Clonal architecture in L3MET....210
Figure 7.9 Clonal architecture of B7MET....210
Figure 7.10 Comparison of SNP array analysis of B3MET....213
Figure 7.11 Chromothripsis identified in brain metastasis derived from FFPE B10MET and C9MET....213
Figure 7.12 Consensus clustering of copy number aberrations....215

I would like to thank my wife Tina, Professor Ian Tomlinson and the bioinformaticians Dr Francesc Castro and Dr Michal Kovac for support throughout my DPhil.

Chapter One

Introduction and aims of thesis

1.1 Metastasis

Metastasis is the process in which cancer spreads from the primary tumour and emerges at other sites. Patients may be asymptomatic, with metastatic tumours being detected through clinical investigations. The symptoms experienced in patients with metastatic cancer will depend on the location and size of the metastatic tumour. Treatment strategies adopted in patients presenting with metastases will depend on the frequency, histological subtype, size and location of the metastases as well as the patient's performance status (a measure of the well being and level of activity of the patient) (Weichselbaum and Hellman, 2011, Palma et al., 2014). Patients presenting with oligometastatic disease may be considered as potential candidates for either curative treatment or long-term disease control using either surgery or stereotactic radiosurgery (SRS) with adjuvant therapy. Despite this, many patients presenting with metastatic disease are generally considered to be incurable due to the number and distribution of metastases. The aim of treatment in this group of patients may include: controlling tumour growth, prolonging survival, and alleviating symptoms, whilst maintaining quality of life. In patients who have had their tumours resected, the goal of adjuvant systemic therapy is to target residual, clinically undetectable tumour cells that may remain following surgery that may result in local or distant recurrence. Future avenues of research to drive advances in cancer treatment for metastatic disease may be directed towards increasing our understanding of the molecular and genetic mechanisms underpinning metastatic disease and characterising the degree of heterogeneity that exists amongst primary and metastatic tumours. Progress in these areas of research may guide the direction of future therapeutic strategies, design of

clinical trials and inform clinical decisions made when treating patients with metastatic disease.

1.2 Overview of the metastatic cascade

Metastasis is a complex multistep process that culminates in the successful colonisation of the distant target organ (Valastyan and Weinberg, 2011, Fidler, 2003, Gupta and Massague, 2006). The metastatic cascade is depicted in Figure 1.1.

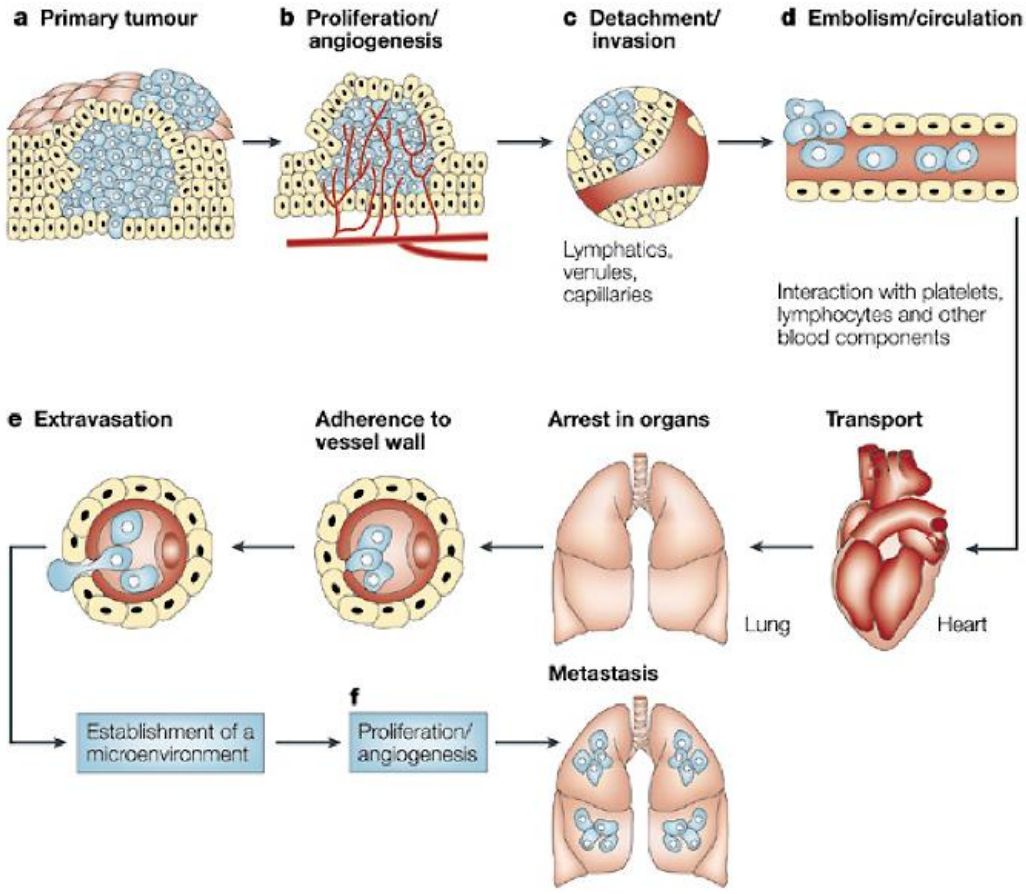


Figure 1.1 l The sequential steps involved in the metastatic cascade. Image derived from (Fidler, 2003). Epithelial cells in the primary tumour proliferate and undergo EMT resulting in increased cell motility coupled with ECM degradation and angiogenesis. Metastatic founder cells detach and disseminate into the blood and lymphatic circulation. Cells that develop into metastatic foci arrest in the target organ, extravasate into the parenchyma, proliferate and colonise the new site.

In brief, epithelial cells in the primary tumour undergo a process of epithelial-mesenchymal transition (EMT). The basement membrane is degraded and invading

tumour cells detach from the primary tumour. Tumour cells enter the extracellular matrix (ECM) and stromal layers and subsequently intravasate into the lumen of blood vessels and disperse. Circulating tumour cells proceed to arrest in capillaries at distant sites where they will extravasate and enter the parenchyma of target organs. Only tumour cells that proceed to successfully adapt, proliferate and colonise the distant organ will emerge as clinically detectable metastases. The distinct biological steps of the metastatic cascade are discussed in further detail in sections 1.3-1.11.

1.3 Organisation of epithelial cells

Carcinomas are malignant tumours that originate from epithelial cells. Epithelial cells form 'sheet like' networks that line cavities and the surfaces of structures in the body, and form a key component in the secretory role of glands and ducts. Epithelia can be classified according to morphology depending on the cell shape, number of layers and specialisation. The coordinated alignment and orientation of epithelial cells in tissue architecture is preserved through several mechanisms, which act in concert (McCaffrey and Macara, 2011, Rodriguez-Boulan and Macara, 2014). The adhesion of neighbouring epithelial cells is regulated through the formation of specialised junctions (Rodriguez-Boulan and Macara, 2014). Neighbouring epithelial cells remain in contact through adherens junctions, which are intercellular junctions, mediated through the calcium dependent cell adhesion molecule epithelial cadherin (E-cadherin). The extracellular domain of the cadherin molecule mediates cell-cell adhesion, whereas the cytoplasmic domain interconnects with the actin cytoskeleton via catenins (Wheelock et al., 2008). In vertebrates, tight junctions oppose the membranes of neighbouring cells at apical and basolateral surfaces mediated principally through the plasma membrane proteins claudin and occludin. Gap junctions are intercellular channels which allow direct communication between adjacent cells permitting the transfer of small molecules and ions (Goodenough and Paul, 2009). Desmosomes are adhesive intercellular junctions that are important in tissues that experience mechanical stress (Delva et al.,

2009). Epithelial cells have an inherent capacity to discriminate between the external and internal environments through apical-basal polarity, which is orchestrated through a series of polarity complex proteins (PCP) (Rodriguez-Boulan and Macara, 2014). The ordered alignment of epithelial cells is further supported by the basement membrane which has a key structural role in maintaining tissue architecture as well as participating in signal transduction (LeBleu et al., 2007).

1.4 Local environment – a barrier to invasion

The process of invasion involves tumour cells breaching the basement membrane to enter the surrounding stroma and normal tissue parenchyma. The local microenvironment has been proposed to serve a role in preserving the stability of the tissue architecture and acting as a barrier to invasion by tumour cells (Valastyan and Weinberg, 2011). Evidence to support this has been observed in mammary tissue where myoepithelial cells form a layer of cells that are in direct contact with the basement membrane and luminal epithelial cells. MCFCDIS mammary cells when injected into nude mice form a ductal carcinoma *in situ* model that resembles a basal-like subtype of breast cancer (Hu et al., 2008). The ductal carcinoma like xenografts proceed to spontaneously develop invasive carcinomas. Using this xenograft model Hu et al demonstrated that administration of co-injected normal myoepithelial cells, (cells in the breast tissue that juxtapose luminal epithelial cells) suppressed the development of invasive cancer.

1.5 Epithelial-mesenchymal transition

During the incipient stages of metastasis, tumour cells undergo the process of EMT, a process which occurs in embryogenesis, tissue repair and fibrosis (Thiery et al., 2009, Kalluri and Weinberg, 2009). Epithelial cells transdifferentiate to form mesenchymal cells gaining increased motility. The process of EMT is characterised by: a loss of apical-basal cell polarity, concurrent disassembly of intercellular junctions, structural

reorganisation of the cytoskeleton with accompanying changes in cell shape and reprogramming of gene expression with activation of genes that define mesenchymal cells (Lamouille et al., 2014). Cells that have undergone EMT acquire resistance to apoptosis and senescence (Thiery et al., 2009).

1.5.1 Disassembly of cell junctions

EMT is characterised by the loss of integrity of cell junctions (Huang et al., 2012). The disintegration of adherens junctions is associated with the down regulation of E-cadherin by transcriptional or post-translational control (Yilmaz and Christofori, 2009). The down regulation of E-cadherin is counterbalanced by the expression of neural cadherin (N-cadherin), a mesenchymal marker (Wheelock et al., 2008, Jeanes et al., 2008). The cell-cell interactions mediated through homotypic N-cadherin interactions are weaker than homotypic E-cadherin cell-cell interactions facilitating cell migration and invasion. Other mesenchymal markers that are up regulated include vimentin and fibronectin. There is also an increased expression of the neural cell adhesion molecule (NCAM). NCAM can modulate the activity of the SRC family of receptor tyrosine kinases to promote focal adhesions and migration (Lehembre et al., 2008).

1.5.2 Loss of apical-basal polarity

In vertebrates the apical and basolateral compartments of a cell are defined by a series of PCP (Figure 1.2). The initiation of EMT is associated with reduced expression and relocation of these proteins leading to increased cellular motility and invasive properties (Lamouille et al., 2014). During the process of EMT the intrinsic apical-basal polarity of the cell is lost and is replaced with the acquisition of front-rear polarity with the front of the cell forming a leading edge. The procurement of front-rear polarity enables the cells to migrate directionally (Petrie et al., 2009). Cells that undergo EMT remodel the cortical actin cytoskeleton and form membranous projections namely filopodia, lamellipodia and invadopodia which aid cell locomotion (Ridley, 2011). Structural reorganisation of the cytoskeleton scaffold results in increased cell contractility.

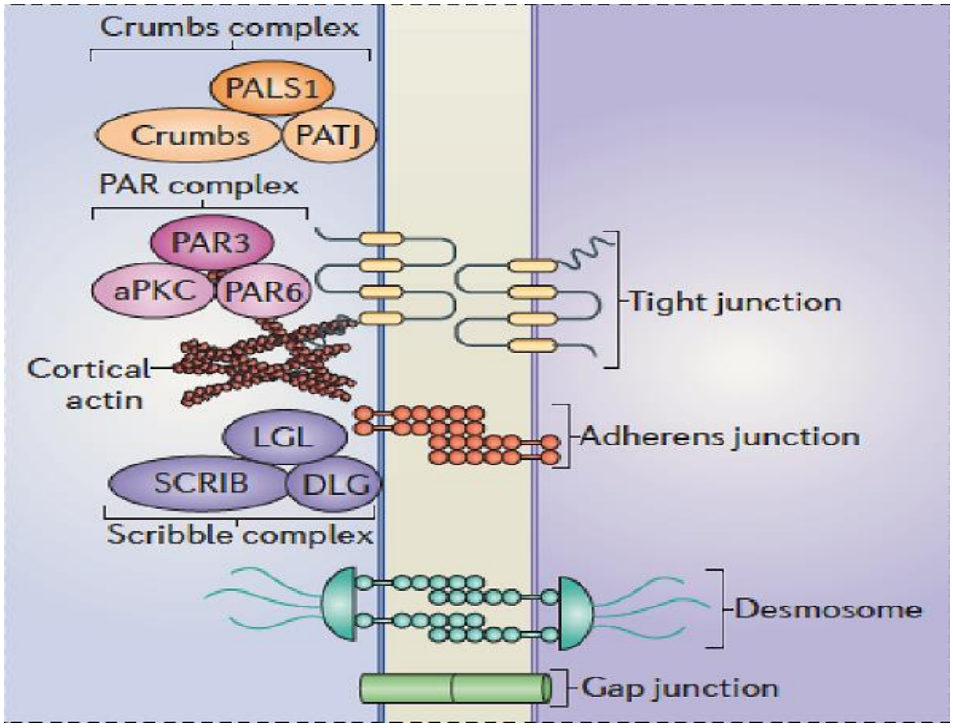


Figure 1.2 l Organisation of cell junction and polarity complex proteins. Image modified from (Lamouille et al 2014). In vertebrates the PAR complex [PAR6, PAR3 and atypical protein kinase (aPKC)] and the Crumbs complex [Crumbs protein (CRB), protein associated with Lin-7 1 (PALS1) and PALS1-associated tight junction protein (PATJ)] define the apical compartments within the cell. The basolateral compartment of the cell is defined by scribble complex [scribble (SCRIB), discs large (DLG) and lethal giant larvae (LGL)]. In the acquisition of front-rear polarity the Scribble and PAR complexes alongside PATJ relocate to the leading edge of the cell. RHOA a RHO GTPase, at the posterior end of the cell promotes the disassembly of adhesion complexes and prevents the formation of PAR complexes.

Numerous signalling pathways are involved in the initiation and progression of EMT including TGF-β, Notch, Hedgehog, Wnt/β-catenin (Ell and Kang, 2013). These signalling pathways converge on core transcription factors including TWIST, SNAIL, SLUG (also known as SNAIL2) and zinc-finger E-box-binding transcription factors (ZEB) which orchestrate EMT by regulating the transcription of E-cadherin (Batlle et al., 2000, Hajra et al., 2002, Comijn et al., 2001). They also mediate the repression of junctional proteins required for cell-cell interactions (De Craene and Berx, 2013). The contribution of a particular transcription factor in the EMT process is dependent on the context of cell/tissue type. These transcription factors induce EMT by suppressing the expression of epithelial genes, whilst promoting the expression of genes associated with a mesenchymal phenotype (Figure 1.3).

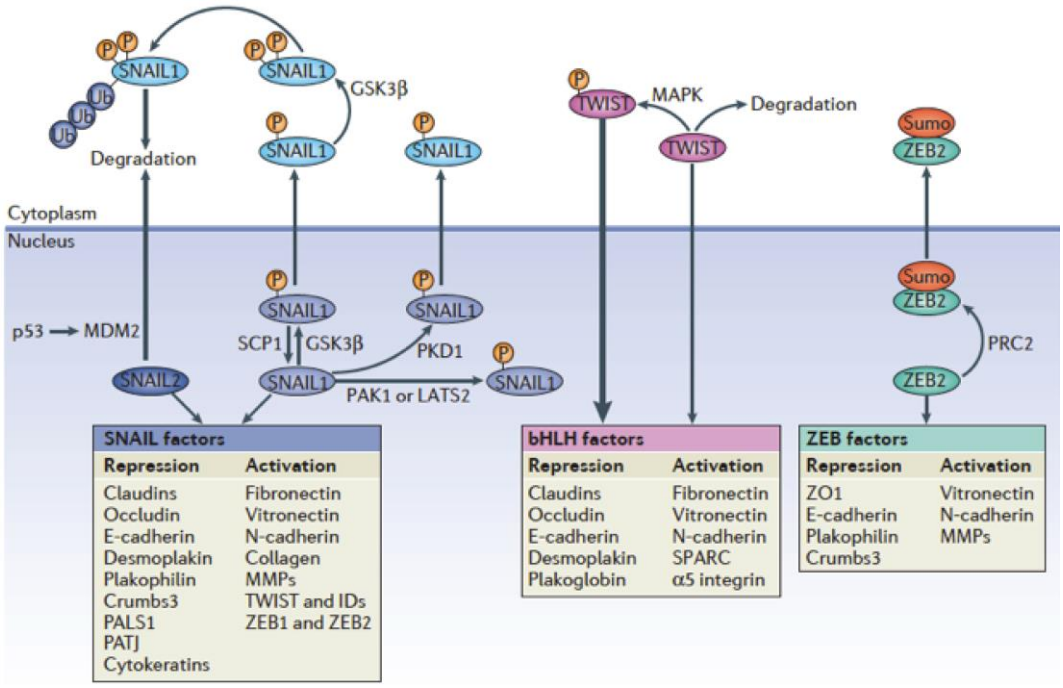


Figure 1.3 I Target genes regulated by EMT core transcription factors. Image derived from (Lamouille et al 2014).

1.5.4 Regulation of EMT at the RNA level

Non-coding micro RNAs (miRNAs) are small nucleotides (20-22) in length, which can regulate post-transcriptional gene expression (Diaz-Lopez et al., 2014, Pencheva and Tavazoie, 2013). They can selectively bind to complementary sequences in the 3' untranslated region of messenger RNA (mRNA) and either inhibit translation or promote degradation. A single miRNA can regulate the translation of several mRNA targets. miRNAs form complex regulatory networks which have been associated with metastasis suppressive and promoting effects (Tavazoie et al., 2008, Ma et al., 2007, Ma et al., 2010, Lin et al., 2013, Cuiffo et al., 2014). Several miRNAs have been associated with the regulation of the key transcription factors orchestrating EMT (Table 1.1).

miRNA	Transcription Factor
miR-29b, miR-30a, miR-34, miR-203	SNAIL1
miR-1, miR-200b	SLUG
miR let-7	TWIST
miR-200 family, miR-205	ZEB1
miR-200 family, miR-205	ZEB2

Table 1.1 I Examples of miRNAs associated with transcription factors regulating EMT

Although miRNAs can regulate the expression of transcription factors, transcription factors can in turn repress the expression of miRNAs resulting in a negative feedback mechanisms (Liu et al., 2013, Moes et al., 2012). Additionally, miRNAs control the expression of genes that define epithelial and mesenchymal phenotypes by modulating the expression of cell adhesion proteins and signal transduction mediators.

1.6 Tumour microenvironment

Invading tumour cells have to overcome physical barriers, chiefly, the basement membrane and ECM. Matrix metalloproteinases (MMP) a family of zinc-dependent endopeptidases are activated by adjacent tumour and stromal cells and are responsible for the degradation of the basement membrane and remodelling the ECM facilitating invasion (Kessenbrock et al., 2010). MMP can modulate the bioavailability of growth factors within the tumour microenvironment. Liberation of growth factors such as epidermal growth factor receptor (EGFR) and insulin-like growth factors (IGFs) promotes tumour proliferation and cell survival. Clinical trials using synthetically generated metalloproteinase inhibitors have yielded disappointing results (Coussens et al., 2002).

Through the process of invasion tumour cells encounter a variety of cell types within the stroma including macrophages, fibroblasts and endothelial cells. These cells have the

capacity to augment the invasive behaviour of tumour cells. Cancer associated fibroblasts secrete cytokines including, vascular endothelial growth factor (VEGF), platelet derived growth factor (PDGF) and hepatocyte growth factor (HGF) which stimulate angiogenesis and proteolytic enzymes (Junttila and de Sauvage, 2013). The degree of recruitment of immune cells to the tumour microenvironment varies amongst cancers. A predominant number of infiltrating T-cells have been associated with a good prognosis in colorectal and breast cancer and melanoma (Fridman et al., 2012). Tumour cells have adapted to inhibit the activity of cytotoxic T-cells by stimulating inhibitory receptors (Pardoll, 2012, Butt and Mills, 2014). They also recruit myeloid derived lineages (neutrophils, dendritic cells and monocytes) fostering an immunosuppressive environment as these cells can inhibit T-cell expansion and activation (Junttila and de Sauvage, 2013).

1.7 Mode of cell motility

At the cellular level, the process of invasion consists of different morphological patterns of cellular motility (Friedl et al., 2012). Tumour cells can either invade collectively in a cohesive group or individually. Giampieri et al utilised intravital imaging to demonstrate that both individual and collective cellular motility can co-exist in a breast cancer model (Giampieri et al., 2009). TGF-β signalling has a role in promoting EMT. To monitor the response to TGF-β signalling, rat mammary carcinoma cells (MTLn3E) were engineered to indicate an activated TGF-β transcriptional response using a cyan fluorescent protein (CFP) reporter. Nearly all cells moving singly were CFP-positive whilst cells moving collectively were CFP-negative. However, some non-motile cells exhibited a CFP-positive response implying that TGF-β signalling alone was insufficient to promote cancer cell motility. To evaluate the role of TGF-β signalling further, MTLn3E cells were constructed to express a dominant-negative TGF-β type-II receptor fused to green fluorescent protein (GFP) [TGF-βRDN-GFP]. Tumour cells expressing TGF-βRDN-GFP could move in a cohesive manner, though the single cell motility was

lost, whereas control cells expressing GFP alone were able to move singly. In contrast, cells ectopically overexpressing TGF-β1 were found to exhibit increased single cell motility. In summation, these findings suggest that TGF-β signalling is required by tumour cells for single cell motility. The distribution of cells and the migratory patterns were analysed using time-lapse analysis. Cells mobilising collectively were found within lymphatic vessels. In contrast cells overexpressing TGF-β1 migrating singly demonstrated increased capacity to enter the vascular system.

1.8 Intravasation and dissemination of tumour cells

1.8.1 Mode of dissemination

Tumour cells have been reported to disseminate through 3 routes: haematogenous circulation, lymphatic invasion and transcoelomic spread. There is a growing body of clinical evidence suggesting that although the presence of lymph node invasion serves as an important prognostic factor, dissemination through the vascular system may be a more important mode by which tumour cells disseminate to distant targets. In breast cancer, patients that have clinically tumour free lymph nodes and a positive sentinel lymph node will have 10-12 lymph nodes removed. Findings from several studies have reported that the removal of clinically negative nodes improved local control in patients receiving mastectomy and radiotherapy compared to mastectomy alone (Fisher et al., 2002). However there was no significant impact on survival in patients or on the occurrence of distant metastasis. (Fisher et al., 2002, Veronesi et al., 1999, Rudenstam et al., 2006). Pelvic lymph node metastasis can occur in approximately 10% of patients with stage 1 endometrial cancer (Creasman et al., 1987). The multinational ASTEC randomised controlled trial evaluated the role of systematic pelvic lymphadenectomy in patients with stage 1 endometrial cancer (Kitchener et al., 2009). Systematic resection of pelvic lymph nodes failed to demonstrate any benefit on overall or recurrence free survival. They concluded that pelvic lymphadenectomy could not be recommended as a routine therapeutic procedure in this subgroup of patients.

1.8.2 Intravasation

The process of intravasation involves tumour cells entering the lumen of blood vessels. The molecular mechanisms involved are being unravelled. Tumour cells can stimulate angiogenesis by releasing cytokines and growth factors. Blood vessels in the tumour microenvironment formed through angiogenesis are tortuous and have increased permeability resulting in irregular blood flow (Weis and Cheresh, 2011). Integrins are transmembrane receptors that are responsible for cell-matrix interactions and coordinate signals from the extracellular matrix (ECM) through signal transduction pathways (Desgrosellier and Cheresh, 2010). Integrins are heterodimers and are composed of α and β subunits that ligate to proteins within the ECM. Reymond et al demonstrated that transendothelial migration of tumour cells is facilitated by the expression of β1 integrin in prostrate cancer cells (Reymond et al., 2012). Using an *in vitro* screen with small interfering RNA (siRNA) in prostate cancer cells they identified Cdc42, a Rho GTPase, as a regulator of transendothelial migration. In addition, they reported that Cdc42-depleted tumour cells expressed reduced levels of β1 integrin. Moreover, they demonstrated transcriptional regulation of the β1 integrin promoter by Cdc42. Sonoshita et al reported increased intravasation of tumour cells compared to that observed in *Apc* mutant mice in an amino-terminal enhancer of split (Aes) knockout (*Apc/Aes*) in a transgenic model of polyposis (Sonoshita et al., 2011). Aes was found to block the intravasation of tumour cells by inhibiting Notch signalling.

1.8.3 Anoikis

Whilst travelling through the vascular system to survive, tumour cells have to confront a number of challenges. Circulating tumour cells have to withstand the haemodynamic stress forces and avoid mechanical destruction. Tumour cells are anchorage independent and evade the process of anoikis through the expression of cell-cell adhesion molecules, protein kinases and integrin signalling (Paoli et al., 2013). Anoikis is a form of programmed cell death that occurs when anchorage dependent cells

detach from the ECM leading to a loss of cell-matrix interactions. It has been proposed to serve as a physiological barrier to metastasis.

1.9 Evasion of immune system

Circulating tumour cells are vulnerable to attack from the innate immune system and have to evade immunological responses from the host. Tumour cells have adapted by using platelets to guard them from immunological surveillance (Gay and Felding-Habermann, 2011). Tumour cells manipulate the coagulation system by expressing tissue factor. This interacts with coagulation factors VII and X which generates thrombin. Activation of thrombin subsequently results in fibrin formation and platelet aggregation through integrin αIIβ3. Activated platelets bind to fibrinogen and become incorporated into fibrin clots promoting thrombus formation. Experimental evidence from mouse models have supported a role for platelets in the development of metastasis. Carmerer et al generated knock-out mice models to examine the role of platelets in a model of metastasis using B16-F10 melanoma cells expressing luciferase (Camerer et al., 2004). They demonstrated reduced metastatic tumour burden in the genetically engineered mice in comparison to wild-type. The genetically engineered mice included Nf-/E2$^{-/-}$ mice which lack circulating platelets, Par4$^{-/-}$ mice which have platelets that are unresponsive to thrombin and fail to aggregate and Fib$^{-/-}$ which are mice that lack fibrinogen. In keeping with this, early microscopic studies conducted by researchers observed tumour cells encapsulated by fibrin and platelets within the vasculature (Crissman et al., 1985, Sindelar et al., 1975). Tumour cells coated with platelets evade immunosurveillance as they are shielded from circulating natural killer cells, which have the capacity to mediate tumour cell lysis upon direct contact.

1.10 Arrest and extravasation at target site

The lifespan of circulating tumour cells in the circulation is unclear (Valastyan and Weinberg, 2011). It has been postulated that following intravasation most circulating

tumour cells will become trapped in capillaries within several minutes given the large diameter of tumour cell (20-30μm) in comparison to capillaries (~8 μm) (Valastyan and Weinberg, 2011). Once trapped in the capillary bed in a distant target organ, circulating tumour cells may penetrate the endothelial layer that separates the vessel lumen from the stromal microenvironment through a process known as extravasation. Unlike the vasculature in the microenvironment of the primary tumour, which is tortuous and permeable, the integrity of the microvascular system at distant organs will be intact and functional. To penetrate the microvasculature tumour cells have developed methods to overcome this obstacle. In breast cancer, TGF-β promotes the expression of angiopoietin-like 4 (*ANGPLT4*) which can dissociate vascular endothelial junctions resulting in an infiltration of tumour cells in to the lung parenchyma (Padua et al., 2008).

1.11 Proliferation and Colonisation

Tumour cells that have successfully extravasated have to adapt to the foreign inhospitable environment of the distant target organ which is often dissimilar to the primary tumour microenvironment (Psaila and Lyden, 2009). Although epithelial cells in the primary tumour are perceived to undergo a process of EMT, tumour cells in the metastasis are frequently differentiated and morphologically resemble the original primary tumour. To account for these findings, it has been proposed that tumour cells have a bidirectional plasticity (Lamouille et al., 2014). They purportedly can interchange between EMT and mesenchymal-epithelial transition (MET) with tumour cells undergoing MET at the distant organ. The regulatory processes governing MET are unclear.

The proponents of the EMT process in cancer have been challenged (Tarin, 2005, Thompson et al., 2005). The conceptual framework of EMT occurring in tumours has largely been derived from findings from *in vitro* studies. Opponents have argued that a distinct change in the morphology of the cell resulting in a definitive change in cell

lineage is arguably absent from observations of clinical specimens. They argue that although tumour cells can be appear dysmorphic these changes do not constitute a change in cell lineage - though it may be difficult to catch them “in the act”.

1.12 Inefficiency of the metastatic process

The process of metastasis is inherently inefficient, with only a few tumour cells overcoming the physical and biological obstacles to successfully develop into macrometastases (Chambers et al., 2002). A key study undertaken by Fiddler and Lieber to identify and quantify metastatic cells estimated that < 1% of tumour cells will proceed to develop into metastases (Fidler and Lieber, 1972). Radioactive labelled living B16 melanoma cells were intravenously injected into mice which were killed at various time points following injection. Several prerequisite criteria were fulfilled by the radioactive label in this study. The radioactive label was bound to the cell whilst it was viable. Upon cell death the radioactive label was released, excreted from the organism and poorly reutilised by normal cells. Moreover, sufficient sensitivity was required to detect a small number of cells *in vivo*. Using this approach, organ radioactivity represented the number of viable tumour cells at different time intervals. They reported that <1% of cells were viable after 24 hours.

1.13 Chronological models defining the metastatic process

Our current understanding of the metastatic process does not enable us to forecast which tumour cells originating from the primary tumour will successfully develop into clinically overt outgrowths. Tumour cells which do form clinically detectable metastases have acquired selectively advantageous traits to proliferate in the hostile foreign environment of the metastatic site. The temporal relationship between the acquisition of these advantageous traits and the time point at which dissemination of metastatic founder cells from the primary tumour occurs has been contentious. Two prevailing

models have been proposed to describe the chronology of events involving the dissemination of metastatic founder cells (Klein, 2009).

1.13.1 The linear progression model

The linear progression model proposes that metastatic founder cells detach from the primary tumour at an advanced stage, in which the primary tumour has established full features of malignancy (Klein, 2009). The model predicts that tumour cells which proceed to develop into metastases will have analogous genetic traits to those found within the primary tumour. Tumour cells within the primary tumour undergo sequential rounds of mutation and selection, to procure advantageous traits and undergo a series of clonal expansions and proliferate (Greaves and Maley, 2012). The model predicts that founding metastatic tumour cells detach from the primary tumour with a fully malignant phenotype. It has been postulated that once metastatic founder cells have successfully adapted and colonised the distant organ, founding cells at the original metastatic sites can subsequently generate further secondary metastases (Weinberg, 2008). This process could continue until systemic progression from malignancy is restricted by the degree of tumour burden encompassed, where the tumour burden ultimately becomes fatal (Friberg and Mattson, 1997).

1.13.2 The parallel progression model

The parallel progression model posits that metastatic founder cells can detach from the primary tumour at an earlier stage in the development of the primary tumour, before the acquisition of a fully malignant phenotype (Klein, 2009). Disseminating tumour cells (DTC) will subsequently acquire advantageous genetic/epigenetic alterations in order to adapt to the niche environment of the target organ. The genomic alterations in the metastatic founder cells occur independently and in parallel to the original primary tumour. The model predicts a greater degree of genetic divergence between the primary tumour and the metastasis in comparison to the linear progression model. The parallel progression model postulates that founding metastatic tumour cells can detach

from the primary tumour at different intervals and that tumour ontogenesis is achieved in parallel.

1.14 Clinical evidence supporting linear/parallel progression models

The relevance of the linear and parallel progression models with relation to clinical practice is emphasised when patients receive treatment based on targeting specific molecular traits found within the primary/metastatic tumour. Evidence supporting and contesting both models are outlined below.

1.14.1 The relationship between tumour size and molecularly advanced cancers

The linear progression model predicts that a fully malignant repertoire is present at an advanced stage in the primary tumour. In support of this proposal, it has been observed that in breast cancer, mutations in the gene *TP53* are seldom found in tumours that are <2cm in size (T1 stage- in accordance with the TNM classification). In comparison, mutations in *TP53* occur more frequently in tumours that are >5cm in size (T3 stage) (Olivier et al., 2006). This finding suggests that the acquisition of mutations in the tumour suppressor gene *TP53* is related to the size of the tumour. This observation is in keeping with the association between tumour size and survivorship (Carter et al., 1989, Narod, 2012). Collectively, these observations imply that surgical resection of the tumour at an early stage, could potentially prevent metastatic dissemination by interrupting the process of clonal expansion.

Several clinical observations are inconsistent with the linear progression model. Firstly 3-5% of malignant epithelial tumours present as a carcinoma of unknown primary where a patient presents with a malignancy with no known detectable primary tumour (Pavlidis and Pentheroudakis, 2012). Moreover, a small but important group of patients

can present with metastatic disease when the primary tumour is small (not exceeding 2cm in size, stage T1M1). The parallel progression model can accommodate for these clinical observations with greater fidelity. In this model metastatic founder cells can detach from the primary tumour at an early stage, even before the primary tumour is clinically detectable.

The parallel progression model which accommodates early dissemination of tumours cells can be inconsistent with clinical concepts. Tumours have significantly improved survival rates when they are staged early (Heintz et al., 2006, Goldstraw et al., 2007). This forms the cornerstone of the tumour, nodes, metastasis (TNM) classification in tumours. The improved survival rates with early surgical intervention do not readily conform with the parallel progression model.

1.15 Genetic evidence supporting parallel/linear progression models

1.15.1 Characterisation of disseminated tumour cells

Researchers have characterised genomic alterations in DTC in patients without clinically detectable metastasis. DTC have historically been identified as cells staining for epithelial markers including epithelial cytokeratins or the epithelial cell adhesion molecule EPCAM (Klein, 2009). Cytokeratin positive cells within the bone marrow are rare in patients without clinically detectable metastasis but have been isolated prior to patients undergoing surgery for their primary tumour. Comparative genomic hybridisation (CGH) in breast and prostate cancer demonstrated fewer genetic aberrations in the DTC in comparison to matched primary tumours (Weckermann et al., 2009, Schmidt-Kittler et al., 2003, Schardt et al., 2005). Unlike primary tumours which demonstrated typical chromosomal genomic aberrations associated with tumour type, DTC karyograms often appeared normal. DTC originating from breast cancer exhibited subchromosomal deletions typical for the breast cancer phenotype confirming their

malignant origin, implying that these tumours had disseminated before a complete repertoire of genome wide instability had been acquired (Schardt et al., 2005).

Importantly in patients with ductal carcinoma *in situ*, a non-invasive lesion, DTC have been found in the bone marrow, questioning when tumour cells disseminate in the development of breast cancer (Klein, 2009). These findings have been recapitulated using (MMTV)-Erbb2 and MMTV-PyMT transgenic mouse models, which develop spontaneously arising mammary cancers (Husemann et al., 2008). DTC were identified in bone and lung tissue before the primary tumour became morphologically invasive.

1.15.2 Concordance of driver mutations

Molecular analysis of driver genes has revealed variable degrees of concordance between primary and metastatic tumours in colorectal and lung cancer. Mutations in the *KRAS* gene have been found to be discordant in 50% of colorectal cancers and 75-80% of lung cancers (Artale et al., 2008, Oudejans et al., 1991). Mutation analysis of the *EGFR* gene in paired primary and metastatic lung adenocarcinomas have reported discordance in 75% of cases (Gow et al., 2009). Taken together, these findings demonstrate the existence of heterogeneity between primary tumours and metastases contradicting the predictions of the linear progression model where the metastasis and primary tumour should share key overlapping genetic aberrations. In contrast the disparity in genetic aberrations between the matched primary tumour and metastasis is accommodated by the parallel progression model which permits dissemination of the metastatic founder cell at an earlier stage and parallel and independent evolution of the primary tumour and metastasis.

1.15.3 Convergent evolution

The parallel progression model can account for the genetic divergence observed in paired primary and metastatic founder cells through the accumulation of genetic traits by the adaptation of tumour cells independently in response to selection pressures in

different niches. Despite the independent and parallel evolution of the primary tumour and metastasis, remarkably a degree of similarity manifests at the gene level. One hypothesis to explain the dichotomous divergence and similarity between the primary tumour and metastasis is that genetic traits acquired by the metastatic founder cells occur within innate constrained mutational selection pressures resulting in convergent evolution of genomic aberrations within tumour cells (Klein, 2009). The existence of 'mutation hotspots' in tumours may support this concept with 'Darwinian like' mutation selection leading to the acquisition of identical mutations. Several examples of mutation hotspots exist. Mutations in the GTPase *KRAS* are present in ~40% of patients in colorectal cancer (Amado et al., 2008, Forbes et al., 2011). *KRAS* mutations are frequently found in codons 12, 13, 61 leading to constitutive activation of the GTPase and subsequent stimulation of downstream pathways (RAS-RAF-MEK-ERK-MAP kinase) (Ellis and Clark, 2000, Smith et al., 2010). *BRAF* which encodes a serine-threonine protein kinase has been implicated in several malignancies (Davies et al., 2002). *BRAF* mutations in codon 600 results in a substitution of valine to glutamic acid resulting in increased kinase activity.

1.16 Tissue tropism of metastases

1.16.1 Seed and soil hypothesis

The predilection for tumours of particular histological subtypes to colonise specific target organs has been well documented (Nguyen et al., 2009a). Metastatic patterns of growth have been observed from autopsy studies and data based on clinical evaluations from clinically detectable metastases (Hess et al., 2006). Ocular melanoma has the propensity to metastasise to the liver (Triozzi et al., 2008). In comparison ovarian and prostate malignancies predominantly metastasise to the abdominal cavity and bone respectively (Hess et al., 2006). Stephen Paget proposed the seed and soil hypothesis in 1889 following his observations of autopsies in patients with advanced breast cancer, noting that the distribution of metastases was not reflective of the blood

supply to organs (Paget, 1989). He posited that tumour cells (the seeds) could only successfully colonise and develop into overt metastases if the tumour microenvironment in the target organ (the soil) was receptive for growth. This hypothesis was subsequently challenged by James Ewing who postulated that the distribution of metastatic outgrowth was determined purely by haematogenous circulation (Talmadge and Fidler, 2010). Our current understanding is that although the vascular distribution may have a partial effect, biological processes have a key role in achieving colonisation. This was first substantiated in a seminal study undertaken by Fidler et al (Hart and Fidler, 1980). B16 melanoma cells injected into syngeneic mice developed metastases in the lung in situ and in implanted pulmonary and ovarian grafts. However, neoplastic lesions failed to develop when a renal graft was implanted or at the site of surgical trauma. An observational study in patients with advanced ovarian cancer with peritoneovenous shunts to reduce ascites noted that patients rarely developed secondary tumours despite a direct connection for tumour cells to enter into the vena cava (Tarin et al., 1984).

1.16.2 Factors influencing organ specific colonisation

The pattern of colonisation of metastatic tumour sites is likely to be governed by a number of factors. Firstly, the haemodynamic circulation will influence the dissemination of tumour cells around the body. Common sites for the development of metastasis involve the lungs, bone marrow, liver and brain. The aforementioned sites have unique environments and distinct physical barriers that need to be successfully overcome by DTC to form macroscopic metastases. The tissue architecture of the bone marrow and liver is composed of sinusoid capillaries, which are fenestrated. The organisation of the capillary structures within these organs facilitates the movement of haematopoietic cells. However this structural organisation also creates a relative permissive environment for tumour cells to invade. In contrast the lung and brain have environments that are less conducive to colonisation. Capillaries within the lung are

lined by endothelial cells that are surrounded by a basement membrane and alveolar cells. The basement membrane forms a physical barrier which needs to be penetrated for tumour cells to enter the lung parenchyma. Transendothelial migration by tumour cells requires the tumour cells to express selective mediators. Invasion of the brain by tumour cells is restricted by the blood brain barrier (BBB) which forms a highly selectively permeable barrier. In recent years there is increasing evidence to suggest that specific biological mediators have a contributory role in the organ specific colonisation of tumour cells.

The mechanisms governing the process of metastatic colonisation in the bone is well characterised in the development of osteolytic lesions in breast cancer. Breast cancer cells derived from a metastatic bone site express higher levels of parathyroid-hormone-related-peptide (PTHrP) compared to breast cancer cells originating from breast and other soft tissue sites.(Powell et al., 1991) Upregulated secretion of PTHrP in concert with interleukins released by breast cancer cells results in bone resorption. A feedback mechanism that stimulates tumour proliferation and the formation of macrometastasis is established (Mundy, 2002).

The model for osteolytic metastasis in breast cancer is depicted in Figure 1.4. Expression of osteoclast activating factors at another metastatic site or in the primary tumour is unlikely to provide a selective advantage for breast tumour cells (Nguyen et al., 2009a). Despite our insight from bone metastasis in breast cancer, the mechanisms driving metastasis to other target organs remains poorly understood.

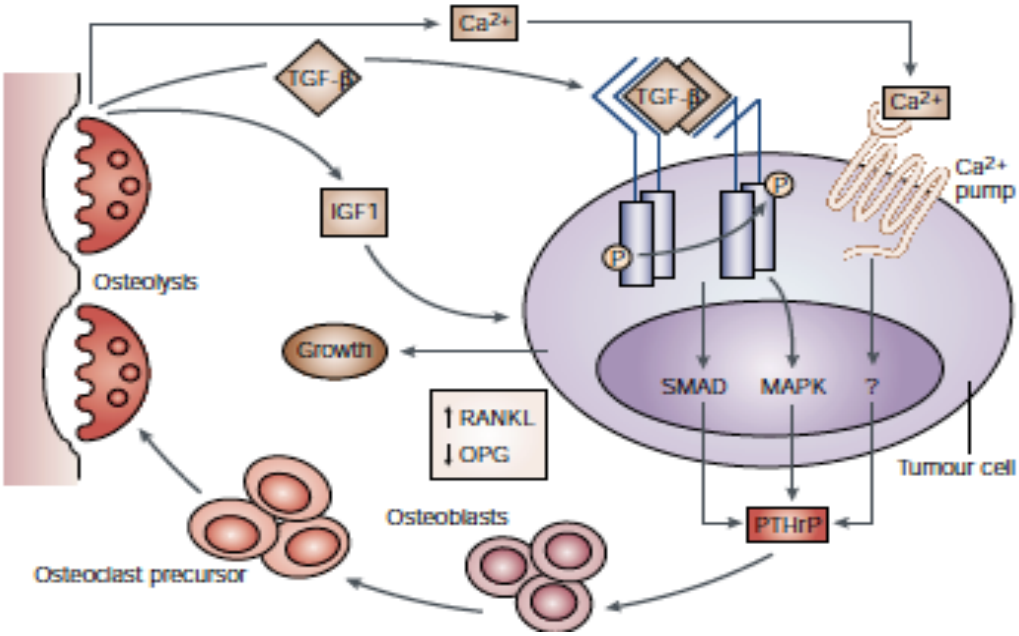


Figure 1.4 l Model for the development of osteolytic metastasis. Image derived from Mundy (2000). Breast tumour cells produce PTHrP and interleukins leading to activation of osteoblasts to produce RANKL (receptor activator of nuclear factor–κB ligand) and down regulation of osteoprotegrin (OPG). This results in the activation of osteoclasts and subsequent osteolysis. As a consequence of osteolysis, transforming growth factor (TGF-β) and insulin-like growth factor 1 (IGF1) are released and the extracellular calcium levels rise. The growth factors bind to the tumour cell resulting in autophosphorylation and signaling through SMAD and mitogen-activated protein kinase (MAPK). This promotes tumour cell proliferation and further PTHrP release resulting in a positive feedback loop

1.17 Tumour latency

Cancers may relapse locally at the primary site or with distant metastasis. The dynamics underpinning why tumours vary greatly in their recurrence free interval is unclear. Lung adenocarcinoma and pancreatic malignancies often relapse swiftly with aggressive disease. In contrast patients with oestrogen positive breast cancer may remain in remission for several years before they relapse (Nguyen et al., 2009a). Studies in breast cancer have demonstrated DTC in the bone marrow in patients with no macroscopic metastatic disease. Two modes of dormancy have been proposed either proliferative and tumour mass (Sosa et al., 2014). In proliferative dormancy tumour cells exit the cell cycle, whilst in tumour mass dormancy tumour cells grow slowly with the rate of cell death counterbalancing the rate of cell division.

1.18 Brain Metastasis

1.18.1 Demographics of brain metastases

Metastases to the brain are more frequent than primary intracranial tumours outnumbering them by 10 fold (Patchell, 2003). The prevalence of brain metastases in cancer patients varies widely and has been estimated to be 10-20% (Schouten et al., 2002, Barnholtz-Sloan et al., 2004). It is likely that the reported prevalence is an underestimate as patients with asymptomatic metastases may go undetected (Owonikoko et al., 2014). In support of this, the prevalence of brain metastases has been found to be higher in autopsy studies of patients with known primary tumours in comparison to the reported prevalence in population based studies (Cascino et al., 1983, Tsukada et al., 1983). The distribution of metastases within the brain appears to be proportional to the vascular supply, with 80%, 15%, and 5% being found in the cerebral hemispheres, cerebellum and brain stem respectively (Norden et al., 2005). In adults, the most common primary tumours responsible for brain metastases include malignancies of lung, breast, renal and colorectal origin as well as melanomas (Barnholtz-Sloan et al., 2004). Delattre et al reported that metastases from the colon, breast and kidney are more likely to be single whereas melanoma and lung cancer frequently presents with multiple metastases (Delattre et al., 1988). The type of primary tumour associated with brain metastases has altered over the last few decades which may reflect the incidence and mortality rates of individual tumour types. One study by Nider et al compared two cohorts of patients from the years of 1983-1989 and 2005-2009 (Nieder et al., 2011). They reported that the incidence of patients presenting with brain metastases with a primary breast tumour remained stable (17%). The number of patients with brain metastases with a lung primary reduced (52% versus 40%). However the number of patients presenting with brain metastases with colorectal renal and melanoma primary tumours increased in numbers (8% versus 24%). The total number of patients presenting with brain metastasis is increasing in number. Several factors are likely to be contributory for this increased prevalence. Improvements in

treatments for cancer has resulted in prolonged survival rates for patients, resulting in more patients presenting with brain metastases who may have previously died from their extracranial disease. Advances in imaging techniques coupled with the greater availability of imaging resources have led to the identification of more cases (Davis et al., 1991, Schaefer et al., 1996, Nieder et al., 2011).

1.18.2 Clinical presentation of brain metastasis

As a brain metastasis continues to grow in size, cerebral oedema, mass effect, or increased bleeding risk, can result in the development of symptoms. These include: headaches, seizures, cognitive impairment and focal neurological deficits including, aphasia, hemiparesis or visual field defects. The symptoms experienced by patients can be debilitating (Pekmezci and Perry, 2013).

1.18.3 Prognostic factors associated with brain metastasis

The treatment approach adopted is influenced by the size, number and accessibility of the brain metastases, the degree of mass effect, performance status and the presence of extracranial disease. Relative good prognostic factors for patients with brain metastases include age <65 years, good performance status, the presence of a single metastasis, a controlled primary tumour and certain histological types of cancer which may be amenable to targeted therapy such as an *EGFR* mutant lung adenocarcinoma (Eichler et al., 2011).

1.18.4 Therapies for brain metastasis

Surgery, whole brain radiotherapy, stereotactic radiosurgery or a combination of these treatment modalities have formed the principal treatment options for patients with brain metastases (Owonikoko et al., 2014). Surgery may be indicated in carefully selected patients when there is a limited number of brain metastases and limited extracranial disease, when immediate decompression from mass effect is necessary, or when histological confirmation is required for a definitive diagnosis. SRS delivers high doses

of radiotherapy using convergent beams to a discrete target (Lippitz et al., 2014). There is sharp fall in dose at the edge of the treated target, limiting damage to adjacent normal tissue. SRS may be an appropriate treatment option in patients with a limited number of small metastases <3cm in size, in patients with tumours that are inaccessible for surgical resection and for patients that are not candidates for surgery. Patients with brain metastasis treated with best supportive care with steroids alone may have a median survival of 1-2 months. The use of whole brain radiotherapy (WBRT) has extended survival to several months (Gaspar et al., 1997). WBRT may be the treatment modality of choice for patients with extensive intracranial disease or a poor performance status. Numerous studies have evaluated the role of WBRT following definitive treatment such as surgery or SRS (Kocher et al., 2011, Aoyama et al., 2006, Sneed et al., 2002, Patchell et al., 1998). Although adjuvant WBRT following definitive treatment reduced the rate of recurrence of brain metastases there was no impact on survival in comparison to patients who received definitive treatment alone.

1.18.5 Targeted therapy for brain metastases

Advances in the understanding of tumour biology has uncovered several molecular drivers of tumourigenesis in different tumours types. Researchers have noted that particular molecular drivers may be represented more frequently in brain metastases when compared to primary tumours. Sun et el reported that brain metastases expressed increased levels of epidermal growth factor (EGF) and phosphorylated EFGR in comparison to primary tumours (Sun et al., 2009). Patients with HER2-positive breast cancers (*ERBB2* amplification) have a propensity to develop brain metastases (Lin and Winer, 2007b, Leyland-Jones, 2009). *BRAF a*nd *NRAS* mutations in melanoma are associated with a proclivity to develop brain metastases (Davies et al., 2009). These findings have initiated interest through retrospective case observations and early phase studies treating patients with targeted therapies (Owonikoko et al., 2014). Patients with non-small cell lung cancer have demonstrated responses to EFGR

inhibitors erlotinib and gefitinib (Ceresoli et al., 2004, Wu et al., 2007, Welsh et al., 2013, Grommes et al., 2011). In breast cancer response rates ranging from 20-30% have been observed with a combination of laptinib, a tyrosine kinase inhibitor, and capecitabine (Sutherland et al., 2010). Early phase studies have demonstrated a response to targeted agents in melanoma (Falchook et al., 2012, Long et al., 2012).

1.18.6 Exclusion of patients with brain metastasis from clinical trials

Patients with brain metastases have historically often been excluded from clinical trials (Carden et al., 2008, Gilbert, 2001). Numerous factors are responsible for the paucity of clinical trials incorporating this group of patients. Firstly patients with brain metastasis may have a poor prognosis or poor performance status and these factors alone may lead to exclusion. Patients with brain metastasis often have received several lines of treatment and may have systemic disease which is less likely to be responsive to treatment due to the acquisition of resistance in response to the previous treatment exposure. Historically the enrolment of patients to phase 1 trials has been hindered due to concerns of assessing toxicity of drugs in patients who may develop complex neurological problems, alongside reservations regarding the possible interactions with antiepileptic medication (Gilbert, 2001). In general they form a group of patients with a poor prognosis that currently have an unmet clinical need for therapeutic strategies.

1.19 Blood Brain Barrier

The blood brain barrier (BBB) is a highly specialised dynamic, selectively permeable barrier which is responsible for maintaining homeostasis within the brain microenvironment separating the systemic circulation from cerebrospinal fluid (Abbott et al., 2006). The BBB facilitates the delivery of nutrients to the brain and the efflux of waste products. Moreover, it restricts the movement of fluid and ions regulating the ionic and osmolar composition within the brain microenvironment. The BBB is principally composed of three structures, namely cerebrovascular endothelial cells,

astrocytic perivascular end-feet and pericytes (Ballabh et al., 2004). Endothelial cells in the BBB form tight junctions which restrict the paracellular movement of hydrophilic molecules across the BBB. Although small molecules such CO_2 and O_2 can freely diffuse across the BBB, the transcellular movement of small hydrophilic molecules is regulated by selective transporters. Large hydrophilic molecules have limited capability to penetrate the BBB. The BBB expresses P-glycoprotein which is a multi-drug resistance protein which can actively efflux some chemotherapeutic drugs (Neuwelt, 2004). To be effective, chemotherapeutic agents have to be delivered to the tumour within a meaningful time frame and at sufficient concentration. In a series of 2000 patients, Lockman et al reported that the BBB was compromised in 89% of patients and cytotoxic concentrations of paclitaxel and doxorubicin were achieved in only 10% of patients (Lockman et al., 2010). Disruption of the BBB is not homogenous and large regions may remain intact and functional.

1.20 Understanding the molecular mechanisms underpinning brain metastasis

The mechanisms underpinning why certain tumours have a propensity to develop brain metastases remains unclear. In recent years, researchers have aimed to identify candidate genes which may serve a role in the process of the formation of brain metastases. Different strategies have been employed to the unravel genes associated with metastasis to the brain.

1.20.1 The role of host stromal cells in the brain

Having entered the brain parenchyma tumour cells encounter host cells that are specific to the brain, namely astrocytes and microglia, which may facilitate colonisation in the brain. Evidence from xenograft models and human specimens of brain metastases have demonstrated that microglia and astrocytes are in direct contact with tumours along the border of tumour and even infiltrating the inner tumour cell mass (Lorger and Felding-Habermann, 2010, Fitzgerald et al., 2008). Several studies have

reported that astrocytes may act to protect growing tumour cells from cytotoxic chemotherapeutic drugs (Langley et al., 2009, Lin et al., 2010, Kim et al., 2011). Langley et al analysed the role of astrocytes in tumour growth in co-culture experiments. The presence of astrocytes reduced the incidence of cisplatin and 5-fluorouracil induced apoptosis in tumour cells. The protective role provided by astrocytes is likely to be mediated through direct contact rather than soluble factors as placement of a trans-well membrane to separate the tumour cells from the astrocytes resulted in the loss of this protective effect.

1.20.2 The role angiogenesis in the formation of brain metastasis

The formation of metastatic outgrowths necessitates the establishment of a vascular supply. Interest in understanding and potentially disrupting this process has been a focus of research. Kienast et al visualised brain metastasis formation in vivo in relation to blood vessels using multiphoton laser scanning microscopy (Kienast et al., 2010). Melanoma cells grew through co-option with the growth of cancer cells along existing vessels, whilst lung carcinoma cells grew in a process which required the sprouting of new vessels. The continued growth of brain tumours may be dependent on the expression of vascular endothelial growth factor (VEGF). Kim et al demonstrated a significant reduction in brain tumour burden in mice following inhibition of the vascular endothelial growth factor receptor (VEFGR) (Kim et al., 2004). However, Juan Yin et al undertook a study evaluating the role of VEGFR inhibition in a murine model using cediranib a VEGFR inhibitor which failed to prevent extravasation of the tumour cells in a prostate cancer metastasis model (JuanYin et al., 2009). Yano et al demonstrated that lung carcinoma cells transfected with antisense VEGF165 injected into the carotid artery of nude mice resulted in reduced numbers of brain metastases developing in comparison to untransfected cells (Yano et al., 2000). Collectively, these findings demonstrate that there is heterogeneity in the response of tumour cells to VEGF. VEGF

expression may be required but is not sufficient alone for the formation of brain metastases.

1.20.3 Comparative analysis of gene expression in humans

Studies in humans analysing the genetic profile of brain metastasis have focused on comparative analysis of gene expression. Grinberg-Rashi *et al* quantified the RNA expression of a panel of genes hypothesised to be involved in metastasis using non-small cell lung cancer brain metastases. They found that *CDH2,* which encodes for N cadherin, a calcium dependent cell adhesion molecule, *KIFC1* a mitotic regulator and *FALZ3 a* neuronal transcription factor, were associated with brain metastasis in early and advanced non small cell lung carcinoma based on the expression level of these genes (Grinberg-Rashi et al., 2009). Ray et al reported the overexpression of transcription factor FOXC1 in basal-like breast cancer was associated with poor overall survival and a higher incidence of brain metastases (Ray et al., 2010). Moreover, the ectopic overexpression of the gene in breast cancer cells resulted in increased cell migration and invasion whilst shRNA-mediated knockdown resulted in the opposite effect. Hexokinase 2 (HK2) has an important role in glucose metabolism. Palmieri et al reported that HK2 was overexpressed in resected brain metastases of breast origin (Palmieri et al., 2009). Overexpression of the gene was associated with poor survival following craniotomy.

1.20.4 Animal models of brain metastasis

Mouse models have been used in recent years to understand the biological process of metastasis *in vivo*. Bos *et al* identified 3 genes that may have a role in breast cancer metastasis to the brain using a mouse model (Bos et al., 2009) Pleural malignant breast cancer cells were injected into the arterial circulation of immunodeficient mice. These mice develop metastases at multiple sites. Tumour cells colonising the brain were isolated and underwent a second round of *in vivo* selection to yield cells with significant brain metastatic activity. Gene expression profiles were analysed and genes known to

be associated with brain relapse were prioritised. The genes that were reported were *COX2,* which is necessary for prostaglandin production, which can increase the permeability of the blood brain barrier during inflammation, *HBEGF* an epidermal growth factor receptor ligand and *ST6GALNAC5* a sialytransferase expressed in the brain which catalyses the addition of sialic acid to gangliosides and glycoproteins. This sialytransferase may have a role in cell interactions. Brain metastatic activity was reduced by RNA interference mediated knockdown of *COX2* and *ST6GALNAC5.* Treatment of mice with cetuximab which targets the EGFR also reduced brain metastatic activity.

Recently, Valiente et al reported that plasmin in the brain stroma acts a defensive mechanism to prevent invasion in a mouse model using lung and breast cancer cells (Valiente et al., 2014). Plasmin suppressed metastatic activity by inactivating L1CAM which serves as a path finding molecule, resulting in disruption of the normal colonisation pattern observed in the brain. Plasmin also promotes tumour cell death by activating membrane bound astrocytic Fas ligand. They demonstrated that human brain metastatic cells expressed serpins which inhibit plasminogen activator suppressing plasmin activity which was required for vascular co-option. Sevenich et al proposed that capthesin S mediated brain specific metastasis in a murine model (Sevenich et al., 2014). Capthesin S is produced by both tumour cells and macrophages. Tumour derived capthesin S was expressed higher in the early stages of brain metastasis and decreased in the later stages. However the inverse was true of stromal derived capthesin S. The combined depletion of capthesin S was required to reduce the brain metastatic burden *in vivo*. Genetic and pharmacological depletion of capthesin S resulted in reduced transmigration of tumour cells across the BBB. Moreover, Sevenich et al confirmed that junctional adhesion molecule B (JAM-B), an organ specific component of the BBB acted as a substrate for proteolytic cleavage.

1.20.5 Limitations of mouse models

Although the mouse models have implicated several genes that may be involved in metastasis, there are number of disadvantages including the degree of fidelity when extrapolating the biological processes to humans. A key problem is that mice that have orthotopically transplanted tumours which give rise to metastatic outgrowths die from their systemic disease before the brain metastasis can be studied reliably (Eichler et al., 2011). To overcome this obstacle, two approaches have been adopted: either direct implantation or iterative haematogenous metastasis models. In direct implantation models, tumour cells are directly injected into the brain parenchyma or into the subarachnoid space to disseminate through the cerebrospinal fluid. Inoculation of tumour cells directly into the brain results in the formation of cerebral metastases, whereas inoculation into the subarachnoid space results in leptomeningeal metastasis. Haematogenous models of brain metastasis have evolved by selecting clones that preferentially develop brain metastasis following haematogenous dissemination in an iterative process. Tumour cells can be injected through the internal carotid artery followed by ligation of the artery or through an intracardiac injection. A common criticism of these approaches is that the processes in the early steps of the metastatic cascade may not be reproduced faithfully (Eichler et al., 2011). Genes that have been associated with metastasis to the brain are represented in Table 1.2.

1.21 Next generation sequencing

Advances in sequencing technology coupled with reducing costs and improvements in bioinformatic software has provided researchers an unrivalled opportunity to undertake high throughput analysis of genomes (Shendure and Lieberman Aiden, 2012). In the era of personalised medicine, next generation sequencing (NGS) promises to revolutionise oncology. The technology is likely to influence the design and use of targeted drugs, aid predictions in outcomes and response to treatment and classify

Gene	*Function*	*Tumour type*	*Comments*
COX2	Important in prostaglandin production, possibly leading to increased permeability of BBB	Breast	Enriched in breast brain metastasis murine model (Bos et al., 2009).
HBEGF	EGFR ligand—increases cell growth, motility, and invasiveness	Breast	Enriched in breast brain metastasis murine model (Bos et al., 2009).
ST6GALNAC5	Sialyltransferases catalyze the addition of sialic acid to gangliosides and glycoproteins, and cell-surface sialylation has been implicated in cell–cell interactions	Breast	Enriched in breast brain metastasis murine model (Bos et al., 2009).
HK2	Role in glucose metabolism, oxidative phosphorylation, and antiapoptosis	Breast	*HK2* expression is associated with poor patient survival after craniotomy (Palmieri et al., 2009)
FOXC1	Transcription factor involved in brain development	*Breast*	Associated with poor overall survival in basal-like breast cancer and a higher incidence of brain metastasis (Ray et al., 2010)
HER2	Tyrosine kinase receptor	Breast	Overexpression increased the incidence of large brain metastases (Palmieri et al., 2007)
CTSS	Capthesin S acts as a lysomal cysteine protease	Breast	Capthesin S cleaves junctional adhesion molecule B perturbing the integrity of the BBB (Sevenich et al., 2014)
VEGFA	Angiogenic growth factor	Breast	VEGFR inhibition decreased brain tumor burden via a reduced number of blood vessels, decreased proliferation and increased apoptosis *(Kim et al., 2004)*
Serpin 1	Encodes for neuroserpin	*Breast and lung*	Inhibits plasminogen activator to suppress plasmin production and foster vascular co-option (Sevenich et al., 2014)
LEF1	A transcriptional effector of the canonical WNT pathway	*Lung*	Associated with lung metastasis to the brain; knockdown inhibited brain metastasis (Nguyen et al., 2009b)
CDH2, KIFC1, and FALZ3	CDH2: N-cadherin is a calcium-dependent cell–cell adhesion molecule *KiFC1* microtubule dependent motor *FALZ3* neuronal transcription factor	Lung	Associated with brain metastasis in lung cancers, causal role is not clear *(Grinberg-Rashi et al., 2009)*
STAT3	Important transcription factor in cellular signaling pathways	*Melanoma*	Reduction suppressed brain metastases, affected angiogenesis *in vivo* and cell invasion *in vitro* (Xie et al., 2006)

Table 1.2 I Overview of the genes that have been implicated with brain metastases (Table adapted from Eichler et al 2011)

tumours based on molecular profiling rather than by morphology alone (Ulahannan et al., 2013). Technically NGS technology is continually evolving, breaking new boundaries with the goal of generating data faster, whilst maintaining accuracy with reduced quantities of input DNA. The field has expanded rapidly resulting in the development of several platforms utilising different sequencing technologies to sequence the entire genome, whole exome, or targeted regions (van Dijk et al., 2014, Mardis, 2013).

1.22 Complexity of cancer genomes

Cancers are biologically diverse heterogeneous entities with complex genetic alterations (Stratton et al., 2009, Yates and Campbell, 2012). One may consider the simplest alteration being a somatic base mutation. However, from this point onwards the complexity increases considerably with mutations resulting from insertions and deletions in the genome, epigenetic modifications, focal/large-scale changes in copy number across the genome, and structural variations such as inter/intra chromosomal translocations. Widespread genomic aberrations can occur in a narrow time window, unlike the traditional model in which successive rounds of selection of advantageous mutations and clonal expansion occurs driving tumour progression. Chromothripsis purportedly occurs in a single event from shattering of the genome and is typified by an oscillating pattern of copy number and loss of heterozygosity (LOH) events present in a single or a few chromosomes (Stephens et al., 2011). Chromoplexy is a phenomenon which results in complex DNA rearrangements throughout the genome (Shen, 2013). The underlying mechanism is unknown, though the process is generally differentiated from chromothripsis based on that it can affect 5 or more chromosomes and can occur in multiple events rather than a single event. The analytical complexity in cancer genomes is further compounded by heterogeneity if present from different clonal populations

1.23 Practical considerations

Surgically excised tumours will have some degree of contamination from normal tissue. Ideally this should be minimised as this may hinder the identification of true somatic variants. Collaboration with a histopathologist is important to ensure that the specimens being analysed are truly representative of tumour. Most archived clinical specimens are formalin fixed and paraffin embedded (FFPE) where nucleic acids can undergo crosslinking and degrade (Gilbert et al., 2007) The processing of FFPE specimens and aging has historically hindered reproducible sequencing analysis of entire genomes. Encouragingly, researchers are attempting to overcome this paradigm by sequencing targeted regions from FFPE specimens (Kerick et al., 2011, Wong et al., 2014).

1.24 NGS analysis of cancer genomes

NGS has been used to characterise a variety of cancers including lung, colorectal and breast cancers (Imielinski et al., 2012, Cancer Genome Atlas Research, 2014, Banerji et al., 2012). Nascent projects in cancer genome analysis involved sequencing a few tumour samples. To identify recurrent mutations of potential significance in primary tumours, international efforts from large-scale consortiums such as The Cancer Genome Atlas (TCGA) (http://cancergenome.nih.gov/) and international genome consortium (http://icgc.org) were established. They sought to sequence and categorise somatic mutations in selected types of tumour to increase our understanding of the genetic diversity within cancers and facilitate the development of future therapies. Integrative analysis of nucleic acid sequencing data with copy number, methylation profiles and gene expression has been undertaken. In recent years several studies have been published reporting their findings in different subgroups of cancers (Muzny et al., 2012, Koboldt et al., 2012). Furthermore, some datasets are publicly available and form a resourceful comparison dataset of primary tumours for cancer genome analysis projects. The

aforementioned large-scale projects have been predominantly confined to whole exome sequencing and have focused on primary tumours.

1.25 NGS analysis of metastases

Currently there are a limited, though increasing number of often small-scale studies that have served to analyse the genomes of metastases using NGS. The technology has been used to profile the genomes of metastases to determine the genomic disparities that exist between primary and metastatic tumours. Progress has been limited in this field as longitudinal studies comparing matched primary and metastatic tumours at different time points have been constrained by the difficulty in sample acquisition and quality of the stored samples. The findings from NGS of matched analysis of primary tumours and metastases have not conclusively resolved the debate amongst proponents of the linear and parallel progression models with reports to date in support of both models. Turajlic et al reported comparable mutation rates in matched primary and metastatic acral melanomas from a lymph node metastasis (Turajlic et al., 2012a). In this study a sample from the primary tumour and the metastasis were biopsied prior to the patient receiving any therapy. Targeted sequencing of a cancer panel of genes in a cohort of matched colorectal primary and metastatic tumours reported a 100% concordance for *KRAS, NRAS* and *BRAF* mutations (Brannon et al., 2014). Recently Xie et al reported their findings from whole genome sequencing of two matched colorectal and liver metastases reporting between 75-95% similarities in clonal mutations between the primary and metastatic tumour (Xie et al., 2014). An alternative approach to understanding the metastatic process is to sequence metastases from different geographical locations and characterise the degree of heterogeneity that exists (Gerlinger et al., 2012). Table 1.3 summarises the recently published NGS findings from sequencing metastatic tumours.

1.26 NGS of brain metastases

The number of sequencing studies characterising the genomes of brain metastases is sparse. This may be partly attributed to the difficulty in obtaining samples as surgery may not be indicated, the metastasis may be inaccessible and constitutional DNA is required as a control. Moreover, obtaining a matched primary may be difficult due to sample quality or due to difficulty in tracing primary tumours retrospectively. Shah et al sequenced a paired primary and brain metastasis from a lobular oestrogen-receptor-α-positive breast carcinoma (Shah et al., 2009). They reported 32 somatic nonsynonymous mutations in protein-coding mutations. Comparison with the primary tumour demonstrated that 11 (34%) were detected in the primary tumour which was resected 9 years earlier. Ding et al reported findings of NGS sequencing from DNA derived from peripheral blood, primary tumour, basal-like breast brain metastasis and a xenograft derived from the original primary tumour (Ding et al., 2010). The bulk of mutations were shared between the primary tumour and the metastasis (n=20), with only 2 *de novo* mutations arising exclusively in the metastasis. The time interval between the primary tumour and the development of the brain metastasis was 8 months.

Tumour type	Investigator	Number of samples	Type of sequencing	Key findings
Acral melanoma primary and lymph node metastasis	(Turajlic et al., 2012b)	1 pair	Whole genome	
Colorectal primary and liver Metastasis	(Xie et al., 2014)	2 pairs	Whole genome	75-95% of similarity in SNV in comparison of primary and metastasis. Kataegis identified in a primary but not in matched metastasis.
Hepatocellular carcinoma and metachronous lung metastases	(Ouyang et al., 2014)	4 pairs	Whole genome	38-79% of somatic mutations detected in the primary were preserved in the metastases. CNV is the primary and metastases were closely related.
Colorectal cancer and primary tumours	(Brannon et al., 2014)	69 pairs	Targeted sequencing of 230 cancer genes	100% concordance for KRAS, BRAF and NRAS mutations
Breast primary lobular oestrogen-receptor-α-positive and brain metastasis	(Shah et al., 2009)	1 pair	Whole genome	34% of the nonsynonymous mutations found in the metastasis were found in the primary.
Breast primary basal-like tumour and brain metastasis	(Ding et al., 2010)		Whole genome	20 mutations were found shared between the primary and metastasis with only 2 *de novo* mutations found within the metastasis.
Renal clear cell carcinoma spatially distinct regions in primary and metastasis	(Gerlinger et al., 2012)	Multiple spatially distinct sites from 4 individuals	Whole exome	Mutational heterogeneity was demonstrated with spatially separated inactivating mutations implying convergent evolution.

Table 1.3 I Summary of recent NGS projects characterising metastasis

1.27 Hypotheses

1) Metastases originating from different tumour types demonstrate organ specific tissue tropism. I postulate that a common driver mutation may exist which drives tissue tropism to the brain.

2) Brain metastases are generally chronologically advanced in the disease trajectory. I posit that the brain metastases differ from primary tumours with regards to mutation burdens, mutation spectra and copy number aberrations.

1.28 Objectives

a) Evaluating brain metastases genomes with regards to mutation burdens, characterising spectra/signatures, profiling copy number aberrations and modelling clonal architectures.

b) Screening for structural variants, including chromothripsis/chromoplexy and kataegis.

c) To characterise the degree of heterogeneity in brain metastases with respect to primary tumours. TGCA primary tumour data served as a comparison data set.

d) To identify drivers of brain metastases and determine if drivers of primary tumours were retained in the brain metastases.

e) Undertaking WGS on FFPE primary tumours where possible and corresponding brain metastases.

1.29 Outline of thesis

1. Bioinformatic analysis and laboratory preparation are discussed in the materials and methods section, presented in Chapter 2.

2. Chapters 3 to 7 reports the results of the thesis. This involves presentation of the brain metastases by tumour type and an amalgamated analysis of the different histological types of brain metastases in a cohort. In addition, I have reported a case presentation of a paired colorectal brain metastasis and a FFPE derived primary tumour.

3. References and appendices and are presented at the end of the thesis

Chapter 2

Materials and Methods

2.1 Sample description of tumour specimens

2.1.1 Sequenced brain metastases

Fresh frozen brain metastases originating from colorectal (4), ductal breast (7), and lung adenocarcinomas (7) primaries and paired buffy coat lymphocytes were obtained from the National Hospital for Neurology and Neurosurgery (London). FFPE primary tumours matched to the brain metastases were traced retrospectively where possible using an independent party to maintain patient anonymity. A paired FFPE colorectal primary tumour (P1) corresponding to the colorectal brain metastasis C1MET was sequenced successfully.

2.1.2 Single nucleotide polymorphism arrays of brain metastases

All fresh frozen brain metastases that were sequenced were genotyped using single nucleotide polymorphism (SNP) arrays. Additionally, 40 FFPE brain metastases were obtained from the Imperial College, healthcare trust tissue bank (London). Samples that passed quality controls were genotyped using Illumina SNP arrays (Human OmniExpress-24 v1).

2.1.3 Liver metastases

A total of 25 liver metastases were obtained from either the Royal Free Hospital (London) or John Radcliffe hospital (Oxford). Of these tumours, 16 were paired liver metastases from 8 individuals. In total 3 pairs of liver metastases underwent whole genome sequencing (WGS), Complete Genomics (Mountain View, CA). Paired normal liver parenchyma was used to provide constitutional genomic DNA. The liver metastases were collected by a colleague, Dr Rajarajan Ramamoorthy. He extracted

genomic DNA for these tumours for sequencing and SNP arrays. The analysis of data for sequencing and SNP arrays was undertaken by myself. Data from the liver metastases served as a comparative data.

2.2 TCGA data

2.2.1 Somatic mutation data from primary tumours

TCGA somatic exome mutation data were retrieved from colorectal, breast, and lung adenocarcinoma primary tumours from the TCGA cBioPortal (https://tcga-data.nci.nig.gov/tcga/) on the 1st February 2014. This data was utilised to characterise the mutation spectra in primary tumours and were processed using IntOGen and Mutsig to identify driver mutations in primary tumours.

2.2.2 SNP array data from primary tumours

TCGA SNP array raw data were obtained from colorectal, breast, and lung adenocarcinomas. Data were downloaded in July 2013 with the assistance of Dr Christopher Yau. TCGA samples were genotyped on the Affymetrix Genome-Wide SNP Array 6.0 platform (SNP density of ~900,000) build GRCh36. The genome coordinates of the SNPs were converted to build GRCh37 using LiftOver (http://genome.ucsc.edu/cgi-bin/hgLiftOver). Data were processed using OncoSNP to obtain information regarding copy number and LOH.

2.3 Ethical approval

Ethical approval was obtained from a local ethics board, Oxfordshire ethics research committee C (project 09/H0606/5) for the analysis of anonymised samples for driver gene discovery. Clinical information where available, pertaining to patients who had their tumours analysed in this project was obtained using an independent party.

Laboratory procedures

2.4 DNA extraction

2.4.1 DNA extraction of fresh frozen brain metastases

Genomic DNA was extracted using the DNeasy Blood and Tissue kit from Qiagen (Alameda, CA). Fresh frozen brain metastases were homogenised and 60µl of proteinase K (PK) and 240µl of ATL buffer solution was added to each sample. The mixture was vortexed and incubated for 24 hours at 55^{o}C. Following this a further 20µl of PK was added and the samples were mixed by vortexing and incubated for an additional 24 hours. At 72 hours an equal volume of AL buffer solution was added to each sample, which was then mixed by vortexing and incubated for 70^{o}C for 10 minutes. The samples were then loaded on to DNeasy mini spin columns and centrifuged at 8000 rpm for 3 minutes. The flow-through in the collection tubes was discarded. Each spin column was washed using 550µl of AW1 buffer solution and centrifuged at 13000 rpm for 5 minutes. The columns were then washed with AW2 buffer solution and centrifuged at 13000 rpm for 3 minutes (repeated twice). Genomic DNA was eluted with 100µl of nuclease free water. Samples were then centrifuged at 14000 rpm for 1 minute. To increase the yield of genomic DNA this step was repeated after reapplying the eluate to the spin column.

2.4.2 DNA extraction of buffy coat lymphocytes

Buffy coat lymphocytes paired to the fresh frozen brain metastases were resuspended and homogenised in phosphate buffer solution 0.1M. Cells were lysed with 40µl of PK and 120µl of ATL buffer solution for a period of 24 hours at 55^{o}C. Genomic DNA was extracted using the Promega Maxwell 16 Tissue DNA purification kit (Madison, Wisconsin) using an automated process. The purification kit utilises prefilled cartridges with 7 wells (1x lysis buffer, 1x MagneSil PMPs and 5x wash buffer solutions). A plunger was placed into well 7 of each slot. The cartridges were appropriately orientated in accordance with the manufacturer's protocol and placed

into the preparation rack in the Maxwell 16 MDx instrument. The lysed buffy coat lymphocyte samples were loaded into well 1 of each cartridge. An elution tube containing 300µl of elution buffer was loaded in to the machine for each sample. The samples underwent automated processing using the Maxwell DNA extraction (buffy coat) mode. Following the procedure elution tubes were placed in a magnetic elution tube rack. The eluate was then transferred for storage. DNA samples were quantified using spectrophotometry and concentrated where appropriate using a speed vacuum concentrator.

2.4.3 Phenol-chloroform extraction of fixed paraffin embedded tumours

Many clinical specimens are routinely fixed in formalin to preserve tissue architecture. The tissue is then embedded in paraffin wax so sections can be cut using a microtome for subsequent histological analysis. This process leads to irreversible cross-linking of nucleic acids. Cross-linking can inhibit DNA polymerase and PCR reactions. Generally the quality of the DNA degrades with age with increasing DNA fragmentation observed over time (Figure 2.1). Formalin fixed tissues have propensity to develop deamination of cytosine bases to thymine (Hofreiter et al., 2001). When I undertook this DPhil, WGS of FFPE tissue samples with the generation of reliable sequencing data was unavailable.

To overcome this obstacle I worked in collaboration with Illumina to sequence FFPE tumours. I prepared a series of various DNA extractions using Qiagen and phenol-chloroform extractions with different conditions to optimise DNA extractions in preparation for sequencing FFPE tumours. The quality of the extracted DNA was initially assessed primarily through visualisation having run the genomic DNA on a 1% agarose gel. The presence of a distinct high molecular weight band with limited fragmentation characterised a sample of good quality for sequencing (Figure 2.1). Specimens archived for several years were degraded upon visualisation on an

agarose gel irrespective of the method of DNA extraction. Having supplied Illumina with several extractions to construct libraries for sequencing we proceeded to utilise the protocol outlined below for DNA extraction for FFPE tumours. Information regarding quality control checks and processing of sample undertaken by Illumina was not shared with myself

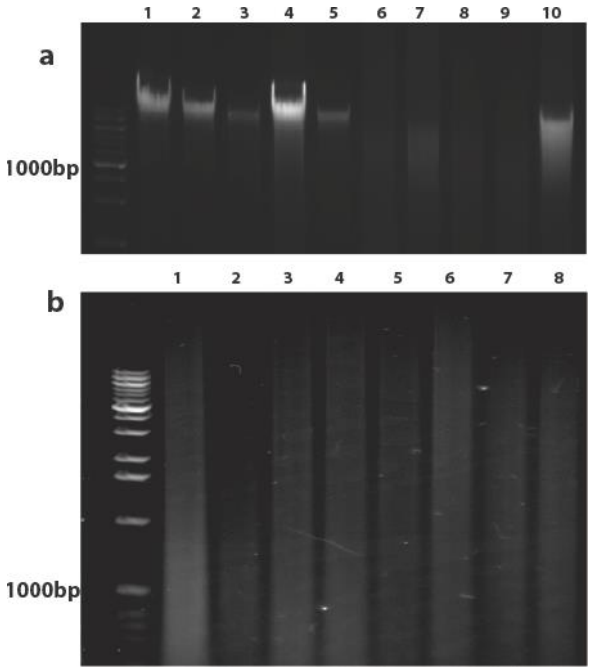


Figure 2.1 I DNA extraction using phenol-chloroform and Qiagen DNA extraction protocols a) Genomic DNA extracted from normal colon tissue which had been fixed and embedded in paraffin immediately after surgery. The genomic DNA was extracted when the paraffin blocks were 4 months old. Lanes 1-5 represents samples extracted using the phenol-chloroform protocol. Lanes 6-10 represent the same samples extracted using the Qiagen extraction protocol. The genomic DNA extracted with the phenol-chloroform extraction protocol forms distinct molecular bands that are preferable for sequencing. b) Genomic DNA extracted from archived colorectal tumour samples where the paraffin embedded blocks are 5 years old. Lanes 1-4 represents tumours that have been extracted using phenol-chloroform extraction, whilst lanes 5-8 represent the same tumours that have been extracted using the Qiagen protocol. DNA degradation is observed with both techniques of DNA extraction.

2.4.4 Phenol-Chloroform extraction protocol

FFPE primary tumours matched to the brain metastases where available, were sectioned at 5µm and 10µm. The 5µm sections were stained with Harris haematoxylin and eosin using standard protocols as outlined in section 2.4.5.

Histological review of the stained slides was used to identify and demarcate the areas of tumour with the assistance of Dr Simon Leedham. This served as a guide to isolate tumour tissue from the 10µm FFPE sections. Paraffin embedded tumour was removed from the slides by scraping with a 21-gauge needle. The retrieved tumour tissue was added to 60µl of PK and 180µl of ATL buffer solution and incubated for 24 hours at 55ºC. Phenol-chloroform DNA extraction was undertaken using the Manual Phase Lock Gel™ kit (5 PRIME). Immediately prior to use, the phase lock gel tubes were centrifuged at 14000 rpm. Lysed samples were loaded into the phase lock tubes with an equal volume of phenol:chloroform:isoamyl alcohol [25:24:1, saturated in 10mM Tris, pH 8.0, 1mM EDTA] (Sigma-Aldrich). The phase lock tubes were mixed by inverting several times and centrifuged at 13000 rpm for 5 minutes. Chloroform (Sigma-Aldrich) of equal volume to the original lysed sample was subsequently added to the top layer in the phase lock tube. The sample was then centrifuged at 13000 rpm for 5 minutes. The aqueous layer from the tube was retrieved and transferred to fresh eppendorf tubes and the volume of this for each sample was quantified. This volume is herein referred to as V. The following solutions were then added to the eppendorf tubes: 1µl of glycogen, 5M NaCl at 1/20 of the volume of V and 100% ethanol twice the volume of V. The samples were then vortexed and incubated overnight at -20ºC. Following incubation samples were centrifuged at 13000 rpm for 30 minutes. The pellet that formed was washed with 500µl of 70% ethanol, centrifuged at 13000 rpm for 5 minutes and air-dried at room temperature. Following this the pellet was resuspended in 100µl of nuclease free water. Extracted DNA was then assessed for quality and quantified.

2.4.5 Harris haematoxylin and eosin staining

Sections of paraffin embedded tissue were cut at 5µm using a microtome and dried at 37ºC for 3 hours. Slides were immersed in xylene for 5 minutes and rehydrated in a series of graded alcohols (100%, 90%, 70%) for 2 minutes sequentially and then

placed in water for 2 minutes. The slides were then immersed in Harris haematoxylin for 30 seconds, washed with water, and briefly immersed in acid alcohol. The slides were then placed in water for a further 5 minutes and immersed in eosin for 3 minutes followed by a further immersion in water. The slides were then dehydrated using a series of graded alcohols (70%, 90%, 100%) 1 minute sequentially. Following this the samples were immersed in xylene for 1 minute and mounted with a cover slip using DPX mounting media (Sigma-Aldrich). Regions of tumour were identified at microscopy and demarcated (Figure 2.2).

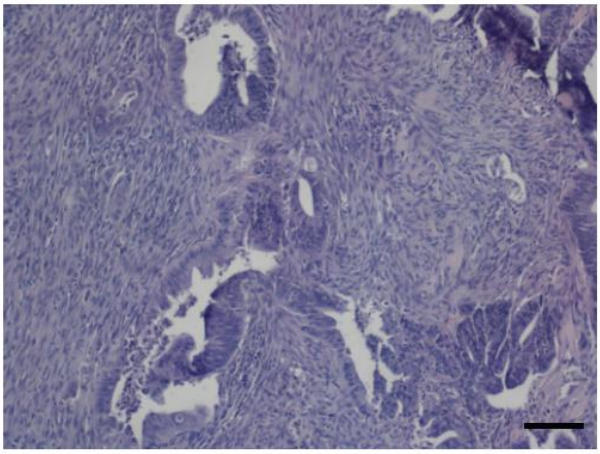


Figure 2.2 | Haematoxylin and eosin staining of sections of primary colorectal cancer. Regions containing tumour were demarcated by staining to delineate tissue for DNA extraction. Scale bar: 100μm.

2.5 Removal of RNA from DNA extractions

Ribonuclease RNase 1 degrades single-stranded nucleic acid molecules into individual nucleoside 3' monophosphates by cleaving each phosphodiesterase bond. The enzyme is inactivated by heating for 20 minutes at 70^{o}C with 5mM dithiothreitol (DTT). DNA extractions were treated with RNase as RNA can form a template for DNA ligase in library preparations resulting in sequencing artefacts

RNA was removed from DNA extractions by using RNase 1 (Epicentre). RNase 1 was diluted 1:10 with enzyme dilution buffer, with 20µl then added to each DNA preparation. The mixture was then incubated for 30 minutes for 37^{o}C to degrade contaminating RNA. Following this DTT was added to achieve a final concentration of 5-10mM. The RNase 1 enzyme was then inactivated by incubating the mixture for 20 minutes at 70^{o}C.

2.6 Assessment of DNA quality and quantification

2.6.1 Agarose gel electrophoresis

The quality of the extracted DNA was assessed using agarose gel electrophoresis, which separates DNA molecules by their size. The degree of fragmentation from the extracted DNA can be evaluated. DNA molecules can be visualised using ethidium bromide dye, which intercalates with DNA and fluoresces in UV light (UV absorbance 285 nm wavelength, emission at 605 nm).

Gel electrophoresis was undertaken with a 1% agarose gel. 20g of agarose powder (Invitrogen) was dissolved in 1X TBE buffer and heated. The agarose solution was then cooled to 60^{o}C and ethidium bromide dye was added at a final concentration of 0.25µg/ml. The agarose solution was poured into a gel casting tray with well combs and allowed to set at room temperature. Once set, the well combs were removed and the agarose gel was placed in DC power pack immersed with 1 X TBE buffer. 2µl of genomic DNA was mixed with 1µl of 5 x Orange Loading Dye (Trevigen) and loaded into the wells. In addition, 3µl of 1 KB Plus DNA Ladder (Invitrogen) was added to the first well to serve as a guide to estimate the size of the DNA molecules being analysed. Electrophoresis was undertaken for 30 minutes at 100V. DNA was visualised using UV light using the UV box transilluminator (Bio-Rad). Images were taken with Image Lab^{TM} image acquisition analysis software (Bio-Rad).

2.6.2 Spectrophotometry

Spectrophotometry is a method to quantify DNA by measuring the amount of DNA UV radiation absorbed (Barbas et al., 2007). The NanoDrop-100 spectrophotometer (Thermo Scientific) was utilised to evaluate DNA quantity and purity at 260nm and 280nm UV wavelengths. The optical density (OD) is a logarithmic ratio of the radiation falling on a material, to the radiation that is transmitted through the material. An OD of 1 at 260nm (OD_{260}) corresponds to 50µg/ml of double stranded DNA. Spectrophotometry quantification can be inaccurate in the presence of contaminating nucleic acids. The purity of the DNA can be assessed by calculating the ratio of OD_{260}/OD_{280}. A ratio of ~1.8 corresponds to a pure sample of DNA, whereas pure RNA sample has a OD_{260}/OD_{280} of ~2.

2.6.3 Quantification of DNA using PicoGreen assay

The PicoGreen assay is a sensitive assay for quantifying double-stranded DNA (dsDNA) in solution using a fluorescent nucleic acid stain. The technique utilises a bisbenzimide dye that is sensitive to double stranded DNA. The assay is more sensitive than spectrophotometry detecting concentrations as low as 25pg/mL. The assay can discriminate between DNA and RNA as bisbenzimide dyes preferentially binds to dsDNA.

A working solution was generated by diluting picogreen reagent, Quant-iT™ Picogreen® (Invitrogen) in a 1:200 dilution of TE buffer (10mM Tris-HCL, 1mM EDTA, pH 7.5). 2 µg/mL of dsDNA standard was diluted in TE buffer to generate solutions of varying concentrations of DNA which would produce an eight point standard curve (final concentrations- blank, 0.025ng/µl, 0.05ng/µl, 0.1ng/µl, 0.3ng/µl, 0.5ng/µl 0.7 ng/µl, 1ng/µl). Experimental DNA was diluted 1:100 in TE buffer solution. The working solution of picogreen reagent was added to the diluted experimental DNA and standard DNA solutions, centrifuged and incubated at room temperature for 5

minutes. The mixture was shielded from light during this time to prevent photodegradation. After incubation the sample fluorescence was measured using a fluorescence microplate reader (fluorescein wavelengths - excitation 485nm, emission 535nm). The fluorescence value of the reagent blank was subtracted from each sample and a standard curve was generated of fluorescence versus DNA concentration. Using the standard curve the concentration of the experimental DNA solutions were determined.

2.6.4 Quantification of DNA using Qubit assay

The Qubit assay is a highly sensitive assay that can quantify DNA, RNA or protein. A fluorescent dye emits a signal when bound to the target molecule (dsDNA in this case) which is detected by a bench top fluorometer.

The Qubit assay was undertaken using the Qubit® dsDNA HS Assay kit (Life Technologies). A working Qubit™ solution was generated by diluting Qubit™ dsDNA HS reagent in a 1:200 dilution in Qubit™ dsDNA HS buffer. To calibrate the Qubit 2.0 fluorometer 2 DNA standards were utilised. 190µl of Qubit™. working solution was loaded into 2 assay tubes. 10µl of the standard DNA solutions was added to the assay tubes. The final volumes in assay tubes equated to 200µl. In addition, experimental DNA of 1µl each was added to 190 µl of Qubit™. working solutions. All mixtures were vortexed, centrifuged briefly and incubated at room temperature for 2 minutes. The Qubit 2.0 fluorometer was calibrated using the standard DNA solutions and upon completion of calibration experimental DNA concentrations were quantified.

2.7 Technical validation of identified somatic variants

NGS using Illumina sequencing is an established reliable method of sequencing germline and somatic variants. Despite this, it is remains common practice at present to validate the findings from NGS using alternative sequencing methods. Technical validation of somatic variants was undertaken using the Ion Torrent Personal Genome Machine (PGM) [Life Technologies]. This platform utilises electrochemical detection technology using semiconductors to infer the DNA sequence by monitoring the release of hydrogen ions during sequencing in DNA synthesis (Rothberg et al., 2011). In this project technical validation was undertaken using a targeted approach which involved screening for mutations in genes which are present on a custom high priority cancer gene panel (refer to appendix A for list of genes). Sequencing library preparation and data analysis was conducted by myself. The steps involved in the construction of the library are outlined below using the HID-Ion Ampliseq™ Library Preparation kit (Figure 2.3).

2.7.1 Amplification of genomic targets of DNA by PCR reaction

Input DNA was quantified and an aliquot of each sample was taken and diluted to achieve concentrations of 5ng/µl. A PCR reaction was then executed using the following components to achieve a final total volume of 10 µl per reaction: 5X ion Ampliseq HIFI™ master mix (Life Technologies), 2X ion Ampliseq primer pool (Life technologies), and genomic DNA (5ng/µl).

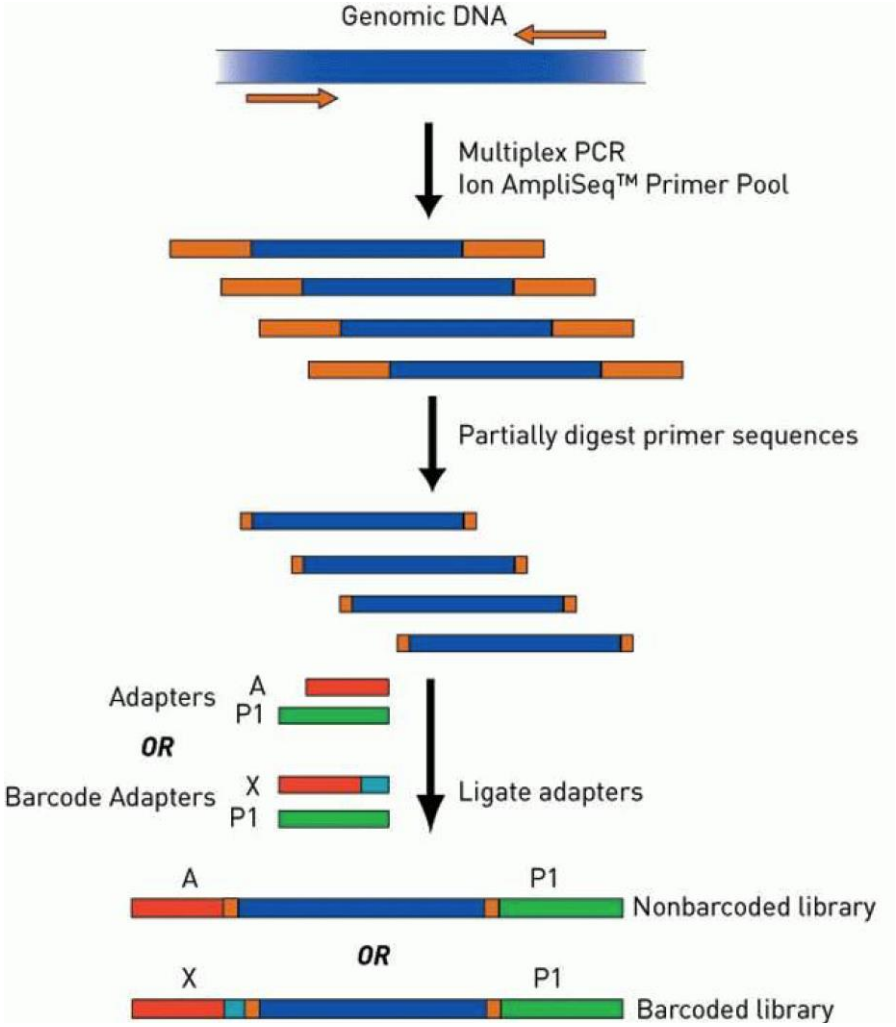


Figure 2.3 I Workflow summarising steps in library preparation for Ion Torrent. Image derived from Life Technologies.

Each experimental sample had 2 reactions created per sample (2x 10 µl per sample of DNA). The thermo cycler profile is outlined below. After the PCR the 2 pools of samples were combined to give a total volume of 20µl.

Thermocycler profile

1. Activation of the enzyme- 99°C for 2 minutes.
2. Denaturing the DNA- 99°C for 15 seconds.
3. Annealing and extension of DNA- 21 cycles of 99°C for 15 seconds/60°C for 4 minutes.
4. Hold the reaction at 10 °C.

2.7.2 Partial digestion of primer sequences

2µl of FuPa reagent was added to each sample. The samples were mixed by vortexing and incubated in the thermocycler according to the following schedule - 50ºC at 10min, 55ºC at 10 min, 60ºC at 20min then hold 10ºC.

2.7.3 Ligation of barcode adaptors

4µl of switch solution, 2µl of diluted barcode adaptor mix, 2µl of DNA ligase was added to each well (containing digested amplicon 22µl) to give a total volume of 30 µl. The samples were then vortexed and spun down briefly and incubated in the thermocycler. The following program was run: 22ºC at 30min, 72ºC at 10 min and then held at 10ºC.

2.7.4 Purification of unamplified library

Agencourt AMPure® XP reagent was brought to room temperature and vortexed. 45µl of the AMPure® XP reagent was added to each library, mixed by vortexing and incubated at room temperature for 5 minutes. The plate was then placed in a magnetic rack (DyanMag™) and incubated till the solution became clear and a pellet formed adjacent to each individual magnet. The supernatant was discarded without disrupting the pellet. Each pellet was then washed by adding 150µl of 70% ethanol and the plate was moved along the rack so that the pellet formed on the opposite side of each well. The supernatant was discarded and the washing process was repeated again. For the second wash residual ethanol was removed and the pellets were left to air-dry for 5 minutes at room temperature.

2.7.5 Amplification of library

Following air-drying of the pellet, the plate was then removed from the magnet and 50µl of Platinum PCR supermix high fidelity mix (Invitrogen) and 2µl of library amplification mix were added to each pellet. The pellet was then resuspended in the mixture by pipetting. Following this the plate was placed back on the magnet for 2

minutes after which 50µl of supernatant was transferred to a new PCR plate. The samples were then incubated in the thermocycler using the following schedule (98°C 2min, then 5 cycles 98°C at 15sec and 60°C for 1min, and then hold at 10°C.

2.7.6 Purification of amplification clean up

Following the PCR reaction 70µl of AMPure® XP reagent was added to the PCR product. The mixture was vortexed and incubated for 5 minutes at room temperature. The plate was then placed in the magnetic rack for 5 minutes at which point the solution became clear, after which the supernatant was discarded. The remaining pellet was then washed twice with 150µl of 70% ethanol. The supernatant was then discarded leaving pellets to air-dry for 5 minutes. The plate was then removed from the magnet and the pellet was dispersed in TE solution, after which the plate was replaced in the magnetic rack and on this occasion the supernatant was retrieved.

2.7.7 Bioanalyzer

Prior to sequencing the quality of the library generated has to be determined. Sequencing libraries were analysed using a bioanalyzer to ensure that DNA molecules of the appropriate size were selected, with the aim to size select DNA molecules in the range of 100-300 base pairs. The Agilent High Sensitivity DNA Kit (Agilent Technologies) was used alongside the Agilent 2100 bioanalyzer to assess the quality of the DNA library. In brief, the chip priming station was adjusted appropriately and the gel-dye mix and 2 DNA ladders were loaded on to the high sensitivity DNA chip in appropriate wells in accordance with the manufacturer's protocol. The chip was then vortexed for 1 minute at 2400 rpm in an IKA vortex mixer (Agilent Technologies). Following this the chip was then loaded into the Agilent 2100 bioanalyzer. The data was processed using the 2100 Expert Software (Agilent) (Figure 2.4).

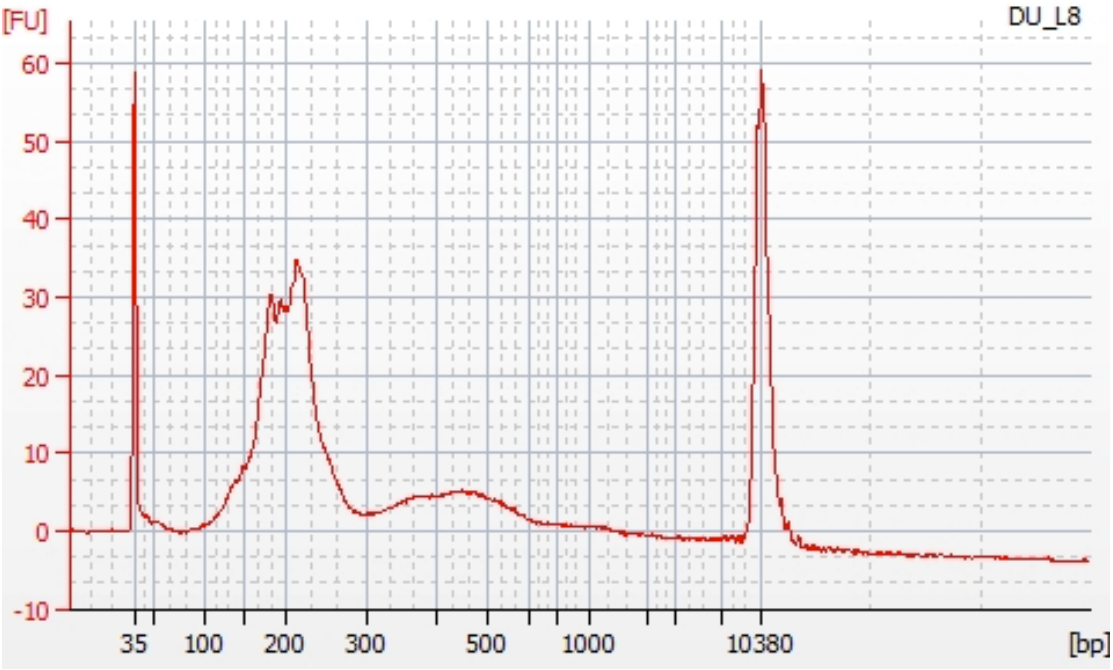


Figure 2.4 I Electropherogram demonstrating quality of DNA library. Following size selection DNA fragments in the range of 100-300 base pairs were selected. The trace also depicts DNA ladders visible at 35 and 10380 base pairs.

2.7.7 Sequencing of libraries

Following library construction sequencing of the prepared library was undertaken in the Wellcome Trust Centre for human Genetics (Oxford). Subsequent data analysis was undertaken by myself.

2.8 Single nucleotide polymorphism arrays

2.8.1 Summary of tumours processed using SNP arrays

DNA obtained from fresh frozen and FFPE tumours were genotyped using SNP arrays on different platforms as summarised in Table 2.1. Fresh frozen samples analysed using SNP arrays include 5 colorectal, 7 breast, 7 lung adenocarcinoma brain metastases and a cohort of 25 liver metastases (of which 8 were paired).

Platform	Sample
Human Omn1 Quad v1	5 Colorectal brain metastases
Human CytoSNP 12 v2.1	7 Lung, and 7 breast brain metastases
Human Omni2.5-4 v1	22 liver metastases
Human 610 Quad v1	4 liver metastases
Human OmniExpress-24 v1	22 FFPE colorectal, breast and lung brain metastases

Table 2.1 I Summary of platforms used for SNP arrays

The liver metastases were processed by a colleague Dr Rajarajan Ramamoorthy. Colorectal, breast and lung FFPE blocks from the original brain metastases where available and a larger cohort of FFPE brain metastases were also processed using the Illumina FFPE SNP array manufacturer's protocol. In total 22 FFPE tumours passed quality control measures. Data analysis was undertaken by myself. The arrays were undertaken in accordance with the Illumina Infinium assays manual protocol. An outline of the process is provided below.

2.8.2 Amplification of DNA

An MSA3 plate was used for amplifying DNA. 20µl of MA1 solution was loaded in to the MSA3 plate alongside 4µl of DNA sample (50 ng/µl). The location of DNA samples in each well was recorded. 4µl of 0.1NaOH was added to each well to denature DNA. The plate was vortexed and centrifuged and incubated at room temperature for 10 minutes. Following this 34µl of MA2 and 38µl of MSM solutions were dispensed into each well containing samples. The plate was again vortexed and centrifuged, then incubated for 24 hours in a hybridization oven (Illumina) at 37^{O}C.

2.8.3 Fragmentation and precipitation of DNA

Amplified DNA generated in section 2.8.1 was fragmented in an enzymatic process. Once removed from the oven 25µl of FMS solution was added to each sample. The plate was then vortexed, centrifuged to 50 xg for 1 minute at 22^{O}C and then placed on a heat block for 1 hour at 37 ^{O}C. 50µl of PM1 was added to each sample. The

plate was then vortexed, incubated for 5 minutes at 37^{O}C, followed by centrifugation at 50 xg for 1 minute at 22^{O}C. After this 155µl of 100% 2-propanolol was added to each well upon which the plate was then inverted ten times to thoroughly mix the contents and left to incubate at 4^{O}C for 30 minutes. Following incubation the plate was centrifuged at 3000 xg for 20 minutes at 4^{O}C. Immediately after centrifugation supernatant from the wells was decanted and then air-dried at room temperature for 1 hour leaving blue pellets at the bottom of the wells.

2.8.4 Resuspension of precipitated DNA

23µl of RA1 was added to each well of the MSA3 plate containing the DNA pellet. The plate was sealed with foil and incubated in the hybridization oven for 1 hour at 48^{O}C. The plate was then vortexed at 1800rpm and pulse centrifuged.

2.8.5 Hybridization on to multi-use beadchip

The hyb chamber (Illumina) was assembled in accordance with the manufacturer's protocol. 400µl of PB2 solution was placed into the reservoirs and the hyb chambers were then closed shut until used. The MS3 plate was the heated for 20 minutes at 95^{O}C to denature DNA and then allowed to rest at room temperature for 30 minutes. Beadchips were then placed into the hyb chambers in appropriate orientation. 12µl of each sample was subsequently loaded into appropriate inlets on the beadchip and locations were recorded along with the bar codes for each group of samples. The beadchips were then loaded into the hyb chambers, which were then sealed and incubated in the hybridization oven for 20 hours. After incubation and cooling at room temperature for 25 minutes, the intelli hyb seals were removed from the beadchips and the beadchips were then placed into the wash rack and immersed in PB1 solution for 1 minute.

2.8.6 Single base extension and staining of beadchips

The chamber racks (Illumina), were heated to 44^{O}C, upon which flow through chambers were inserted in to the rack. The reagents were sequentially placed in to the flow through chamber as outlined in Table 2.2.

Processing step	Reagent	Incubation time
1	150µl RAI	30 seconds then repeat step 1 for a total of 5 times
2	450µl XC1	10 minutes
3	450µl XC2	10 minutes
4	200 µl TEM	15 minutes
5	450µl 95% formamide/1mM EDTA	1 minute then repeat step 5

Table 2.2 l Incubation steps in single base extension

Following this the samples were then incubated for 5 minutes. For staining a further series of solutions were loaded for this process (Table 2.3).

Processing step	Reagent	Incubation time
1	250µl STM	10 minutes
2	450µl XC3	1 minute, repeat once the wait 5 minutes
After completing step 2, step 1 and step 2 were repeated for a total of 5 times sequentially. The beadchips were then removed from the flow through chambers.		
3	PB1	Move beadchip up and down in solution 10 times
4	XC4	Move beadchip up and down in solution 10 times
Following step 4 the beadchips were placed in a tube rack and placed in a vacuum desiccator (Sigma-Aldrich) to dry in a vacuum for 50 minutes.		

Table 2.3 l Sequential steps in the staining process

2.8.7 Imaging beadchips

The iScan reader system (Illumina) was used to image the beadchip. The reader excites the fluor of the single base extension product on the beads on the beadchip which is then recorded. Following calibration of the reader and imaging of the

beadchip, intensity data (idat) files were generated. Genome Studio (Illumina) was utilised for further analysis of idat files.

2.9 Analysis of FFPE brain metastases using SNP arrays

FFPE SNP arrays were run using the Illumina Human Omniexpress 24 v 1.0 platform. Prior to running the array the extracted DNA samples had to be assessed for quality. Samples that passed this evaluation were restored using the Illumina FFPE restoration protocol before being run on an array. The arrays were subsequently processed similarly to the description given in section 2.8.

2.9.1 Illumina FFPE qc protocol

Extracted FFPE DNA was assessed for quality using the Illumina FFPE QC kit (Illumina) (Figure 2.5). An aliquot of DNA quantified by PicoGreen assay (section 2.6.3) was diluted to 1ng/ µl. The QCT (standard sample) was diluted to form a 1:100 dilution. The following reactions were then created using a final volume of 20µl using either genomic DNA, the standard or control alongside SYBR Green PCR master mix (Applied Biosystems). The qPCR was run with the following thermal profile (50^{o}C 2 minutes, 95^{o}C 10 minutes, then 40 cycles of 95^{o}C 30 seconds, 57^{o}C 30seconds, 72^{o}C 30seconds). Each genomic DNA sample was processed thrice. The Cq values from the FFPE genomic samples were averaged. Samples were determined to be of sufficient quality for restoration based on a comparison of the average Cq value of the DNA samples run in triplets and Cq value of the standard as defined in the manufacturer's protocol.

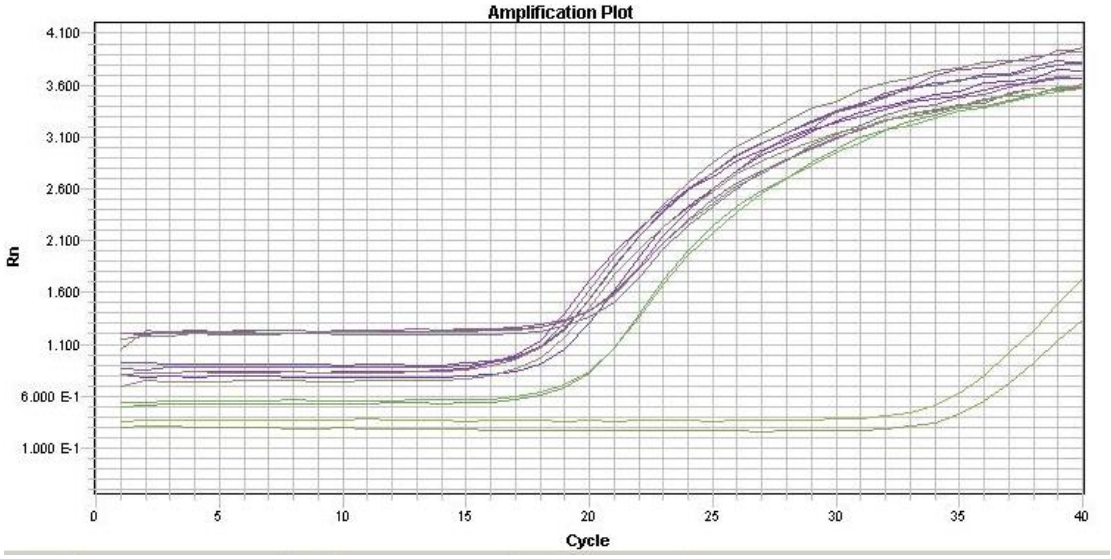


Figure 2.5 l qPCR trace using Illumina FFPE qc kit. The graph represents the amplification efficiency of the multiple DNA samples.

2.9.2 Illumina restoration protocol for FFPE arrays

The Illumina restoration kit restores degraded DNA to a state that is compatible for amplification for FFPE SNP array genotyping. DNA was purified using the Zymo purification kit (Zymo Research). 8µl (100ng) of genomic DNA from each sample was added to each well in the PRAS plate. 4µl of 0.1N NaOH was dispensed to each well, followed by incubation at room temperature for 10 minutes. Following this, 34µl of PPR reagent and 38µl of AMR reagent was added to each sample. The plate was sealed, inverted ten times, centrifuged at 280 xg for 1 minute and then incubated at 37°C for 1 hour. Following this, 560 µl of Zymo DNA binding buffer was added to each well and mixed by pipetting. Samples were transferred to the Zymo-Spin Plate and centrifuged at 2250 xg for 2 minutes. The flow through from the collection plate was discarded. 600µl of Zymo wash buffer was added to each well and the plate was centrifuged again at 2250 xg for 2 minutes. The Zymo-spin plate was placed over an RST plate and 13µl of ERB solution was added to each sample followed by incubation at room temperature for 5 minutes. The plates were centrifuged again at 2250 xg for 1 minute. The eluant was collected and incubated for 2 minutes at 95°C

in the heat block. Immediately after incubation, the plate was cooled in ice for 5 minutes. 10 µl of CMM solution was added to each well. The plate was sealed, vortexed, centrifuged for 1 minute at 280 xg and incubated at 37°C for 1 hour. After incubation 140 µl of Zymo DNA binding buffer was added to each well and the plate was centrifuged at 2250 xg for 2 minutes. The plate was then washed with Zymo wash buffer administered to each well. Finally 10 µl of nuclease free water was applied to each column and after incubation for 5 minutes at room temperature DNA was eluted after centrifugation at 2250 xg for 1 minute.

Analysis of sequencing data

2.10 Next generation sequencing

Advances in NGS in recent years have been rapid leading to the development of different generations of sequencing platforms utilising different sequencing technologies (Metzker, 2010, Liu et al., 2012a). This progress has been coupled with advances in bioinformatic analysis. The analysis of cancer genomes has proven to be technically challenging (Ulahannan et al., 2013). Specific advances in this field include: the development of cancer specific variant callers, identification of copy number and LOH directly from sequencing data, evaluation of the clonal architecture of tumours and sequencing FPPE tumours.

Three different approaches to sequencing the genome have typically been adopted. Firstly one may analyse the entire genome interrogating coding and non-coding regions. Alternatively, whole exome sequencing (WES) may be undertaken which relies on a capture based method. A third approach involves targeting specific regions, or genes. This approach may be useful when screening for known variants in the regions of interest. The work presented in this DPhil involves WGS of a cohort of brain metastases originating from different types of primary tumours. Additionally, sequencing data from paired colorectal liver metastases and multi-region analysis of

primary colorectal tumours were used for comparative analysis. The brain metastases and multi-region samples derived from primary colorectal tumours were sequenced using Illumina HiSeq 2000 technology (Illumina San Diego, CA), whereas the liver metastases were sequenced using the Complete Genomics platform (Mountain View, Ca). In addition, technical validation of the brain metastases variants identified by Illumina sequencing was undertaken. Somatic variants identified were cross-validated using the Ion Torrent PGM (Life Technologies), using a high priority cancer gene panel. Bioinformatic analysis in this project was undertaken by myself.

2.11 Analysis of Illumina whole genome sequencing data

2.11.1 Summary of sequencing characteristics

Extracted tumour and paired constitutional DNA from patients with brain metastases were sequenced using the Illumina HiSeq 2000 sequencing platform. Paired end reads were utilised (100 base pairs) with the aim to achieve a median depth of coverage of 30X for constitutional DNA and 60X for tumour samples. The tumours were sequenced in a collaborative effort as part of the Oxford-Illumina WGS500 initiative. Library preparation and sequencing was undertaken by both Illumina and the Core Genomics department at the Wellcome Trust Centre for Human Genetics (Oxford). The NEBNext DNA Sample Prep Master Mix Set1 kit (New England Biosciences) was utilised to construct libraries for Illumina sequencing. A total of 1µg of DNA was required for sequencing. An overview of the multiple steps involved in analysing sequencing data is represented in Figure 2.6.

2.11.2 Mapping of sequencing data

Information generated from sequencing platforms is produced in FASTQ files. These files contain information regarding the sequence of each read and quality scores pertaining to each base. BAM files are generated in the process of mapping where

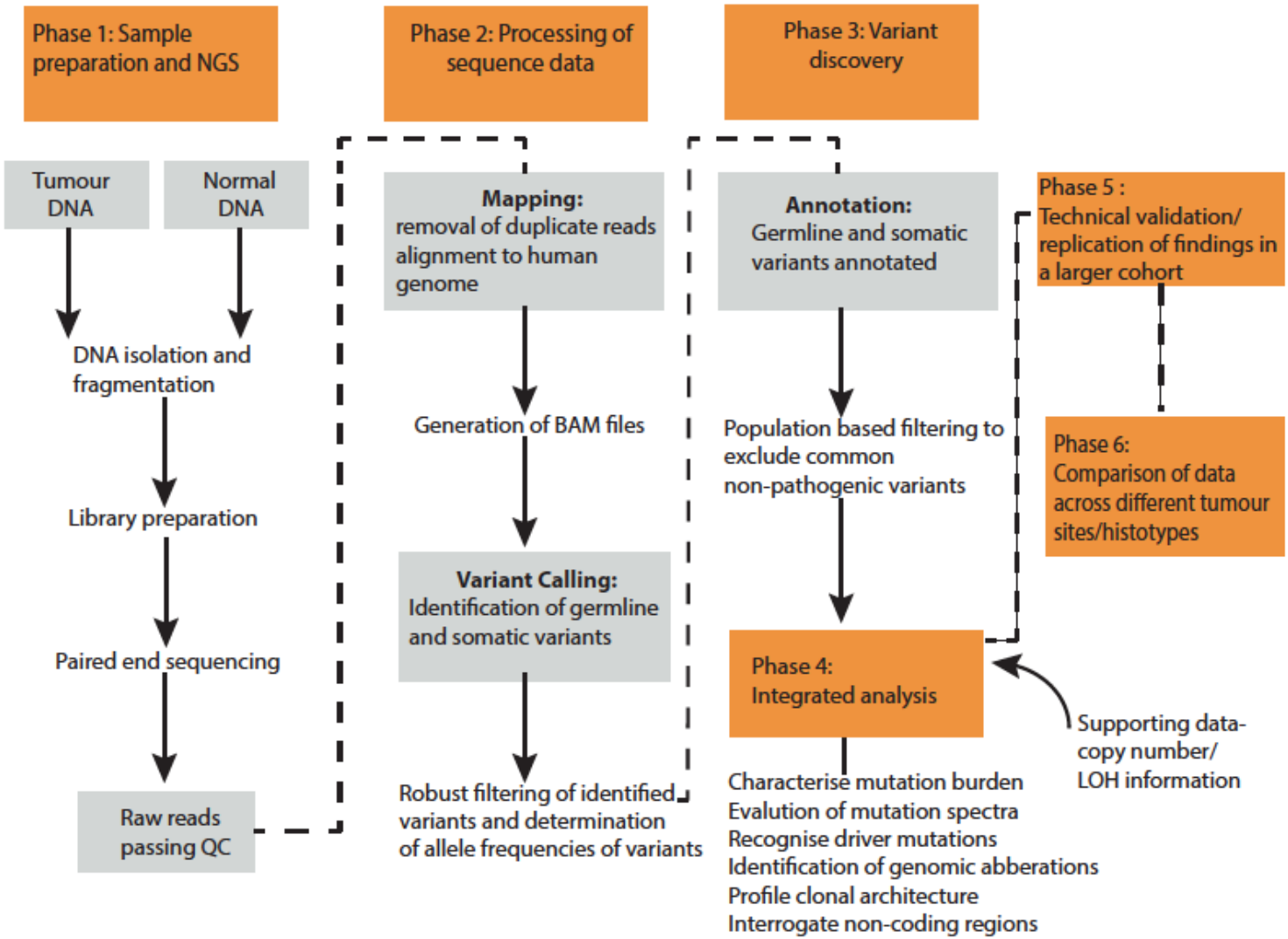


Figure 2.6 I Outline of processing steps involved in cancer genome analysis.

Image adapted from Ulahannan et al 2013

reads are aligned to a reference genome through comparative assembly. Caveats in this process involve mapping complex regions including regions containing repetitive sequences. Dedicated mapping programs are used to generate BAM files (Horner et al., 2010). The selection of the software utilised is dependent on the accuracy, sensitivity, computing capacity and time efficiency (Bao et al., 2011). The Illumina whole genome data sets were mapped by Dr John Broxholme in the Oxford-Illumina WGS 500 initiative. PCR duplicates were removed from BAM files and reads were subsequently mapped to the Human Reference Genome (GRCh37d5) using stampy v1.013 and BWA v0.5.6 (Li and Durbin, 2009, Lunter and Goodson, 2011). All subsequent downstream analysis of sequencing data was undertaken by myself.

2.11.3 Variant calling and filtering of somatic mutations

Having mapped reads to a reference genome variant calling programs are utilised to identify putative mutations which are recorded in variant call format (VCF) files in a standardised data format. The VCF file format contains information regarding the location of the putative mutation, the quality of the call and characteristics pertaining to the variant. In this project several variant calling programs were utilised to analyse the data in the initial stages of analysis including Platypus, Mutect and Somatic Sniper (Rimmer et al., 2014, Cibulskis et al., 2013, Larson et al., 2012). All known established single nucleotide variant (SNV) driver mutations were identified by all three variant specific callers. However, analysis subsequently focused using the variant caller Platypus for several reasons. Firstly, as Platypus was developed within the department at the Wellcome Trust Centre for Human Genetics by Dr Andrew Rimmer thresholds for variant calling could be adapted more readily if required. Unlike the other two variant callers Platypus had the ability to identify indels. Moreover, Platypus has the ability to simultaneously undertake variant calling on several BAM files which was useful when I analysed a trio data set involving BAM files from a control, primary FFPE tumour and colorectal brain metastasis (Chapter 4). The processing steps outlined below represent the most robust framework

achieved for data analysis and pertain to variant calls generated by Platypus. Measures for assessing quality control in this process involved confirming the presence of identified somatic variants in the VCF files using IGV (Thorvaldsdottir et al., 2013). Somatic variants were subsequently technically validated using the Ion Torrent PGM

Variant calling was undertaken by Platypus v0.4 (Rimmer et al., 2014). Tumour and normal BAM files were analysed together to identify germline and somatic mutations. VCF files were filtered to retain high quality calls for somatic mutations using a series of steps.

1) Variant calls with allele bias and pass calls were retained.
2) Calls with posterior probability less 20 than were excluded.
3) Regions of segmental duplications were removed. Segmental duplication coordinates were obtained from the University of California Santa Cruz (UCSC).
4) A minimum depth of coverage of 10X was required in both the tumour and normal at a given loci.
5) Mutations with variants in the normal VCF file at those loci were excluded.
6) Somatic mutations with a frequency less than 3% were excluded.

2.11.4 Annotation and further filtering of somatic variants

The VCF file contains information regarding the position of the mutation. However from this file one can not directly determine whether the mutation is exonic and the functional significance of the mutation. The software engine ANNOVAR was used to annotate the filtered variant calls by assigning mutations to a gene where applicable (Wang et al., 2010). Each mutation was annotated using SIFT and PolyPhen2 scores to predict their functional significance (Ng and Henikoff, 2003, Adzhubei et al., 2010). Moreover, mutations were annotated with population databases using ANNOVAR to

identify the frequency of the mutations in these databases. Annotations included the single nucleotide polymorphism database (dbSNP v135) http://www.ncbi.nlm.nih.gov/SNP/, 1000 genomes project allele frequencies http://www.1000genomes.org and the 6000 National Heart Lung and Blood Institute (NHLBI) exome project. Variants in these databases were excluded if they were present at >10% frequency.

2.11.5 Analysis of trio dataset

WGS of a single colorectal FFPE primary tumour was successfully processed in collaboration with Illumina. When I commenced my DPhil, WGS of FFPE samples with reliable data quality was not an available technique. This FFPE tumour was sequenced in a pilot study achieving a median depth of coverage of 30X. Platypus v0.4 (Rimmer et al., 2014) was utilised to simultaneously analyse BAM files from the matched control, FFPE primary tumour (P1) and fresh frozen brain metastasis (C1MET). The VCF files were then processed similarly to the process outlined in sections 2.11.3-2.11.4. A minimum depth of coverage of 10X was required in the control, FFPE primary tumour and fresh frozen brain metastases at a given loci. To reaffirm the sequencing results generated from this analysis and to ensure a fair comparison, the metastasis BAM file was down-sampled at random using Picard http://sourceforge.net/projects/picard/ to achieve a comparable median depth of coverage of 30X. Variant calling was repeated on the BAM files of the control, FFPE primary tumour and the down sampled C1MET brain metastasis. Subsequent filtering was undertaken as described above. I have presented results from the original analysis and subsequent analysis following down sampling in Chapter 4.

2.12 Analysis of Complete Genomics data

Complete Genomics sequencing amplifies fragmented DNA using a nanoball amplification method. The nanoballs are then packed onto a silicon chip. A combinatorial probe ligation system is then used to infer the sequence of DNA by

detecting the emission from a fluorescent probe. Complete Genomics provide a service to sequence DNA samples and produce partially processed bioinformatic data using the CGAtools package (Complete Genomics). Analysis of the data downstream of the VCF files was undertaken by myself.

3 pairs of matched liver metastases obtained from the Royal Free Hospital underwent WGS by Complete Genomics with the aim of achieving a median depth of coverage of 60X in the tumour and 30X in the control. Data obtained from sequencing was processed and analysed by myself to be used as a comparison to the colorectal cancer brain metastases. Gapped reads of 35 base pairs in length, were mapped to the GRCh37d5 Human Reference Genome using the CGAtools package.

VCF files were processed by myself using the programs SnpEff and SnpSift to generate a tab-delimited data frame (Cingolani et al., 2012). The allelic proportions for each variant were calculated. This data frame was subsequently filtered to extract calls called with confidence and annotated using ANNOVAR. The filtering processes are outlined below.

1) Regions of segmental duplications were removed. Segmental duplication coordinates were obtained from UCSC.
2) A minimum depth of coverage of 10X was required in both the tumour and normal at a given loci.
3) Mutations with variants in the normal VCF file at those loci were excluded.
4) Variant calls were filtered to exclude variants: present in DBSP v135, 1000 genome project and 6000 exome project >10%.

2.13 Multi-region analysis of primary colorectal tumours

Two primary colorectal tumours derived form chemotherapy naive patients were sampled from spatially separated sites and underwent WGS using the Illumina HiSeq 2000 platform. In total 6 and 7 regions were analysed from the tumours MR1 and MR2 respectively. The samples were part of an independent project conducted by Dr Simon Leedham, though served as comparison data set in my study. Data analysis was undertaken by myself. The VCF files were then processed similarly to the process outlined in sections 2.11.3-2.11.4.

2.14 Downstream Analysis of sequencing data

2.14.1 Analysis of mutation spectra

Analysing the mutation spectra of cancer genomes can be informative as it may reflect biological processes such errors in DNA repair pathways, or exposure to carcinogenic agents (Alexandrov et al., 2013, Lee et al., 2010).

There are 6 classes of base substitutions (C>A/G>T, C>G/G>C, C>T/G>A, T>A/A>T, T>C/A>G, T>G/A>C). The proportion of mutations for each base substitution was calculated using a bespoke computer script. Base substitutions across the genome and within the exome are reported where applicable.

2.14.2 Analysis of mutation signatures

Tumours may be exposed to numerous biological processes which may result in characteristic base substitutions. Somatic signature (http://bioconductor.org) was utilised to extract somatic motifs from multiple samples to generate mutation signatures. Base substitutions were analysed in the sequence context by identifying the 5' and 3' base flanking each mutation. In total 96 possible trinucleotide sequence motifs exist using this method. Data were normalised to account for there natural frequency of motifs in the genome.

2.14.3 OncoSNP-SEQ

OncoSNP-SEQ is a bioinformatic tool to directly obtain information regarding copy number and LOH from WGS data by using allele specific read counts (Yau, 2013). The tool provides an estimate of the haploid read depth from sequencing data taking into account the possible ploidy and degree of normal cell contamination producing models which best fit the sequencing data. OncoSNP-SEQ was utilised on both Illumina and Complete Genomics sequencing data sets. The output qc files generate information for different models for each tumour with regards to mean copy number, haploid read depth, fraction of estimated normal cell contamination. Each model is assigned a log-likelihood ratio; a measure of how best each model fits the sequencing data. Output files were systematically analysed by manual visual inspection to select the most appropriate model to obtain information regarding copy number and LOH. Data from OncoSNP-SEQ was used to verify the findings obtained from SNP arrays.

2.14.4 Analysis of clonal architecture

PyClone is a software program that can infer the structure of clonal populations in cancers using a Bayesian clustering method to group sequenced mutations into candidate clonal groups (Roth et al., 2014). The program estimates cellular prevalence of the mutations accounting for allelic imbalances resulting from copy number changes. The clonal architecture of the tumour samples were analysed using PyClone v0.12.3.1. Allelic frequencies for each mutation at a given loci were determined by the total read depth and total number of reads supporting a variant at a given loci. Information regarding copy number for each locus was obtained from OncoSNP after processing genotyped SNP array data. The clonal architecture of each tumour was modelled using PyClone to ascertain the complexity of the clonal structure and to determine the position of key driver mutations within this framework.

2.14.5 IntOGen analysis of somatic mutations

The IntOGen software engine was used to identify driver mutations in a cohort of colorectal, breast and lung primary tumours from TCGA (Gonzalez-Perez et al., 2013). The Oncodrive FM component identifies genes that are biased towards the accumulation of mutations with significant functional impact. It detects genes that are likely to be positively selected in tumour development and are more likely to be drivers. Only genes with a false discovery rate (FDR) < 0.05 were deemed significant. The candidate driver genes from the TCGA analysis were used to screen the brain metastases to determine if drivers of primary tumours were retained in the brain metastases. IntOGen was also run using the sequenced brain metastases independently.

2.14.6 Mutsig analysis of somatic mutations

The Mutsig program analyses mutations from a cohort of mutations to identify candidate driver genes (Lawrence et al., 2013). Mutsig was utilised on TCGA primary colorectal, ductal breast and lung adenocarcinomas to identify drivers in these primary cancers. Candidate driver mutations were used to screen the brain metastases. Only genes with an FDR<0.05 were considered to be of significance.

2.14.7 Analysis of structural variants

The algorithm BreakDancer was used to identify structural variations in cancer genome analysis (Chen et al., 2009). BreakDancer version 1.1.2 was utilised to try and identify regions of genomic rearrangement in the brain metastases. Reads were filtered to achieve a minimum of 20 supporting read pairs with a threshold quality score of 50. Analysis was subsequently focused on break point regions to reduce false positive calls. Candidate regions fulfilling these criteria were subsequently visualised on integrative genome viewer (IGV).

2.14.8 Identification of regions of kataegis

Kataegis are localised regions of hypermutation are characterised by short intermutation distances between somatic mutations. The mechanisms underlying these mutation pattern remains unclear, though the APOBEC family of cytidine deaminases have been linked with this phenomenon.

Regions of kataegis were identified using the R Bioconductor program Somatic Signature version 3.0 (http://bioconductor.org) (Figure 2.7). Whole genome mutations were filtered as described in section 2.11.3-2.11.4. Somatic SNVs were exclusively selected to increase the confidence of identifying regions of kataegis. Mutations across the whole genome were subdivided according to chromosome and the intermutation distance of mutations of each chromosome was plotted using Somatic Signature. The plots from each chromosome were visualised and screened manually. Data from chromosomes with regions demonstrating kataegis were the analysed in further detail.

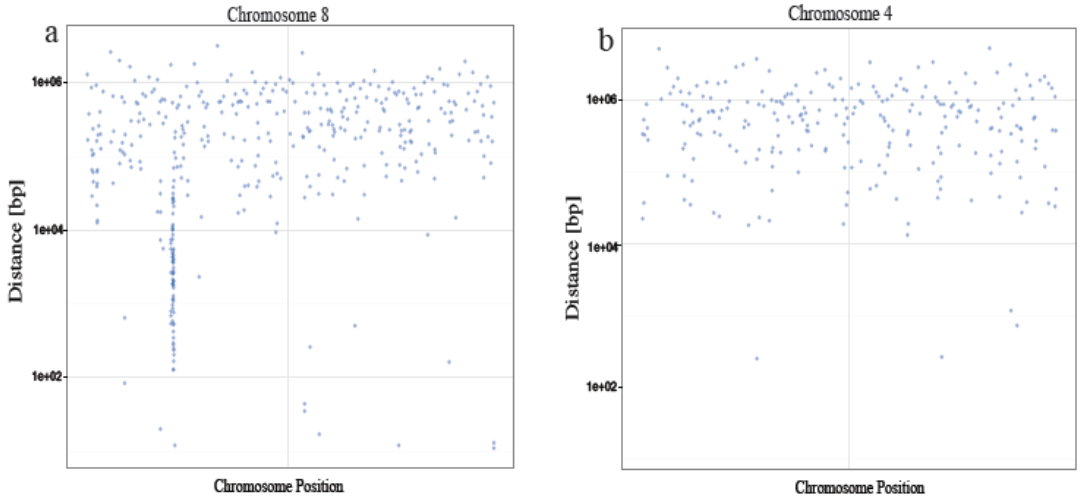


Figure 2.7 l Example of kataegis identified in breast brain metastasis. A region of kataegis was identified by plotting the intermutation distance of somatic mutations against chromosome position. a) Region of Kataegis identified in chromosome 8 with a cluster of somatic mutations over a short interval. b) No evidence of kataegis is present in chromosome 4. Intermutation distance was plotted for all chromosomes and manually inspected

2.14.9 Identification of microsatellite instability

Microsatellite instability (MSI) is associated with hypermutability which results from defects in the DNA mismatch repair system (Martin et al., 2010). DNA mismatch repair enzymes correct errors that arise spontaneously in DNA namely, single base mismatches and short insertions and deletions. DNA mismatch repair enzymes identify these errors and excise the erroneous nucleotides and repair the strand. Consequently defects in the DNA mismatch repair (MMR) pathway results in the accumulation of these errors. DNA MMR impairment results in a failure to preserve the genomic DNA sequence when DNA is replicated. Microsatellites are repeated sequences of genomic DNA consisting of 1-6 nucleotides in length. Each individual has a set length of microsatellites but there can be variability amongst individuals. Microsatellite instability MSI has been associated with several cancers. In particular in colon cancer the presence of MSI has been associated with a good prognosis in a subgroup of patients. Although historically MSI has been identified using PCR-electrophoresis based methods, improvements in bioinformatic techniques enables researchers to identify microsatellite instability from WGS data.

To screen for MSI in the whole genome using sequenced data from the colorectal brain metastases, MSIsensor a C++ program was utilised (Niu et al., 2014). The program functions to detect the length of microsatellites in tumour and normal BAM files at each site. The observed distributions across the genome were compared statistically to determine the presence or absence of MSI. In parallel BAM files were screened by manual inspection to examine well documented sites of microsatellite instability (BAT 26, BAT 25, D2S123, D5S346, and D17S250) to reaffirm the MSI

status of the tumour. The MSI status of the liver metastases was available and had been confirmed through PCR-electrophoresis.

2.14.10 GISTIC analysis

Beroukhim et al developed an algorithm to assess the significance of chromosomal aberrations in cancer genomes (Beroukhim et al., 2007). The genome identification of significant targets in cancer (GISTIC) is an algorithm that statistically identifies regions of copy number aberrations over background aberrations by taking into account the frequency and magnitude of copy number aberrations in a cohort of tumour samples. GISTIC analysis was undertaken using the Nexus Biodiscovery version 7.0. Tumour and normal BAM files were run simultaneously using discrete windows of 10KB. Threshold parameters for analysis included a G-score of <1 and q-value 0.05. To increase the reliability of candidate regions identified loci corresponding to segmental duplications were excluded, as this can be a source of mapping artefacts resulting in false positive results.

2.14.11 STAC analysis

Diskin et al developed a method to characterise the significance of copy number changes in cancer genomes (Diskin et al., 2006). The algorithm significance testing for aberrant copy number (STAC) is method to identify recurrent regions of copy number aberrations that are unlikely to occur randomly. The aberrations identified were computed for the likelihood of finding that aberration at that frequency at a given loci. The likelihood of the event is statistically evaluated generating a p-value. Regions corresponding to segmental duplications were excluded.

2.14.12 Interrogation of non-coding regions

2.14.12.a Transcription factor binding sites

Somatic mutations across the genome were interrogated for recurrent non-coding mutations. Evaluation of non-coding regions focused on targeting recurrent mutations that occurred within known transcription factor binding sites using the ENCODE database (https://www.encodeproject.org). Candidate loci were analysed further to determine the degree of mappability in these regions using mappability scores from UCSC.

2.14.12.b MicroRNA database

Somatic mutations identified in the brain metastases were screened to identify recurrent mutations in the microRNAs. For the purpose of this analysis the microRNA database miRBase v20 (http://www.mirbase.org) was utilised using BEDTools suite v2.17 (Quinlan and Hall, 2010) to identify recurrent mutations.

2.15 Analysis of technical validation

The Ion Torrent PGM analyser was utilised to technically validate somatic mutations where available using a custom gene panel of 82 cancer related genes, with the aim to achieve a mean read depth of ~1000. Appendix A contains the list of genes on the cancer panel. Somatic variants detected using the Ion Torrent platform were identified through automatic calls from the Torrent server output (Life Technologies) and through visual inspection on IGV.

2.16 Bioinformatic analysis of SNP arrays

2.16.1 Method to obtain comparable SNP densities across platforms

Fresh frozen brain and liver metastases were processed using several SNP array platforms as outlined in section 2.8. The platforms used and SNP densities are summarised in Table 2.4.

Platform	SNP Density	Build
Human Omn1 Quad v1	1,134,514	hg build 37
Human CytoSNP 12 v2.1	298,563	hg build 37
Human Omni2.5-4 v1	2,450,000	hg build 36
Human 610 Quad v1	620,902	hg build 37

Table 2.4 l Summary of SNP array platforms and density of probes.

I devised a method to compare samples from different platforms [Human Omn1 Quad v1, Human CytoSNP 12 v2.1, Human Omni2.5-4 v1 and Human 610 Quad v1] at a comparable resolution. Firstly the co-ordinates for platforms in GRCh36 were converted to GRCh37 using LiftOver (http://genome.ucsc.edu/cgi-bin/hgLiftOver). SNP probes were filtered to identify probes in common amongst the 4 platforms with an rs id yielding 121,484 SNPs. The number of remaining SNPs on the Human CytoSNP 12 v2.1 (the platform with the lowest probe density) with a minor allele frequency >10% was identified to be 111,297. Combining 121,484 and 111,297 SNPs yields a potential density of 232,781 SNPs. For each of the other platforms SNPs with a minor allele frequency greater than 10% were identified and 111,297 probes were selected at random and combined with the probes directly in common for each platform. This approach yielded a comparable resolution of 230,000 probes amongst the genotyped arrays.

2.16.2 OncoSNP analysis

OncoSNP v1.3 a software tool for characterising copy number alterations and LOH from cancer samples from SNP genotyping data was run using the filtered array files section 2.14.1 (Yau et al., 2010). The software infers ploidy and degree of normal cell contamination producing models which best fit the array data. Each model is

assigned a log-likelihood ratio, a measure of how best each model fits the array data. Output files were systematically analysed by manual visual inspection to select the most appropriate model to obtain information regarding copy number and LOH. Raw array data files obtained from TCGA primary breast, colorectal, lung adenocarcinoma primary tumours were obtained. These had been processed by the TCGA on the Affymetrix Genome-Wide SNP Array 6.0 platform (SNP density of ~900,000) build hg 36. The build was converted to build 37 using LiftOver and SNP density was downsized to approximately 230,000 probes and run on OncoSNP v1.3. The TCGA data set was used for comparison.

2.16.3 Consensus clustering of SNP arrays

SNP arrays from FFPE tumours were processed using OncoSNP to evaluate copy number alterations. Segmental gains and losses for each tumour were identified in relation to the mean copy number of the tumour. The genome was divided in to bins of 500,000 bases and each segment was assigned a copy number state (gain, loss and neutral). Copy number alterations identified in this process were clustered using consensus clustering, and the optimal number of clusters was identified from a consensus matrix identifying an optimal k value (Wilkerson and Hayes, 2010).

2.17 Statistical analysis

Statistical analysis in this thesis was conducted using the statistical package GraphPad Prism version 6.

2.18 Summary of samples used

I have summarised the tumour samples that were sequenced and had arrays performed in Tables 2.5-2.6. An overview of the comparisons amongst tumours are outlined in the tables.

Sample	Tumour type	Platform Read depth	Comparison with TCGA primaries (n=300)	Mutation burden analysed	Mutation Spectra and Signatures	Driver Mutations in brain metastasis and primary identified	Clonal architecture reconstructed	Identification of structural Variants	Identification of copy number aberrations from sequencing data
C1MET	Colon	Illumina Hiseq2000 60X	Yes	Yes	Yes	Yes	Yes	Yes	Yes
C2MET									
C3MET									
C4MET									
P1	Primary Colon (pair to C1MET)	Illumina Hiseq2000 30 X	Compared with paired brain met C1MET down sampled	Yes	Yes	Yes	Yes	Yes	Yes
L1MET	Lung	Illumina HiSeq 2000 60X	Yes	Yes	Yes	Yes	Yes	Yes	Yes
L2MET									
L3MET									
L4MET									
L5MET									
L6MET									
L7MET									
L8MET									
B1MET	Breast	Illumina Hiseq 2000 60X	Yes	Yes	Yes	Yes	Yes	Yes	Yes
B2MET									
B3MET									
B4MET									
B5MET									
B6MET									
B7MET									
LiMET8a	Colorectal liver metastases	Complete Genomics 60X	Comparison with Colorectal brain metastases	Yes	Yes	N/A	No	No	No
LiMET8b									
LiMET9a									
LiMET9b									
LiMET15a									
LiMET15b									
MR1	Primary colorectal Multiregion analysis	Illumina HiSeq 2000 60X	Comparison with C1MET/P1	Yes	No	No	No	No	No
MR2									

Table 2.5 l Summary of sequenced tumours and overview of analysis

Sample Set	Number	SNP array platform	Type of tissue	Comparison with primaries	Short segment copy number alteration analysis	Large scale copy number analysis
colorectal brain metastases	4	Human CytoSNP 12 v2.1	Fresh frozen Discovery cohort	Yes	Yes	Yes
lung brain metastases	7			Yes	Yes	Yes
breast brain metastases	7			Yes	Yes	Yes
colorectal liver metastases	25 (8 are paired)	Human Omn1 Quad v1 and Human CytoSNP 12 v2.1	Fresh frozen	Yes	Yes	Yes
TCGA colorectal primaries	MSI positive (72) MSI negative (502)	Affymetrix Genome-Wide SNP Array 6.0	Fresh frozen	N/A	Yes	Yes
TGCA lung primaries	304			N/A	Yes	Yes
TCGA breast primaries	530			N/A	Yes	Yes
FFPE colorectal brain metastases	6	Human Omniexpress 24 platform	FFPE	No	No	Yes
FFPE lung brain metastases	7			No	No	Yes
FFPE breast brain metastases	9			No	No	Yes

Table 2.6l Summary of tumours analysed for copy number alterations

Chapter 3

Analysis of colorectal cancer brain metastasis genomes

3.1 Introduction

Brain metastases are the most frequently occurring intracranial neoplasm in adults (Patchell, 2003). Tumours of particular histological subtypes have a propensity to metastasise to the brain (Owonikoko et al., 2014). The incidence of brain metastasis in patients with colorectal cancer can range from 1-4% (Tan et al., 2009). With improvements in systemic therapies coupled with improved imaging and access to such resources the number of patients presenting with brain metastasis is likely to increase.

Colorectal cancer is the third most common cancer worldwide (Siegel et al., 2013). Metastatic dissemination in colorectal cancer frequently occurs in a hierarchical order (Wan et al., 2013). The majority of metastases are directed to the liver occurring in 35-55% of patients, with lung metastases forming the next most common site of metastasis. In comparison, metastasis to the brain is less frequent. The unequal distribution of metastatic deposits has been partially attributed to the vascular architecture, where the colon is directly connected to the liver through the mesenteric circulation (Nguyen et al., 2009a). There is a growing body of evidence to suggest that although the anatomical vasculature may play a role in the initial distribution of metastatic founder cells, biological processes have a key role for successful colonisation of target organs (Nguyen et al., 2009a).

To my knowledge, WGS analysis of colorectal cancer brain metastases (CCBM) has not been published to date. In this chapter, I report my analysis of 4 CCBM that have undergone WGS and compare my findings with primary colorectal tumours from published reports and my own analysis of primary tumours that was obtained from TCGA. In addition, I have extended my analysis to a cohort of liver metastases of primary colorectal origin that I have used as a comparative data set.

3.2 Methods

3.2.1 Sequencing of CCBM genomes

DNA from 4 CCBM and paired buffy coat lymphocytes were extracted, quantified and underwent quality assessment as outlined in sections 2.4-2.6. 1µg of DNA was utilised for WGS using the Illumina HiSeq 2000 platform. Paired-end reads of 100 base pairs were used with the aim to achieve a median depth of coverage of 30X and 60X for constitutional and tumour DNA respectively. Reads were mapped using BWA and Stampy. Variant calling and downstream annotation of (single nucleotide variants) SNVs and small indels was undertaken using Platypus and the software engine ANNOVAR. A stringent pipeline was devised to filter sequencing data to provide confident calls. Processing of data is described in detail in sections 2.11.

3.2.2 Downstream analysis of WGS data

Sequencing data were analysed by myself using bioinformatic programs and bespoke analytical scripts. A comprehensive report of this is provided in Chapter 2 in section 2.14, including the parameters that were utilised in data analysis. Analysis of data involved: evaluating mutation burdens and spectra/signatures, screening for driver mutations and structural variants and modelling the clonal structures of the CCBM.

3.3 Comparison Datasets

3.3.1 Somatic mutations from TCGA primary colorectal tumours

Somatic mutations from primary colorectal tumours were obtained from the TCGA (section 2.2.1). This information was utilised to characterise the mutation spectra in 300 primary tumours. Candidate driver mutations in primary colorectal cancers were obtained using IntOGen and Mutsig analysis (FDR <0.05) using colorectal tumours obtained from the TCGA cBioPortal (sections 2.14.5-2.14.6). This list of primary colorectal cancer drivers was used to screen the CCBM to determine if these drivers were retained in the brain metastases. To identify putative drivers of brain metastasis in the CCBM, IntOGen analysis was undertaken using the sequenced CCBM (FDR <0.05).

3.3.2 Copy number information from primary colorectal tumours

The sequenced CCBM were genotyped using Illumina SNP arrays. Copy number information from primary colorectal cancers was used as a comparison data set. Raw data from primary colorectal tumours from TCGA, which were genotyped on Affymetrix arrays were obtained. All SNP array data were processed using OncoSNP to obtain information regarding copy number and LOH. The processing of these data and the method to establish a common SNP density amongst the array platforms for comparative analysis is reported in detail in section 2.14. In addition for validation, copy number information from sequenced tumours was inferred directly from sequencing data using OncoSNP-SEQ.

3.3.3 Colorectal cancer metastases to the liver

Colorectal cancer liver metastases (CCLM) were obtained from the Royal Free Hospital (London) and John Radcliffe hospital (Oxford) by a colleague Dr Rajarajan Ramamoorthy. 3 paired CCLM underwent WGS using the Complete Genomics platform with a median depth of coverage of X60 X30 in tumour and constitutional

DNA. Data analysis of the liver metastases was undertaken by myself. A detailed description of the processing steps is outlined in section 2.12.

In addition to sequencing, a cohort of CCLM were genotyped using Illumina SNP arrays. I developed a method to compare data from the colorectal brain and liver metastases and primary colorectal tumour SNP arrays genotyped on different platforms. A detailed description of the bioinformatic processing steps involved is outlined in section 2.16. Of the paired liver metastases, those with a suffix “a” denotes the first metastasis in the pair in time. Information regarding the time interval between the paired liver metastases was unavailable.

3.4 Technical validation of sequencing data

Selected somatic variants underwent targeted verification using the Ion Torrent PGM analyser. A custom panel of 82 cancer related genes was utilised. Library construction and data analysis was undertaken by myself (section 2.7 and 2.15). 100 ng of DNA was used for sequencing, with a mean read depth of ~1000. Variants detected using the Ion Torrent platform were identified through automatic calls from the Torrent server output and through visual inspection on IGV.

3.5 Patient Demographics

Clinical information pertaining to the patients who had their brain metastases sequenced is reported in Table 3.1. Information was collected where available using an independent party in accordance with ethics to maintain patient anonymity. Information pertaining to the CCLM that were analysed is summarised in Table 3.2. Treatment data for the CCLM was unavailable.

Colorectal brain metastasis	Age of patient (Years)	Gender	Histology of primary tumour	Time interval between resection of brain metastasis and primary tumour (Years)	Treatments received by patient following primary tumour
C1MET	54	M	Adenocarcinoma (moderately differentiated)	2	Adjuvant Folfox. On progression exposure agents including irinotecan and capecitabine and avastin
C2MET	64	M	Adenocarcinoma (moderately differentiated)	3	Adjuvant Folfox
C3MET	60	F	Adenocarcinoma	6	No treatment data available
C5MET	45	M	Adenocarcinoma	1	Rectal cancer chemoradiotherapy adjuvant capecitabine.

Table 3.1 | Summary of clinical information for patients who had their brain metastasis sequenced

Liver Metastasis	Age of patient	Gender	Location of primary colorectal tumour	Histological grade of primary colorectal tumour
LiMET1	54	F	Left descending colon	Unknown
LiMET2	65	M	-	Unknown
LiMET3	59	M	Hepatic flexure	2
LiMET4	54	M	Rectum	Unknown
LiMET5	59	F	-	Unknown
LiMET6	71	F	Rectum	Unknown
LiMET7	66	M	Rectum	Unknown
LiMET8a	53	M	Sigmoid	2
LiMET8b				
LiMET9a	30	M	Sigmoid	2
LiMET9b				
LiMET10a	21	F	Sigmoid	2
LiMET10b				
LiMET11a	71	M	Ascending colon	2
LiMET11b				
LiMET12a	65	F	Sigmoid	2
LiMET12b				
LiMET13a	44	F	Recto-sigmoid	2
LiMET13b				
LiMET14a	62	M	Ascending colon	2
LiMET14b				
LiMET15a	65	M	Recto-sigmoid	Unknown
LiMET15b				
LiMET16	54	M	Recto-sigmoid	2
LiMET17	73	M	Sigmoid	Unknown

Table 3.2 l Summary of clinical information for patients who had their liver metastases resected. Paired tumours are identified with a suffix with the letters a and b.

Results

3.6 Mutation burdens of CCBM

Across the genome the CCBM had a median number somatic mutations of 31,319 [range 13,447 to 69,984]. Of these mutations, the median number of SNVs was 27,168 [range 10,492 to 63,111]. The mutation burdens in the CCBM are represented in Figure 3.1. The number of small indels in the brain metastases ranged from 2,955 to 6,873 [median 4,150]. C1MET had the greatest mutation burden amongst the CCBM (69,984 somatic mutations).

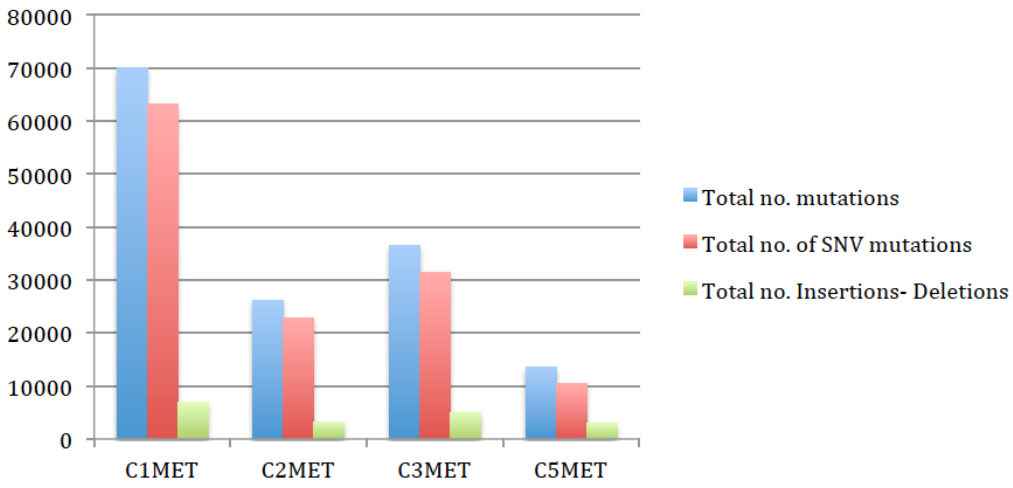


Figure 3.1 I Mutation burdens in the CCBM in the whole genome. The total number of mutations, SNV mutations and the number of small indels are depicted. C1MET has the greatest mutation burden.

The distribution of mutations across protein-coding and non-coding regions were assessed. The majority of somatic SNVs were located in intergenic regions [range 8,775 to 43,340] forming 65% to 70% of the overall mutation burdens. The proportion of mutations in intronic regions ranged from 30% to 34%. Mutations located within the exome formed ~1% of the overall mutation burdens in the CCBM. The number of mutations in the exome ranged from 2.5 to 20 per 10^6 (Figure 3.2).

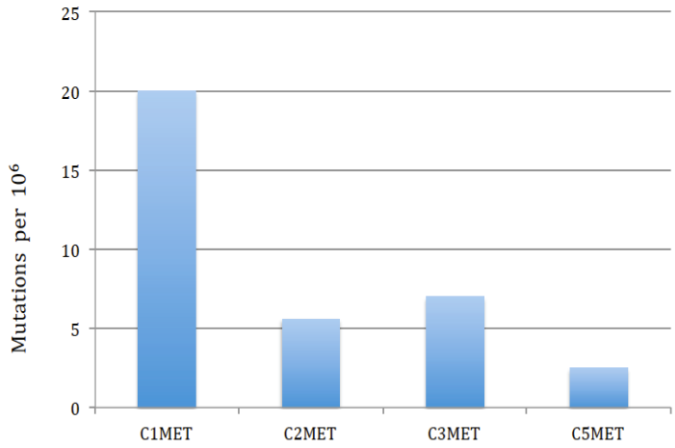


Figure 3.2 I Number of mutations in exome in CCBM. Mutations are represented per megabase. C1MET has the greatest mutation burden.

In coding regions the proportion of synonymous mutations ranged from 19% to 32% [median 23%]. To determine the number of potentially functional mutations within protein-coding regions sequencing data were filtered (nonsynonymous, frameshift, splice-site and nonsense mutations were retained). A median of 131 [range 60 to 469] mutations were potentially functional. Nonsynonymous mutations were the most common, forming 22 to 75% of the overall exonic mutation burden (Figure 3.3). Proportionately amongst the CCBM, the number of nonsynonymous mutations was greatest in C1MET [75% range 22-75%]. Splice-site mutations were not identified amongst the CCBM. The number of small indels, within coding regions ranged from 4 to 15 events per tumour [median 12.5]. The number of nonsense mutations detected ranged from 4 to 15 events per tumour [median 7].

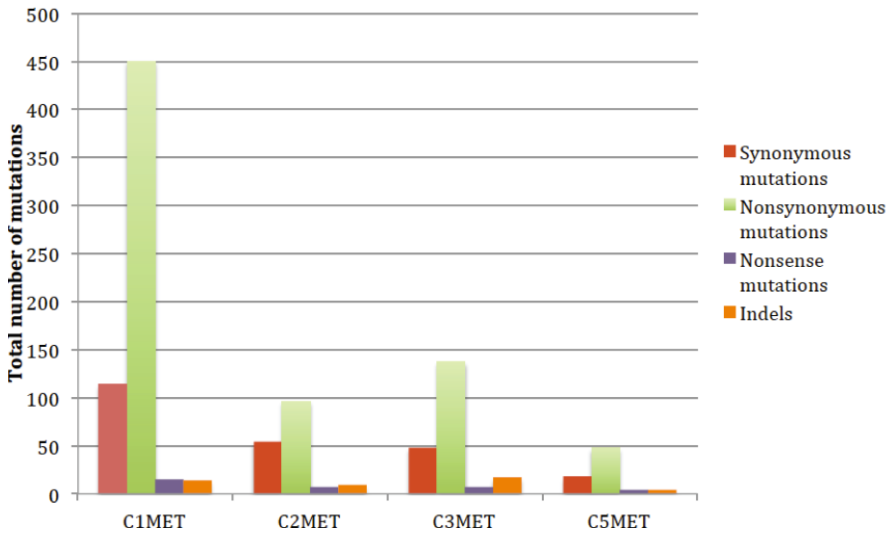


Figure 3.3 I Distribution of types of protein-coding mutations in CBBM.

Nonsynonymous mutations are the most frequent type of mutation.

The mean nonsynonymous:synonymous mutation ratio was 2.7:1 The transition transversion ratios ranged from 0.4 to 0.9.

Collective analysis of the mutation burdens demonstrated that C1MET had the highest mutation burden amongst the CCBM. C1MET was hypermutated with a mutation burden of 20 per 10^6 in the exome and an overall mutation burden of 69,984. The time interval from resection of the primary tumour (2 years) was less than other CCBM [range 1 to 6]. This suggests that the increased mutation burden was not reflective of the duration in time between the occurrence of the primary tumour and resection of the brain metastasis.

3.7 Screen for microsatellite instability and *POLE* mutations

Hypermutated samples in primary colorectal tumours are typically associated with defective DNA mismatch repair machinery, resulting in microsatellite instability (MSI), or with somatic *POLE* mutations (Cancer Genome Atlas, 2012). C1MET was confirmed to be MSI negative by using two methods in concert. MSI sensor, a

bioinformatic tool which identifies genome wide somatic microsatellite changes in paired tumour-normal sequences was run using tumour and normal BAM files (Niu et al., 2014). In addition, using IGV I screened 5 established microsatellite loci (BAT25, BAT26, D2S123, D5S346, D17S250) manually by visual inspection (Boland et al., 1998). Both methods provided confirmatory evidence that C1MET was MSI negative. Somatic *POLE* mutations were not detected in C1MET. Analysis of germline variants failed to identify *POLE* and DNA mismatch repair gene mutations. Extension of the analysis to the remaining CCBM confirmed that they were MSI negative.

3.8 Analysis of mutation spectra

The mutation spectra of tumours can provide insight with regards to an imprint of biological processes in action, namely, defective DNA repair pathways or exposure to mutagenic agents (Alexandrov et al., 2013, Lawrence et al., 2013). To evaluate this further the proportion of the 6 classes of base substitutions (C>A/G>T, C>G/G>C, C>T/G>A, T>A/A>T, T>C/A>G, T>G/A>C) were characterised in the CCBM using somatic SNVs across the genome.

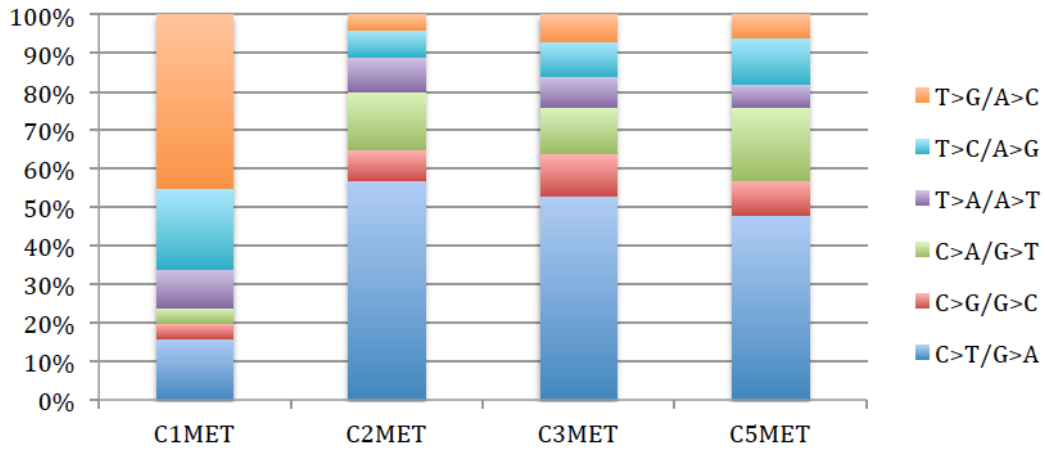


Figure 3.4 I Proportion of base substitutions in CCBM in the exome. T>G/A>C mutations predominate in C1MET.

Analysis of mutations across the genome confirmed that generally C>T/G>A mutations were the predominant mutation profile in the CCBM median 23% [range 9 to 30%]. This proportionately increased in exome where the median number of C>T/G>A mutations increased to 51% [range 16 to 57%] (Figure 3.4). This is consistent with previously reported findings in colorectal cancer with 59% of mutations in coding regions consisting of C>T/G>A transitions which predominantly occurred at 5'-CpG-3' sites (Sjoblom et al., 2006). Deamination of 5-methylcytosine results in the formation of TpG dinucleotides (Simmen, 2008, Waters and Swann, 2000). Cazier et al reported a proportionate increase in C>T/G>A mutations in the exome compared to the whole genome in a cohort of primary bladder cancers which was attributed to an increase in G:C content in the exome (Cazier et al., 2014a). A notable exception amongst the CCBM was the mutation footprint of C1MET, where T>G/A>C mutations predominated forming 51% of base substitutions in the whole genome and 45% of mutations in the exome.

3.9 Analysis of mutation signatures

To gain further insight into the context in which base substitutions occur in the CCBM, SNV mutations were analysed in sequence context taking into account the base flanking the 5' and 3' end of each SNV. This provides a total of 96 possible motif combinations. Mutation signatures were generated which take into account the relative contribution of each motif within a tumour using Somatic Signature. Through this process two signatures were generated, S1 and S2. (Figure 3.5). C1MET was associated with signature S1, where T>G/A>C mutations in C1MET were overrepresented in the context of a C[T]T motif. In contrast, the mutation signatures observed in the remaining CCBM resembled signature S2. Alexandrov *et al* undertook a comprehensive analysis of mutation signatures in a variety of tumours of

different histological subtypes (Alexandrov et al., 2013). Signature S1 in my analysis resembles signature [17] in the analysis undertaken by Alexandrov et al. This signature constituted 1.8% of all signatures detected in the pan-cancer analysis.

Cohesin, a protein complex that regulates the separation of sister chromatids in DNA replication, co-localises with several transcription factors including CTCF (CCCTC-binding factor)(Parelho et al., 2008, Rubio et al., 2008, Kagey et al., 2010). This provides a regulatory mechanism to control genomic stability, gene expression and epigenetic regulation. Recently Katainen et al characterised the mutation footprint of 213 colorectal cancers using WGS in combination with chromatin immunoprecipitation sequencing (ChIP-seq). They identified 3 mutations signatures in their analysis. The most prevalent signatures correlated with signatures [S1] and [S17] as reported by Alexandrov et al. They reported that amongst the colorectal tumours exhibiting signature 17 there was a predominance of A>C/T>G and A>G/T>C mutations in CTCF/cohesion binding sites. Moreover, mutations in these sites had a tendency to occur in regions of the genome that were late replicating particularly in the S phase of DNA replication. They analysed the mutation footprint in hypermutated tumours with pathogenic mutations in the proof reading domain of *POLE* (Pol ε exo$^-$). Pol ε exo tumours demonstrated a reduced frequency of mutations in CTCF/cohesion binding sites implying that these sites are not replicated by *POLE*. The underlying aetiology for signature 17 remains unknown and though it has been proposed that oxidative stress could result in this signature (Katainen et al., 2015). Signature S2 in my analysis did not closely resemble the other signatures generated by Alexandrov et al.

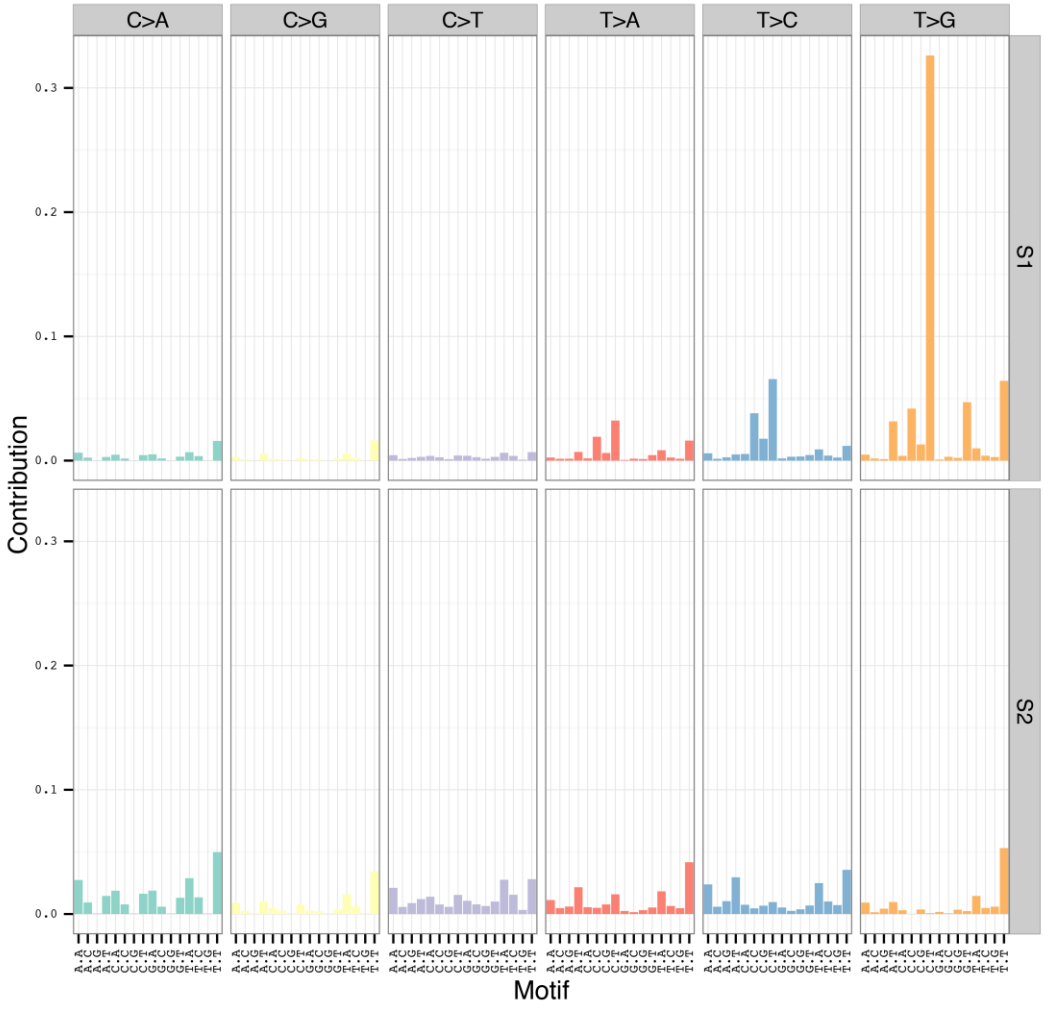


Figure 3.5 I Mutation signatures detected in CCBM. C1MET is associated with signature S1, whereas the remaining CCBM are associated with signature S2. T>G/A>C mutations in C1MET occurred in the context of a C[T]T motif.

3.10 Identification of putative drivers of CCBM

Two approaches were adopted to identify putative drivers of brain metastasis in the CCBM. Firstly, I ran the IntOGen algorithm on the sequenced brain metastases (FDR <0.05). Secondly, I screened all truncating mutations using a threshold of 3 more occurrences in the same gene. No putative drivers of brain metastasis were identified through these processes in the CCBM.

3.11 Screen for drivers of primary colorectal cancer

Although drivers of CCBM were not identified, I sought to determine if drivers of primary colorectal cancer were present in the CCBM. Candidate drivers of primary colorectal cancer were identified through IntOGen and Mutsig analysis using a large cohort of primary tumours (n=389) obtained from TCGA cBioPortal (FDR <0.05). The list of candidate genes was used to screen somatic mutations in the CCBM.

Candidate somatic mutations were filtered further to exclude base substitutions with potentially benign or moderate functional effects. (Polyphen 2 > 0.8 and SIFT score < 0.2). Mutations fulfilling the aforementioned criteria were authenticated by visual inspection on IGV and technically validated where possible. Copy number and LOH information obtained from genotyped SNP arrays processed with OncoSNP were integrated with the mutations identified.

From the sequenced CCBM, *APC* (n=3), *TP53* (n=3), and *KRAS* (n=3) mutations were detected from the 'classical' drivers of colorectal tumourigenesis (Table 3.3) (Fearon, 2011). *KRAS* mutations were found in activating mutation hotspots, namely codons 12, 13 and 117 which result in constitutive activation of the RAS GTPase and downstream signalling (Ellis and Clark, 2000, Smith et al., 2010). Other putative drivers of colorectal cancer identified in this process in the CCBM include *PIK3R1* (n=1), *ARID1A* (n=1), and *TCF7L2* (n=2). A recent comprehensive analysis of 276 primary colorectal tumours identified *ARID1A* and *TCF7L2* as putative drivers of primary colorectal cancer using the software MutSig (Muzny et al., 2012).

Gene	Sample	Location of mutation	Type of mutation	LOH at locus of mutation	Copy Number at locus of mutation	PolyPhen2 Score	SIFT Score	Mutation frequency in gene TCGA cBioPortal %	Mutation frequency in gene at that loci -TCGA cBioPortal
APC	C2MET	exon8:c.646 C>T:p.R216X	Nonsense	No	2	--------	--------	76% (161/212)	7
	C3MET	exon17:c.4216 C>T:p.Q1406X	Nonsense	Yes	1	--------	--------		2
KRAS	C1MET	exon2:c.35 G>C:p.G12A	Nonsynonymous	No	2	0.95	0	42% (89/212)	1
	C2MET	exon2:c.38 C>T:p.G13D	Nonsynonymous	Yes	4	0.83	0		10
	C3MET	exon4:c.351 A>T:p.K117N	Nonsynonymous	No	3	1	0.1		2
TP53	C1MET	exon5:c.524_535del:p.175_179del, GGTGGGGGCAGC	In frame deletion	Yes	1	--------	--------	52% (112/212)	0
	C3MET	exon5:c.527 G>T:p.C176F	Nonsynonymous	Yes	1	1	0		1
	C5MET	exon6:c670 G>T:p.E224X	Nonsense	Yes	2	--------	--------		0
PIK3R1	C1MET	exon9:c.1040 T>G:p.L347R	Nonsynonymous	Yes	1	1	0	6% (12/212)	0
ARID1A	C3MET	exon2:c.1339_1340 insATAC:PY447fs	Frameshift insertion	Yes	1	--------	--------	9% (20/212)	0
TCF7L2	C1MET	exon13:c.1337 A>Cp.K446T	Nonsynonymous	No	1	0.96	0	13% (29/212)	0
	C2MET	exon7:c.805 A>Tp.R269W	Nonsynonymous	No	3	0.96	0		0

Table 3.3 I Primary driver mutations in colorectal cancer in the CCBM

3.12 Clonal architecture of CCBM

Pyclone which uses a Bayesian clustering method was utilised for inferring the structure of clonal populations in the CCBM (Roth et al., 2014). A similar approach has been adopted by other researchers (Shah et al., 2012, Cazier et al., 2014b). The number of clonal populations reported in each brain metastases is listed in Table 3.4. The number of clonal populations ranged from 7 to 35 [median 20.5].

Colorectal brain metastasis	Number of clonal clusters	Number of somatic SNVs
C1	35	64460
C2	17	22930
C4	24	31407
C5	7	10492

Table 3.4 | Number of clonal populations in CCBM. C1MET had the most complex clonal structure.

Amongst the CCBM, C1MET had the most complex clonal architecture with 35 predicted clonal populations. In comparison, C5MET was relatively clonally "simple" with 7 clonal populations. The complexity of the genomes increased with the number of somatic mutations. The clonal architecture of the brain metastases are depicted in Figures 3.6-3.9. The position driver genes of primary colorectal cancer are highlighted within the framework of the clonal structures. Driver genes of primary colorectal cancer are highlighted in blue alongside randomly selected somatic mutations.

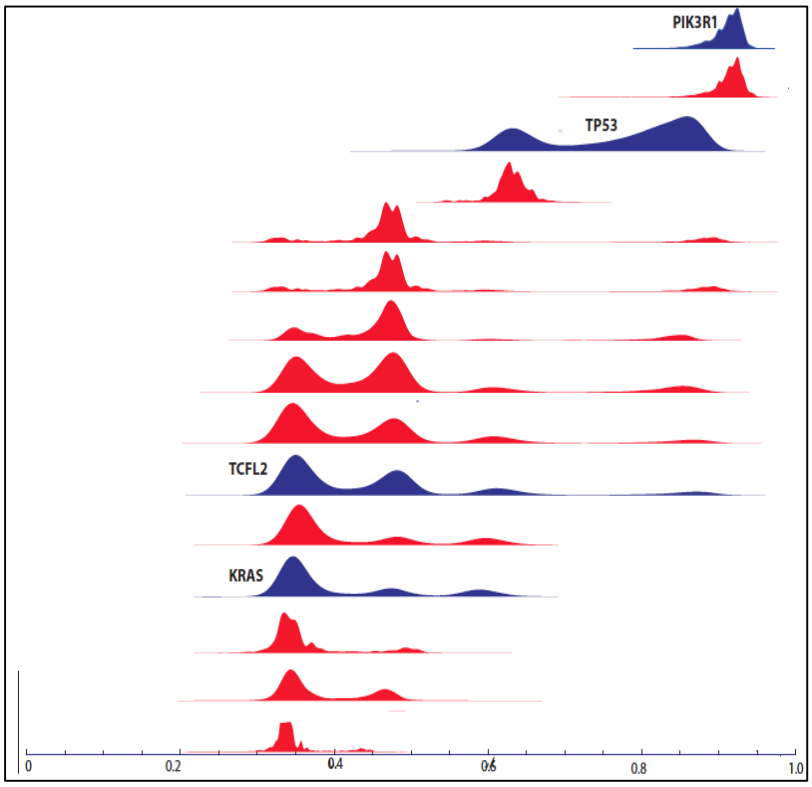


Figure 3.6l Clonal architecture of C1MET. *PIK3R1* is located in the major clone of C1MET, whereas *KRAS* and *TCF7L2* are located in subclonal positions. The clonal positions of randomly selected mutations are depicted (driver mutations are depicted in blue).

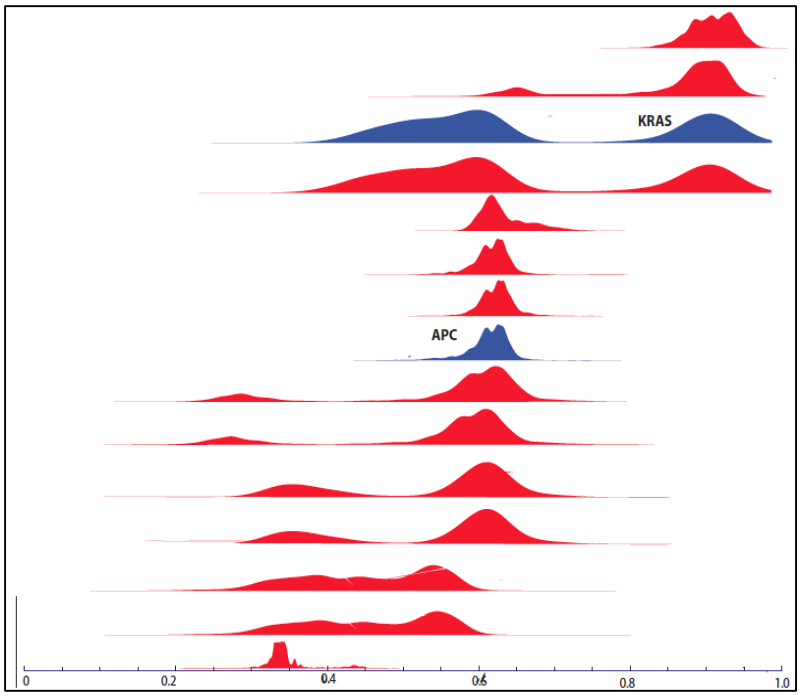


Figure 3.7l Clonal architecture of C2MET. *KRAS* is found within a major clone in C2MET whereas *APC* is subclonal.

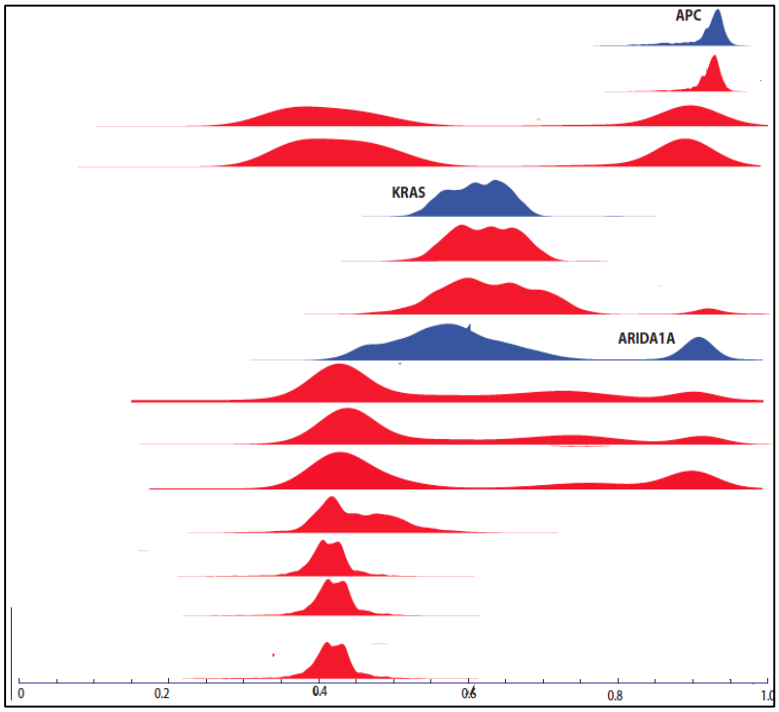


Figure 3.8 Clonal architecture of C3MET. *APC* is located in major clone. *KRAS* and *ARID1A* are subclonal.

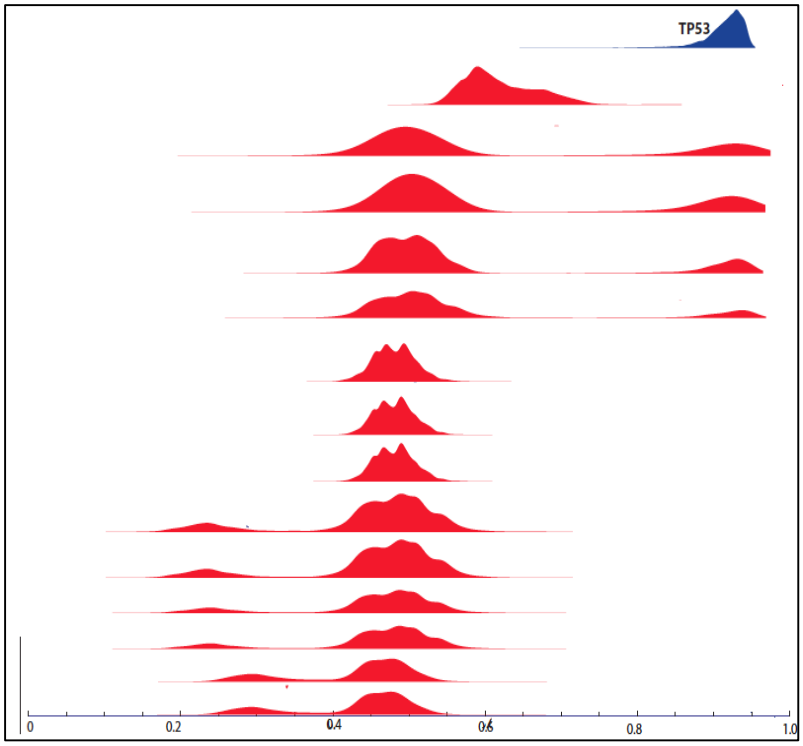


Figure 3.9 l Clonal architecture of C5MET. *TP53* is located within the major clone in C5MET.

Driver mutations of primary colorectal cancer were located in both major clonal groups and subclones. In C1MET, which had the most complex clonal structure *PIK3R1* and *TP53* were located in major clones, whereas *KRAS* and *TCF7L2* were subclonal. *TCF7L2* has been recently identified as a driver of colorectal cancer (Bass et al., 2011). *PIK3R1* and *TP53* mutations were found in regions encompassing LOH and this was reflected in the variant allele frequencies at these loci 88% and 84% respectively. The variant allele frequency (VAF) in the *TCF7L2* and *KRAS* were 33% and 35% in C1MET. *APC* was located in major clone in C3MET, which harboured a pathogenic nonsense mutation Q1406X encompassing a region of LOH. In comparison, in C2MET *APC* was positioned in a subclone. The pathogenic nonsense mutation R216X in C2MET encompassed a diploid region of the genome. *KRAS* was present in a major clone in C2MET encompassing a region of copy number gain. *ARID1A* has recently been considered as a putative driver of several cancers including primary colorectal adenocarcinomas (Cancer Genome Atlas, 2012). A frameshift mutation in *ARID1A* was identified in C3MET, which was located predominantly in a subclonal position in the genome.

3.13 Evaluation of copy number alterations

Genotyped SNP arrays of the CCBM were analysed using OncoSNP. Copy number variation can be considered as a surrogate for chromosomal instability. Deviations from the average copy number were identified in each tumour, herein defined as a copy number alterations (CNA). The brain metastases had a median of 28 CNA [range 20 to 31]. The proportion of the genome that deviated from the mean copy number ranged from 19 to 43 [median =27] with the least deviation observed in C1MET.

Chromosome-scale aberrations were characterised in each tumour (defined as >90% involvement of the arm). Due to the modest sample size, recurrent chromosome-scale

changes were categorised as those present in 3 or more tumours. Recurrent chromosome-scale gains were observed in 7p (n=4 C1MET, C3MET, and C5MET), 13q (n=3 C2MET, C3MET, C5MET), 20q (n=3 C2MET, C3MET, C5MET). Recurrent chromosome-scale deletions were observed in 18p (n=3 C2MET, C3MET, C5MET). Whole chromosome gains were observed in C3MET in chromosome 7 and 16, whereas whole chromosome deletions were present in C3MET in chr4. Amongst the CCBM, C1MET had the least number of chromosome-scale aberrations.

Tumour	Number of CNA	Proportion of genome deviating from mean CN	Chromosome-scale deletions	Chromosome-scale gains
C1MET	20	19	------	7p
C2MET	31	28	4p, 17p, 18p 5q, 18q, 21q, 22q	13q, 20q
C3MET	26	43	chr4, 5p, 8p,18p, 2q, 4q, 15q, 22q	chr7, chr16, 12p, 20q 13q,
C5MET	30	26	17p, 18p 6q, 14, 20q	7p, 12p, 13q, 20q

Table 3.5 l Summary of copy number aberrations in individual CCBM

3.14 Identification of recurrent focal copy number aberrations

Recurrent focal copy number aberrations (< 2MB) were identified in a genome wide screen of the CCBM (Figures 3.10-3.11). The aim was to identify deletions or amplifications which could serve to activate or inactivate oncogenes or tumour suppressor genes respectively. Focal deletions were detected in *RBFOX1* (C1MET and C5MET) and *MACROD2* (C1MET and C2MET) in the CCBM.

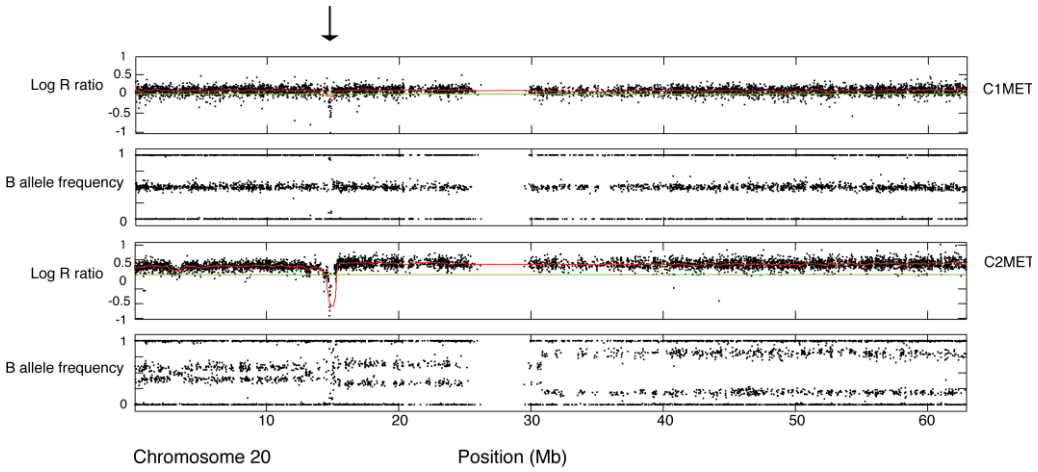


Figure 3.10 l Focal deletions in *MACROD2*. Deletions were detected in C1MET and C2MET

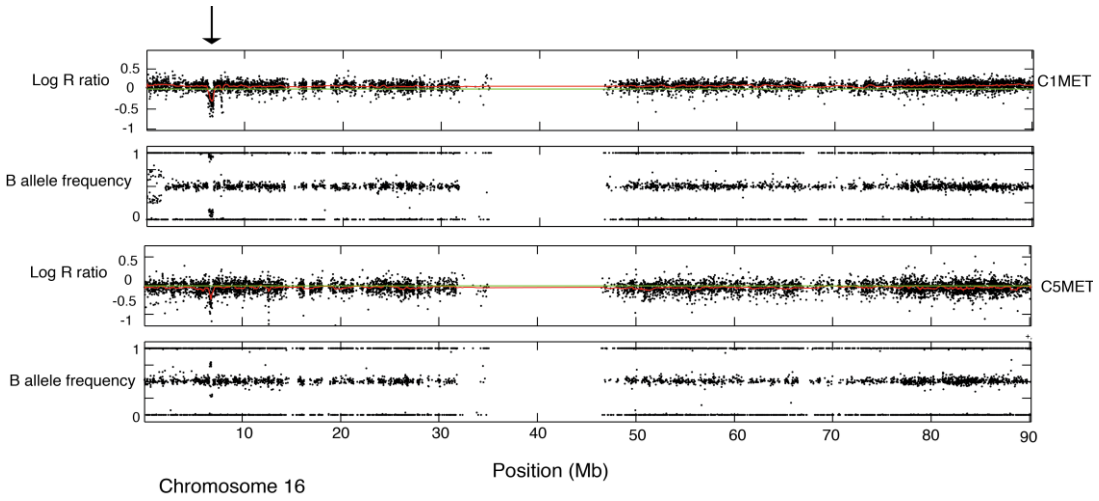


Figure 3.11 l Focal deletions in *RBFOX1*. Deletions were detected in C1MET and C5MET.

Chromosome fragile sites are loci in metaphase chromosomes that have a predisposition to develop breakpoints when the cell is exposed to replication stress. They are frequent sites of chromosome breakage and rearrangements in cancer cells (Durkin and Glover, 2007). The TCGA reported deletions in chromosome fragile sites in *FHIT*, *RBFOX1* and *WWOX* in a comprehensive analysis of primary colorectal tumours (Cancer Genome Atlas, 2012).

3.15 Concordance of copy number information from SNP array and sequencing data.

To validate findings from SNP array data, copy number information and LOH was inferred directly from sequencing data using OncoSNP-SEQ. Focal deletions in *RBFOX2* and *MACROD2* were verified in the OncoSNP-SEQ analysis (Figure 3.12).

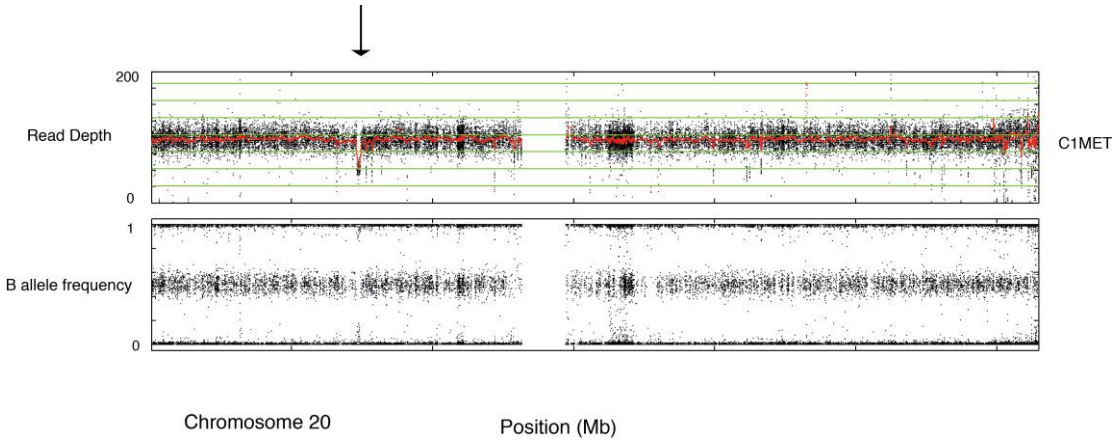


Figure 3.12 l Focal deletion in *MACROD2* in C1MET detected by OncoSNP-SEQ.

The deletion in C1MET in *MACROD2* was verified directly from sequencing data.

The degree of concordance between copy number calls generated using SNP array information and that directly inferred from sequencing data were evaluated. SNP arrays calls generated by OncoSNP were used as a reference to assess the degree of concordance of calls with OncoSNP-SEQ. The degree of concordance ranged from 70 to 79% [median 78%].

3.16 Structural Variants

A genome-wide screen of chromosomal aberrations identified chromothripsis-like alterations in C5MET restricted to a single arm in chromosome 10q (Figure 3.13).

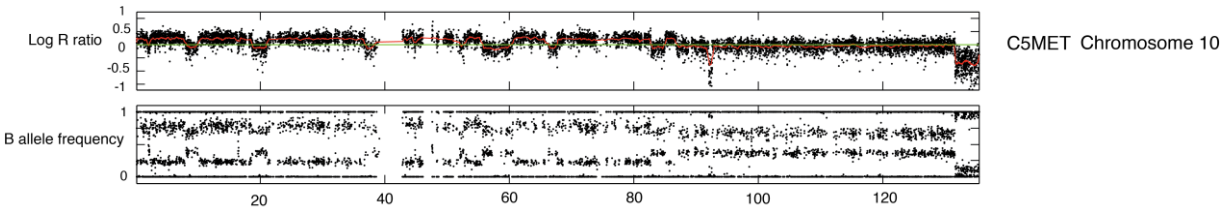


Figure 3.13 I Chromothripsis-like alterations in C5MET

For the purpose of this analysis chromothripsis-like alterations were defined as 10 ≥ alterations in copy number and/or LOH transitions involving ≤3 arms. Using mate pair sequencing Kloosterman et al demonstrated that chromothripsis can be identified in paired colorectal primaries and liver metastases (Kloosterman et al., 2011). The incidence of chromothripsis in colorectal cancer is currently undetermined due to the paucity of large-scale studies addressing this.

Interchromosomal and intrachromosomal translocations were screened for using BreakDancer. Searches focused on breakpoint regions, and were filtered stringently for quality and read depth. This systematic approach failed to identify candidate loci. Kataegis events were excluded in all the CCBM. One possibility for the inability to identify translocations using BreakDancer may result from reads being filtered out in the mapping process.

3.17 Comparison of colorectal brain metastases with other data sets.

To appreciate the degree of diversity in the CCBM, data from the CCBM were compared with published findings in primary colorectal tumours and primary tumours analysed by myself from the TCGA. In addition, liver metastases of colorectal origin were used for comparative analysis.

3.17.1 Comparison of mutation burdens with primary tumours

Although comparisons between sequencing studies can be affected by various factors: the type of sequencing technology utilised, sample size, software analysis programs and filtering criteria, meaningful insight and a broad overview can be gained through such comparisons. Bass et al reported a total number of mutations ranging from 10,445-19,931 (exonic mutations 30 to 147) from WGS of 9 colorectal cancer samples (Bass et al., 2011). TCGA reported exome sequencing data from a large cohort of primary colorectal cancers with a wide range in mutation frequencies, with the majority of tumours (84%) having a mutation frequency less than 8.24 per 10^6. They considered tumours with a mutation frequency greater than 12 per 10^6 to be hypermutated (Cancer Genome Atlas, 2012). Hypermutated tumours were confirmed to be either MSI positive or had somatic mutations in the *POLE* gene. In comparison, the mutation frequency of the CCBM in protein-coding regions ranged from 2.5 to 20 per 10^6. The magnitude of mutation burdens in the CCBM is comparable to that reported in primary colorectal tumours. C1MET is hypermutated with a mutation burden of 20 per 10^6 in the absence of *POLE* mutations. The increased mutation burden observed in C1MET may be related to environmental exposure.

3.17.2 Mutation burdens of CCLM

In total 3 pairs of liver metastases were sequenced using the Complete Genomics platform. The total number of mutations in the CCLM ranged from 8,383 to 15,671 [median 13,000] (Figure 3.14). The number of SNVs ranged from 7,937 to 15,106 [median 12,394].

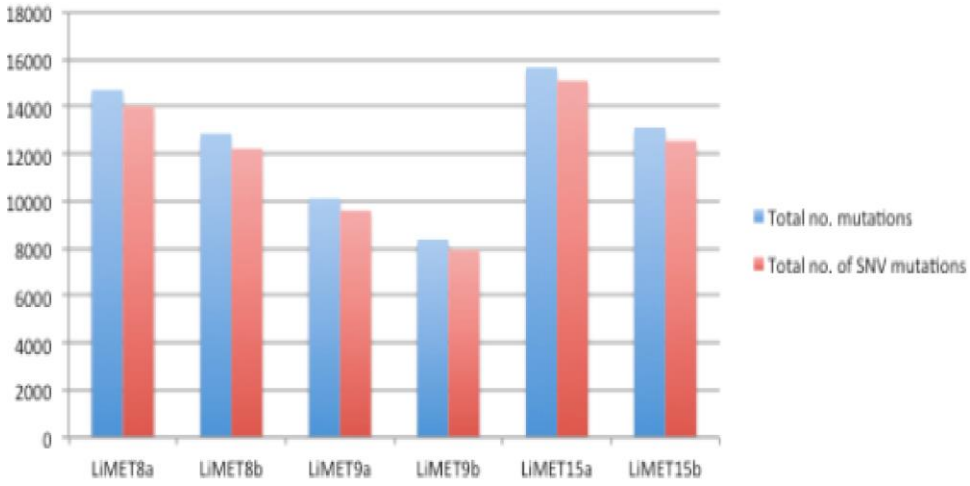


Figure 3.14 I Mutation burdens in CCLM in the whole genome. The mutation burden in the paired liver metastases were comparable. The suffix a represents the first metastasis in the pair.

The total number of small indels in the CCLM ranged from 565 to 705 [median 565]. The number of small indels reported in the CCBM and CCLM differed by 10 fold [CCBM range 2,955 to 6,873]. This may reflect the differing abilities in sequencing platforms and variant callers in identifying complex mutations. A similar observation has been reported by Kovac et al in an analysis of renal papillary cancers (Kovac et al., 2015). Using SNVs as a reliable marker for comparative analysis of mutation burdens, it is evident that the mutation burdens in the CCBM [range 10,492 to 64,460] and CCLM [range 7,397 to 15,106] were comparable.

Similarly to the CCBM the majority of somatic mutations in the CCLM were located in intergenic regions, with exonic mutations forming ~1% of the mutation burden. Within protein-coding regions a median of 97 mutations [range 64 to 118] were found to be potentially functional. Nonsynonymous mutations were the most common type of mutation, ranging from 49 to 105 [median 84] events per tumour. The number of nonsense mutations ranged from 2 to 8 mutations [median 3.5]. The number of small indels in protein-coding regions ranged from 3 to 8 [median 5]. The mutation burdens in the exome in the CCLM ranged from 2.6 to 5 per 10^6 (Figure 3.15), which is comparable to the mutation burdens in most CCBM.

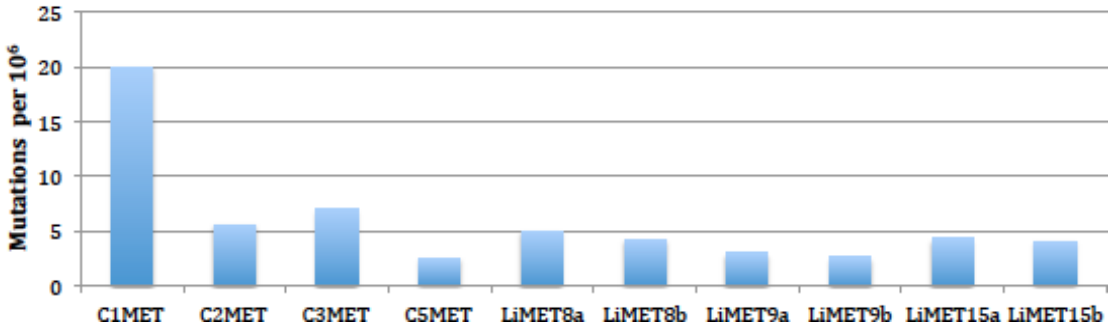


Figure 3.15 | Exonic mutation burdens in CCLM and CCBM in whole genome. Mutation burdens are expressed per megabase.

The mean nonsynonymous:synonymous mutation ratio was 3:1. Using the CCLM as a comparison, with the exception of C1MET which was hypermutated, the mutation burdens of the CCBM and CCLM were comparable. Moreover, comparison with published literature suggests that the metastases have a mutation frequency which is of comparable magnitude to primary tumours.

3.17.3 Comparison of mutation spectra

To characterise the heterogeneity in mutation spectra in the CCBM, the mutation spectra were evaluated in the CCLM and a cohort of primary colorectal tumours obtained from the TCGA cBioPortal (n=300) (Figure 3.16). In the CCLM C>T/G>A formed the predominant base substitution in the whole genome ranging from 28 to 36% of base substitutions [median 31%], whereas it constituted 42 to 59% of mutations in the exome [median 51%]. In the CCBM, C>T/G>A mutations in the exome ranged from 16 to 57% [median 51%]. Similarly, analysis of the mutation spectra in a large cohort of primary colorectal tumours from the TCGA which underwent exome sequencing confirmed that C>T/G>A substitutions predominated in colorectal cancer. These findings are consistent with previous reports in colorectal adenocarcinomas (Sjoblom et al., 2006).

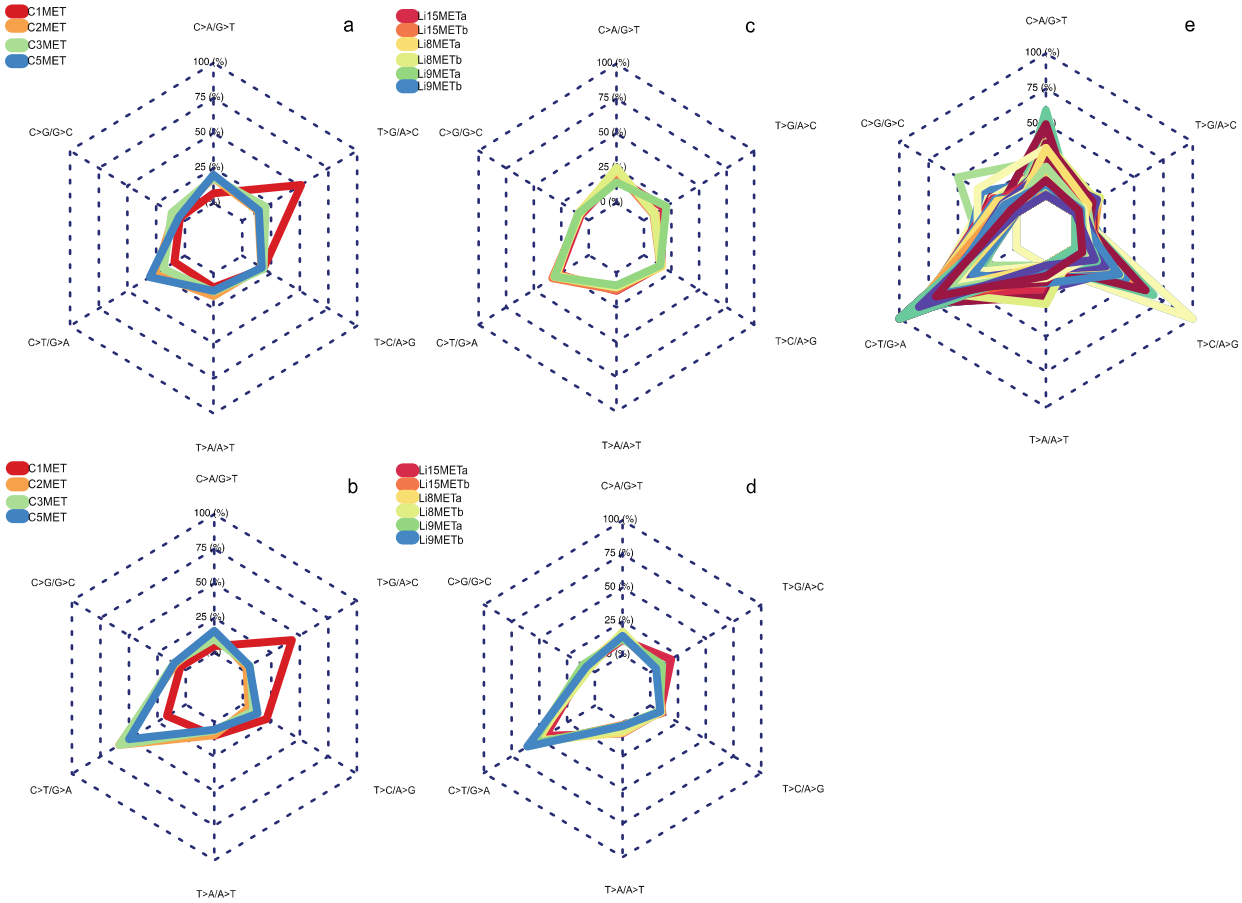


Figure 3.16 | Mutation spectra in primary and metastatic colorectal cancers a) Whole genome brain metastases b) Exome brain metastases c) Whole genome liver metastases d) Exome liver metastases e) TCGA primary colorectal tumours (exome sequencing). C>T/G>A mutations predominate in primary and metastatic colorectal cancer.

3.17.4 Evaluation of copy number alterations

The relative contribution of somatic copy number variation (a surrogate for chromosomal instability) in the colorectal metastases was compared to data from primary tumours obtained from TCGA. The cohort of metastases analysed was increased in size to a total of 5 CCBM and 25 CCLM. Of the 25 liver metastases, 16 were paired (resected from 8 individual patients). The colorectal metastases were genotyped on 4 different Illumina SNP array platforms. In brief, to compare the metastases SNPs directly in common amongst the 4 Illumina platforms were identified, yielding 121,484 probes. From the remaining SNPs on the least dense platform those with a minor allele frequency >10% were identified as they would be

more informative, herein termed MAF-probes (n= 111,297). Using direct SNPs in common amongst each platform and an equivalent number of MAF-probes randomly distributed across the genome a common SNP density of ~230,000 was achieved for all platforms. For comparison, TCGA primary tumours which were genotyped on Affymetrix arrays were downscaled to achieve a comparable SNP density. The brain metastases had a median CNA of 28 [range 20 to 31]. In comparison, the liver metastases had a median number of CNA of 26 [range 12 to 43]. The primary colorectal tumours were subdivided into MSI positive (n=72) and MSI negative tumours (n=502). The median number of CNA in MSI positive tumours was 4 [range 0 to 29], whereas 42% of tumours had no CNA and 95% of tumours had less than 13 CNA. In comparison, the MSI negative tumours had a median of 15 CNA [range 0 to 64]. Of these tumours 5% had no CNA and 80% of tumours had less than 25 CNA. There was a statistically significant difference between CNA in the CCBM and MSI positive primary tumours (Mann-Whitney U test, $p<0.01$). In comparison, there was a statistically significant difference between CNA in the CCLM and MSI positive and MSI negative primary tumours (Mann-Whitney U test, $p<0.01$ and $p<0.01$, respectively). There was no statistically significant difference between CNA in CCBM and MSI positive colorectal tumours though this may reflect the small sample size of CCBM.

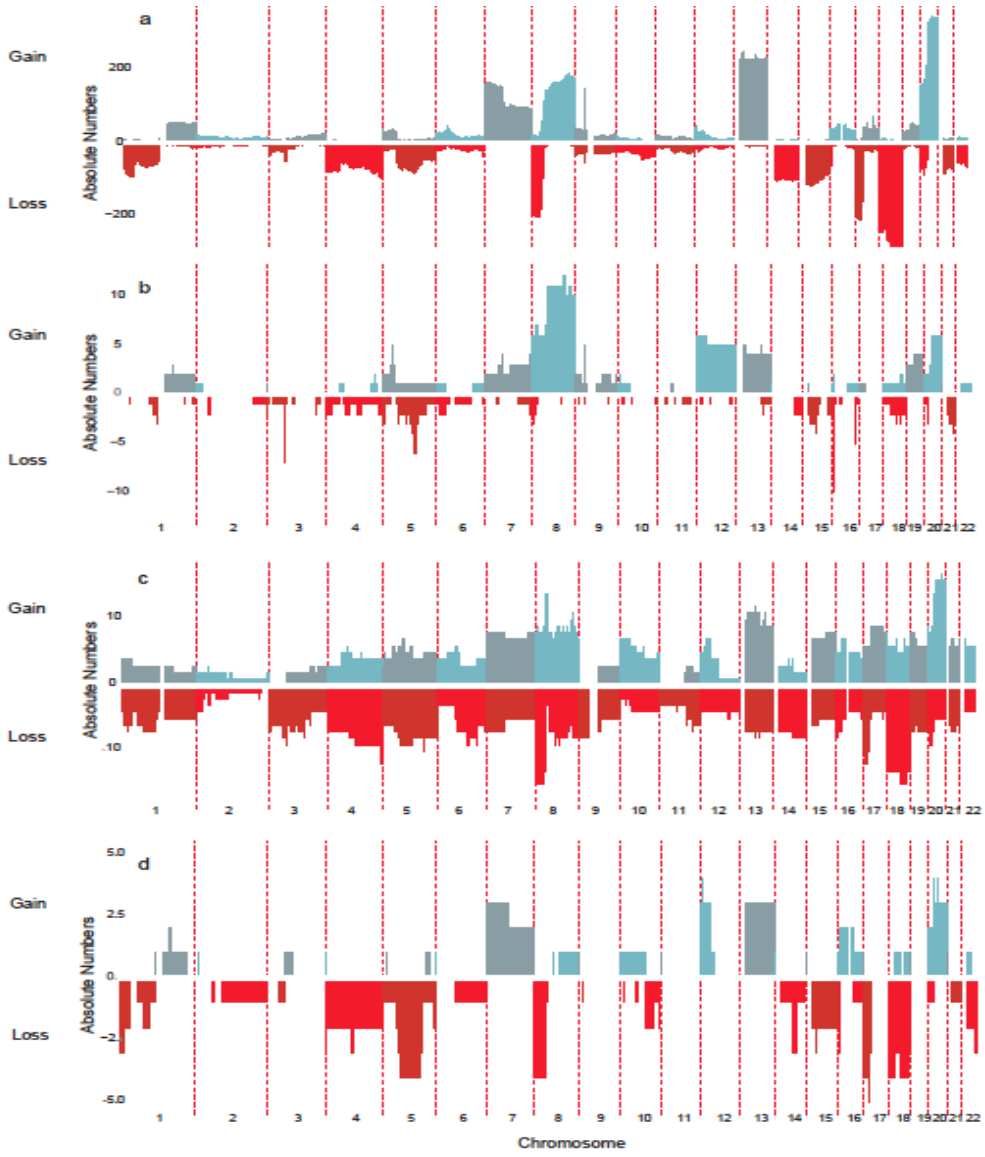


Figure 3.17 I Copy number variation in colorectal primary tumours and metastases Representation of deviation from average copy number in colorectal primary tumours and metastases. a) MSI negative primary tumours b) MSI positive primary tumours c) liver metastases d) brain metastases. Chromosomal distribution of gains and losses in primary tumours and metastases are depicted (Figure 3.14).

The colorectal metastases, which were MSI negative, were compared to MSI negative primary colorectal tumours (Figure 3.17). Whole arm losses were identified in each tumour (arbitrarily defined as >90% involvement of the arm). Notable chromosome-scale changes in the MSI negative primary tumours were defined as those with a frequency >10%. Deletions were detected in 17p (n=166), 18p (n=211), 14q (n=80),

15q (n=74) 18q (n=211) and 21q (n=49). Gains were found in 7p (n=129), 20p (n=107), 8q (n=101), 13q (n=144) and 20q (n=253). To identify significant alterations in the CCBM and CLM that were statistically significant when compared to primary MSI negative tumours, Fisher's exact test was undertaken, with analysis restricted to arms with a minimum of 40% chromosome-scale changes in the metastases due to the modest sample size. No significant differences were observed between the primary MSI negative tumours and the CCBM, though this may reflect the small number of brain metastases. Statistically significant deletions in the CCLM in comparison to the primary tumours were found in 4p, 5p, 7p, 1q, 4q, and 5q. Gains were observed in 8p, 17p, 18p, 19p, 15q, 17q and 19q.

3.17.5 Identification of recurrent focal copy number aberrations

Using the SNP arrays, recurrent focal copy number aberrations (<2MB) were identified in the CCLM. Table 3.6 summarises the recurrent focal copy number aberrations detected in the CCBM and CCLM. Recurrent focal deletions were detected in *RBFOX1* and *MACROD2* in the CCBM. A genome wide screen of the liver metastases uncovered focal deletions in *RBFOX1* in 3 tumours. No focal deletions were identified in *MACROD2*. However focal deletions were identified in the *FHIT* and *PARK2* in the CCLM.

Gene	Type of aberration	Number of samples with focal aberrations	Sample
RBFOX1	Deletion	3	C1MET C5MET Li7MET
MACROD2	Deletion	2	C1MET C2MET
FHIT	Deletion	3	Li7MET Li9METa Li9METb
PARK2	Deletion	3	Li6MET Li7MET Li16MET

Table 3.6 l Summary of focal copy number aberrations in CCBM and CCLM

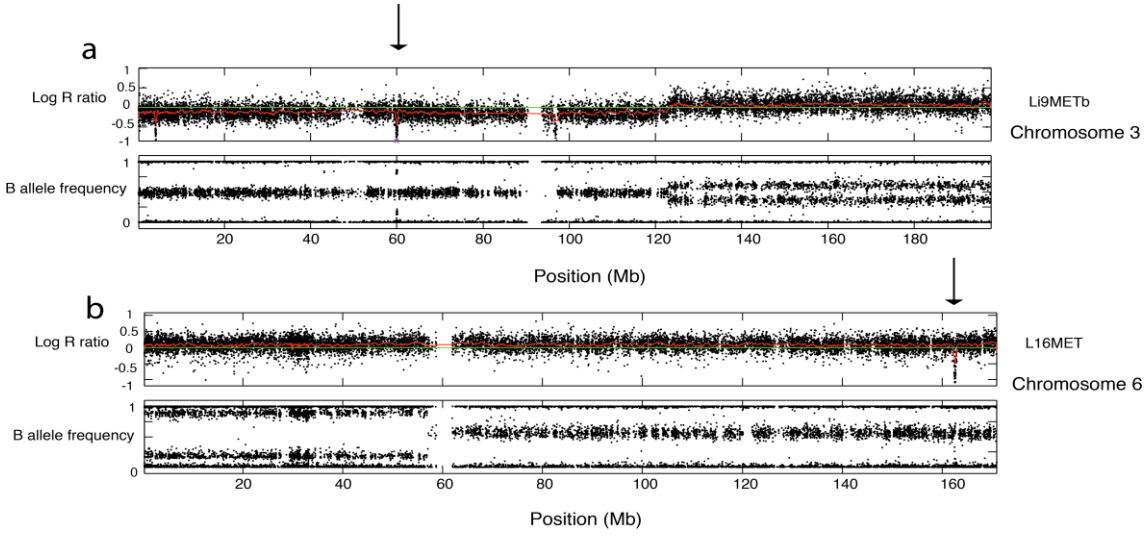


Figure 3.18 l Focal deletion detected in the CCLM. a) Focal deletion in *FHIT*. b) Focal deletion in *PARK2*.

Focal copy number aberrations in the aforementioned genes have been reported in primary colorectal tumours, which are known common fragile sites (Muzny et al., 2012). No focal copy number amplifications were detected in either the CCBM or CCLM.

3.17.6 Identification of chromosomal aberrations

Operational definitions of chromothripsis-like alterations were defined as 10 ≥ alterations in copy number and/or LOH transitions in ≤3 arms. Table 3.7 depicts the tumours that exhibited chromothripsis-like changes with 1 CCBM and 3 CLLM involved. The chromosomal aberrations identified would be in keeping with chromothripsis-like alterations as they involved a single arm.

Tumour	Chromosome arm involved
C5MET	10q
Li8METa	8q
Li12METb	10p
Li15METa	8q

Table 3.7 I Summary of chromothripsis-like changes in CCBM and CCLM

Interestingly, chromothripsis-like alterations in the paired liver metastases were detected in Li8METa and Li12METb. However these alterations were absent in the other liver metastases in the pair. In comparison, the chromothripsis-like alterations present in 8q in Li15METa was present in Li15METb. However there were new extended regions of LOH in the 8q arm of Li15METb in keeping with evolution of the tumour (Figure 3.19).

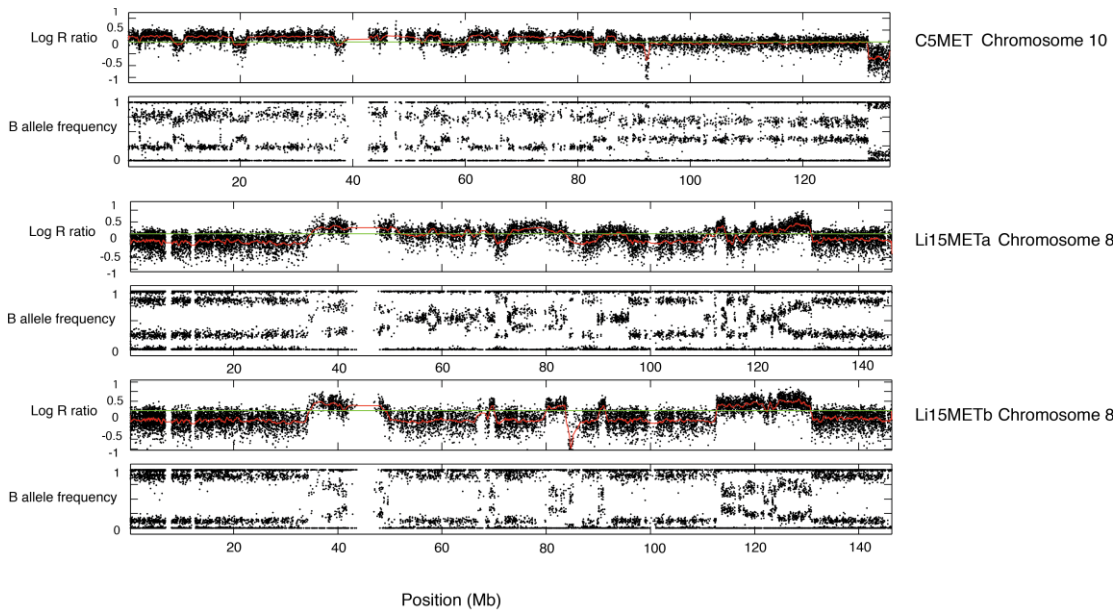


Figure 3.19 | Chromothripsis-like alterations in CCBM and CCLM

3.18 Technical validation of sequencing data

To validate the somatic variants identified through the sequencing analysis pipeline the sequenced CCBM and CCLM were validated using targeted sequencing with the Ion Torrent custom cancer gene panel, which is composed of 82 genes (Table 3.8). In total 15 somatic variants were screened in the CCBM. Of these, 2 variants were non-informative due to insufficient coverage, though the remaining 13 variants were validated. In comparison, in the liver metastases from 10 candidate mutations screened, 8 were technically validated. A complex frame shift deletion in *APC* was not validated.

Tumour	Gene	Mutation	Type of mutation	Validated
C1MET	*ATM*	exon18:c2705 A>GpK902R	Nonsynonymous	Yes
C1MET	*KRAS*	exon2:c.35 G>C:p.G12A	Nonsynonymous	Yes
C1MET	*MTOR*	exon43:c5972 C>TpA1991V	Nonsynonymous	Yes
C1MET	*SMO*	exon1:c286 T>G pS96A	Nonsynonymous	Yes
C1MET	*TP53*	exon5:c.524_535del:p.175_179del, GGTGGGGGCAGC	In frame deletion	Yes
C1MET	*MLH1*	exon9:c709 A>CpT237P	Nonsynonymous	Yes
C2MET	*APC*	exon8:c.646 C>T:p.R216X	Nonsense	Yes
C2MET	*KRAS*	exon2:c.38 C>T:p.G13D	Nonsynonymous	Yes
C2MET	*N4BP2L2*	exon7:c.1078delA:p.I360fs	Frameshift deletion	Insufficient coverage
C3MET	*APC*	exon17:c.4216 C>T:p.Q1406X	Nonsense	Yes
C3MET	*KRAS*	exon4:c.351 A>T:p.K117N	Nonsynonymous	Yes
C3MET	*TP53*	exon5:c.527 G>T:p.C176F	Nonsynonymous	Yes
C3MET	*PTEN*	exon4:c.223delC:p.H75fs	Frameshift deletion	Insufficient coverage
C5MET	*ALK*	exon11:c.G1923T:p.E641D	Nonsynonymous	Yes
C5MET	*TP53*	exon6:c670 G>T:p.E224X	Nonsense	Yes
Li8METa	*TP53*	exon13:c.128 G>A:pR43H	Nonsynonymous	Yes
Li8METb	*TP53*	exon6:c.38 C>T:p.G13D	Nonsynonymous	Yes
Li8METa	*APC*	exon14:c.3809_3834del:p.1270_1278del	Frameshift deletion	No
Li8METb	*APC*	exon14:c.3809_3834del:p.1270_1278del	Frameshift deletion	No
Li9METa	*APC*	exon16:c3925 G>T:pE1309X	Nonsense	Yes
Li9METb	*APC*	exon16:c3925 G>T:pE1309X	Nonsense	Yes
Li9METa	*TP53*	exon6:c.628 C>T:pR210X	Nonsense	Yes
Li9METb	*TP53*	exon6:c.628: C>TpR210X	Nonsense	Yes
Li15METa	*TP53*	exon2:c.245: A>GpH82R	Nonsynonymous	Yes
Li15METb	*TP53*	exon2:c.245 A>G:pH82R	Nonsynonymous	Yes

Table 3.8 I Summary of technical validation of CCBM and CCLM

3.19 Discussion

3.19.1 Mutation burden and spectra/signatures

The mutation burdens in the CCBM and CCLM were found to be of similar a magnitude to that of primary colorectal tumours. The brain metastasis C1MET was found to be hypermutated in comparison to the other colorectal metastases sequenced with a mutation frequency of 20 per 10^6 in the absence of MSI or *POLE* mutations. C1MET exhibited a mutation signature with a predominance in T>G/A>C base substitutions in the context of a C[T]T motif. Alexandrov et al reported the mutation signatures of a large cohort of tumours in a pan-cancer analysis using a hierarchical clustering algorithm (Alexandrov et al., 2013). They reported that T>G mutations occurring in the context of C[T]T motif [signature 17]. This signature was detected in liver, gastric, oesophageal carcinomas and B cell lymphoma. More recently, Katainen et al reported that this signature in colorectal cancer was associated with CTCF/cohesin binding sites. The increased mutation burden and unusual signature may have been driven in response to exposure to environmental agents. Interestingly, despite the increased mutation burdens observed in C1MET, it had the least number of CNA.

3.19.2 Clonal architecture of the colorectal cancer brain metastasis

The clonal architectures of the CCBM varied in complexity ranging from 7-35 clonal groups. Although no drivers of CCBM were identified it was evident that drivers of primary colorectal cancer retained in the brain metastases. In addition to the classical drivers of colorectal cancer *APC, TP53* and *KRAS*, more recently identified drivers of colorectal cancer were detected including *TCF7L2, AR1D1A* and *PIK3R1.* These drivers of primary colorectal cancer were generally found in major clones in the CBBM though this was not universal.

3.19.3 Profiling copy number alterations in colorectal metastases

Historically a variety of techniques have been utilised to elucidate the copy number profile in colorectal cancers including, FISH and genomic hybridisation techniques (Garcia et al., 2003, De Angelis et al., 1999, Jones et al., 2005). In recent years analysis of copy number in cancers has focused on using SNP arrays and directly inferring copy number information from sequencing data.

The number of CNA (defined as the number of deviations from the average copy number) was determined in the CCBM and CCLM and primary colorectal tumours. The median CNA in CCBM and CCLM was 28 and 24 in the brain and liver metastases whereas this was 4 and 15 in primary colorectal MSI positive and MSI negative tumours respectively. The difference in CNA between MSI positive tumours and CCBM and CCLM was statistically significant. However, there was only a statistically significant difference between MSI negative primary colorectal tumours and CCLM. A comparison of CCLM and primary tumours demonstrated deletions in 4p, 5p, 7p, 1q, 4q, and 5q. Gains were observed in 8p, 17p, 18p, 19p, 15q, 17q and 19q. Munoz-Bellvis *et al* reported findings from a comparison of colorectal primary tumours and paired liver metastases. They reported that aberrations were often shared amongst the primary and metastatic tumours. Losses were reported in 1p, 8p, 17p and 18q in both primary and metastatic tumours whereas gains were observed in 8q, 13q and 20q. New genetic alterations acquired in the liver metastases included losses in chromosome 4q, 5q and 10q and gains of 2p, 5p, 6p, 7q and 11p. Gutenberg et al reported their findings of comparative genomic hybridisation of paired primary and colorectal brain metastases (Gutenberg et al., 2010). They found an increase in the number of gains in 8q, 12p, 12q, and 20q with losses in 5q in the brain metastases.

3.19.4 Screen for chromosomal aberrations

Chromothripsis-like alterations restricted to a single arm were identified in 4 metastases, 1 CCBM and 3 CCLM. The arms involved were 8q, 10p and 10q. Chromothripsis has been reported in primary and metastatic colorectal cancer using mate pair sequencing (Kloosterman et al., 2011). The incidence of chromothripsis-like alterations in colorectal cancer is largely unknown.

In summary, CCBM and CCLM often share characteristics of unpaired primary tumours including mutation burden and spectra. However, copy number analysis reported an increase in CNA in the metastases when compared to unmatched primary tumours. Furthermore, chromothripsis-like alterations were identified in 1 CCBM and 3 CCLM. No kataegis events were detected in the CCBM.

Chapter 4

Characterisation of a paired colorectal primary tumour and brain metastasis

4.1 Introduction

Advances in nucleic acid sequencing technology have revolutionised the analysis of cancer genomes (Meyerson et al., 2010). The advances include the identification of new driver mutations, insights into cancer heterogeneity (Zhang et al., 2014, Gerlinger et al., 2012) and the classification of tumours according to their mutation signatures (Lawrence et al., 2013). This has been coupled with the ability to use targeted therapies according to cancer genotype (Mok et al., 2009, Chapman et al., 2011). Despite these efforts, WGS of FFPE tumours has remained challenging due to the integrity of DNA and ability to generate reliable data (Ulahannan et al., 2013).

An increased understanding of the degree of heterogeneity that exists between primary tumours and brain metastases may help direct future therapeutic options for patients. In chapter 3, I reported the genomic landscape of 4 unpaired CCBM. DNA from 3 FFPE primary colorectal tumours that matched with the sequenced CCBM was extracted. One of these samples was of sufficient quality to undertake WGS in a pilot study in collaboration with Illumina. The primary tumour sequenced P1, corresponds to the paired brain metastasis C1MET. Interestingly, as reported in chapter 3 the brain metastasis C1MET had unusual characteristics in comparison to the other CCBM. C1MET was hypermutated and exhibited an uncommon mutation signature. Herein, I report a comprehensive characterisation of the genomes of C1MET and its matched primary tumour P1.

4.2 Methods

4.2.1 WGS of a matched CCBM and primary tumour

Genomic DNA from C1MET and P1 was extracted and sequenced from fresh frozen and FFPE tumour respectively, with the aim to achieve a median depth of coverage of 60X, 30X and 30X in the brain metastasis, paired primary tumour and constitutional DNA respectively. Variant calling was undertaken with BAM files of the primary tumour, metastasis and normal in concert (section 2.11). To account for the differing median depths of coverage between the metastasis and primary, the C1MET BAM file was stochastically downsampled to achieve a comparable coverage (section 2.11.5). Analysis of sequencing data is outlined in (section 2.14).

4.3 Comparison Datasets

4.3.1 Analysis of multi-region sampling from primary colorectal adenocarcinomas

WGS data (Illumina HiSeq 2000) from multiple biopsies, spatially separated, from 2 primary colorectal adenocarcinomas from chemotherapy naive patients were obtained from Dr Simon Leedham. The colorectal adenocarcinomas, MR1 and MR2, had 6 and 7 biopsies taken respectively. Sequencing data were analysed by myself (section 2.13). The sequencing data provided information regarding the degree of heterogeneity that existed within primary colorectal adenocarcinomas. This served as a comparison when appreciating the context of the degree of heterogeneity that existed between C1MET and P1.

4.4 Case Description

A 49 year old male presented with a moderately differentiated colorectal adenocarcinoma (T4N2M0) located at the splenic flexure. The patient underwent a left hemicolectomy and adjuvant chemotherapy with FOLFOX (folinic acid, 5-fluorouracil and oxaliplatin). Histopathological assessment of the primary tumour revealed that the tumour had infiltrated through the muscularis propria and reached the serosal surface, with accompanying perforation. Extramural lymphovascular invasion was present, with 5/10 examined lymph nodes containing metastatic tumour. The patient subsequently relapsed and received several lines of treatment including exposure to irinotecan, and bevacizumab. After a period of 2 years had elapsed following the primary presentation, the patient developed Jacksonian seizures. An MRI scan demonstrated a left frontal brain metastasis that was resected. Immunostaining of the resected metastasis was in keeping with a colorectal origin.

Results

4.5 Mutation burdens

4.5.1 Mutation burdens-assessment of original dataset

WGS was undertaken to achieve a median depth of coverage of 60X, 30X and 30X in C1MET, P1 and constitutional DNA. Using these depths of coverage the mutation burdens were analysed. The total number of mutations in the metastasis and primary were 69,015 and 18,193 respectively of which 63,090 and 14,687 were SNVs respectively. The number of small indels in C1MET and P1 were 5805 and 2922 respectively. Mutations located within exonic regions in C1MET and P1 consisted of 0.8% of the overall mutation burden in both tumours. The bulk of mutations across

the genome were located in intergenic locations constituting 65% (P1MET) and 62% (C1MET) of the overall mutation burdens.

In the exome, the mutation burdens were 20 per 10^6 in C1MET and 5.1 per 10^6 in P1. The number of synonymous mutations in C1MET and P1 were 114 and 36 respectively. Evaluation of protein-coding regions confirmed that there were 450 non-synonymous, 15 nonsense mutations and 15 indels in the metastasis, compared with 85, 8 and 13 respectively in the primary tumour. Notably, 365 SNV mutations were exclusively found in the metastasis, whereas only one mutation, *GNAS* P212L, was exclusively found in the primary. Genome-wide, 14,081 somatic SNVs were shared by the primary and the metastasis, constituting 96% of the SNV mutations in the primary and 22% of SNV mutations in C1MET. Genome wide, there were 606 SNVs exclusively found within the primary tumour. The transition/transversion ratio in C1MET and P1 was 0.4 and 0.7 respectively.

4.5.2 Mutation burden-assessment following downsampling

To account for the differing median depths of sequencing coverage, reads in the metastasis BAM file were downsampled stochastically to achieve a comparable coverage to the primary. Reassessment of the mutation burdens resulted in a small reduction in the total number of SNVs in the metastasis and primary 59,210 (93%) and 14,430 (98%) respectively (Figure 4.1). The number of SNVs shared by the primary and the metastasis genome wide still constituted ~96% of the SNV mutations in the primary. The number of small indels reduced in C1MET (n=4465, 77%) and P1 (n=2614, 89%).

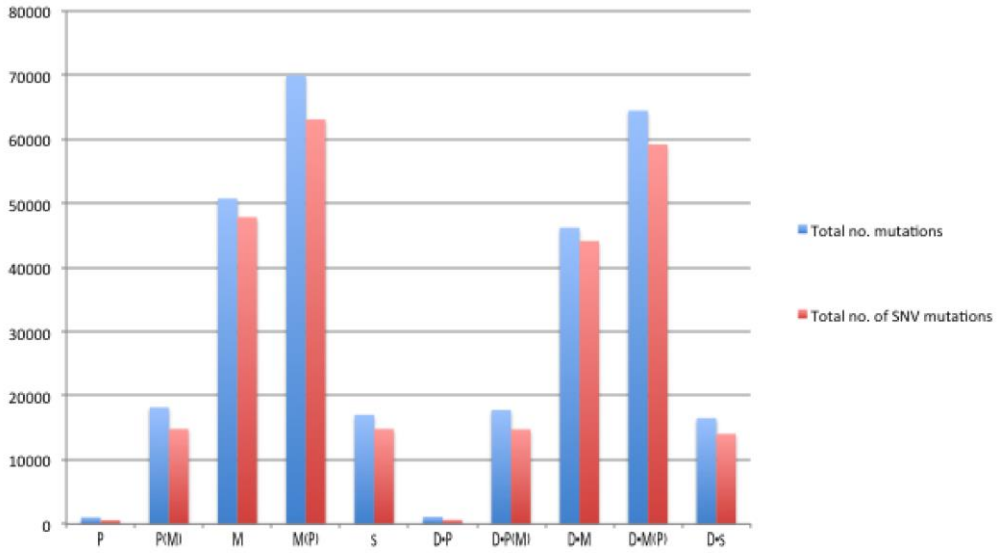


Figure 4.1 Mutation burdens in C1MET and P1 in original and downsampled analysis (whole genome) I P-mutations exclusive to primary, P(M)-all mutations in primary, M-mutations exclusive to brain metastasis, M(P)-all mutations in the brain metastasis, S-mutations shared by the primary and metastasis. The prefix D in the aforementioned categories represents mutation burdens after downsampling analysis. The colorectal brain metastasis has a greater mutation burden than the primary tumour in the original and downsampled analysis.

Genes in protein-coding regions were evaluated and reassuringly all mutations shared between the tumours were retained [85 nonsynonymous, 8 nonsense and 5 indels]. A small number of mutations (8%) in the metastasis were now unidentified. In comparison, mutations in the primary tumour were largely retained (Figure 4.2).

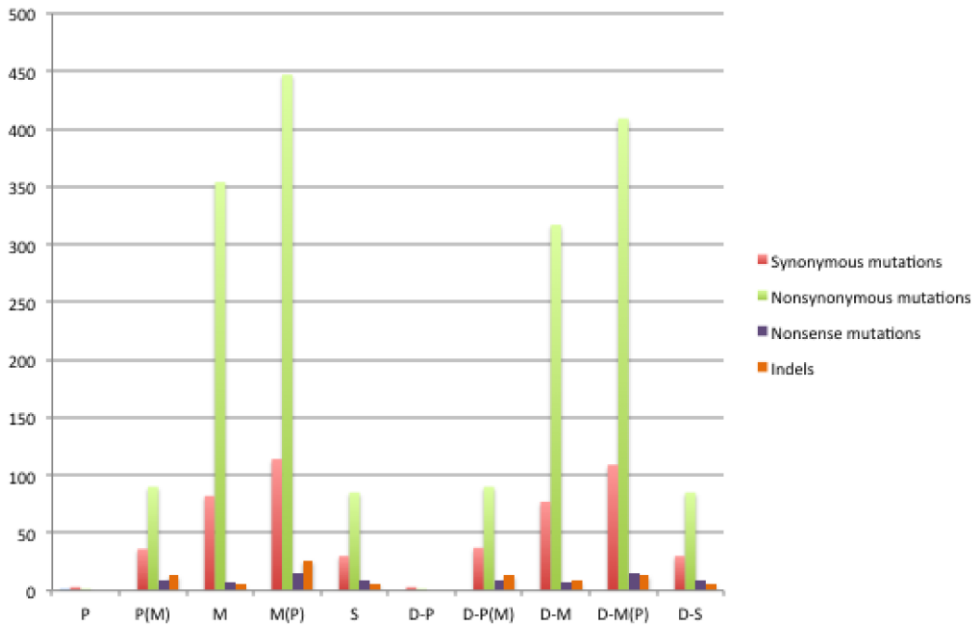


Figure 4.2 Distribution of mutations in the exome in C1MET and P1 in original and downsampled analysis I P-mutations exclusive to primary, P(M)-all mutations in primary, M-mutations exclusive to brain metastasis, M(P)-all mutations in the brain metastasis, S-mutations shared by the primary and metastasis. The prefix D in the aforementioned categories represents mutation burdens after downsampling analysis. There is a small reduction in the number of mutations in the metastasis after downsampling (8%), though mutations in the primary are largely retained.

Collectively, analysis of mutation burdens of C1MET and P1 demonstrate that the bulk of mutations in the primary tumour are found within the metastasis. Furthermore, it is apparent there has been limited divergence in the primary since the origin of the metastasis, whereas the metastasis has continued to evolve. From these findings one could postulate that the primary tumour was resected soon after the metastatic founder cell had detached.

Hypermutated tumours in colorectal cancer have been associated with MSI and somatic mutations in *POLE (Cancer Genome Atlas, 2012)*. As described in detail in chapter 3, C1MET was confirmed to be MSI negative directly from sequencing data using MSI sensor. Moreover, somatic and germline mutations in *POLE* were absent. The paired primary, P1, shared these characteristics. Conceivably, the increased mutation burden observed in C1MET may have resulted from environmental

exposure to therapeutic agents. In support of this, WGS of a matched primary and metastatic acral melanoma in a chemotherapy naive patient demonstrated comparable mutation frequencies between the primary and metastasis which were sampled synchronously (Turajlic et al., 2012a). They reported that the bulk of mutations in the primary were found in the metastasis whereas the metastasis had only 2 exclusive *de novo* mutations.

4.6 Evaluation of variant allele frequency

Given that a small number of SNV mutations (n=606) were found exclusively in the primary, variant allele frequencies (VAF) were analysed to determine whether these mutations were subclonal in origin. Furthermore, this provided insight into the evolution of the primary tumour following detachment of the metastatic founder cell. VAFs were determined for the mutations in the primary and metastasis (Table 4.1). For the purpose of this analysis, WGS data from the downsampled analysis was utilised for comparable coverage in the primary and the metastasis.

Tumour	Number of SNV	VAF		
		Range	Median	Mean
Mutations exclusive to P1	606	7-58	20	22
All mutations in P1	14701	5-88	21	23
Mutations exclusive to C1MET	40856	5-100	24	28
All mutations in C1MET	53977	5-100	28	32

Table 4.1 I Variant allele frequencies in the primary (P1) and brain metastasis (C1MET).

The median VAF of all mutations in C1MET and those exclusive to C1MET was 24 [range 5 to 100] and 28 [range 5 to 100] respectively. In comparison, the median VAF of mutations exclusive to the primary was 20 [range 7 to 58], whereas it was 21 [range 5 to 88] when considering all mutations in the primary. These findings suggest that the SNVs that were exclusive to the primary are not subclonal in origin. Moreover, following the detachment of the metastatic founder cell, the primary has continued to evolve independently to the metastasis.

4.7 Characterisation of mutation signatures

Genome wide there was an increased proportion of T>G mutations in the metastasis (51%). Notably the proportion of T>G mutations (15%) in the primary tumour was not increased, with C>T substitutions predominating (28%). C>T substitutions typically predominate in colorectal cancer (Sjoblom et al., 2006). Using somatic signature somatic base substitutions were analysed in sequence context taking into account the nucleotide flanking the 5' and 3' and of each mutated base. This yields a total of 96 possible sequence motifs. Through this process 2 signatures were identified S1 and S2. T>G mutations in C1MET predominantly occurred in the context of a C[T]T motif, as depicted in signature S1 (Figure 4.3). Alexandrov *et al* previously identified this as signature [17] in a large pan-caner analysis (Alexandrov et al., 2013). They reported the signature in 1.8% of all signatures detected in oesophageal, gastric and hepatocellular carcinomas and B cell lymphomas. Further insight into the characteristics of signature 17 has been obtained through analysis of the mutation footprint of a large cohort of colorectal cancers using WGS and ChIP-seq technology. Analysis of non-coding regions demonstrated that tumours with signature 17 had a predominance of A>C/T>G and A>G/T>C mutations within CTCF/cohesion binding sites. Cohesin regulates the separation of sister chromatids in DNA replication and co-localises with several transcription factors including CTCF (CCCTC-binding factor)(Parelho et al., 2008, Rubio et al., 2008, Kagey et al., 2010).

The underlying aetiology for signature 17 is unknown, though oxidative stress has been suggested as possible causative agent . (Katainen et al., 2015).

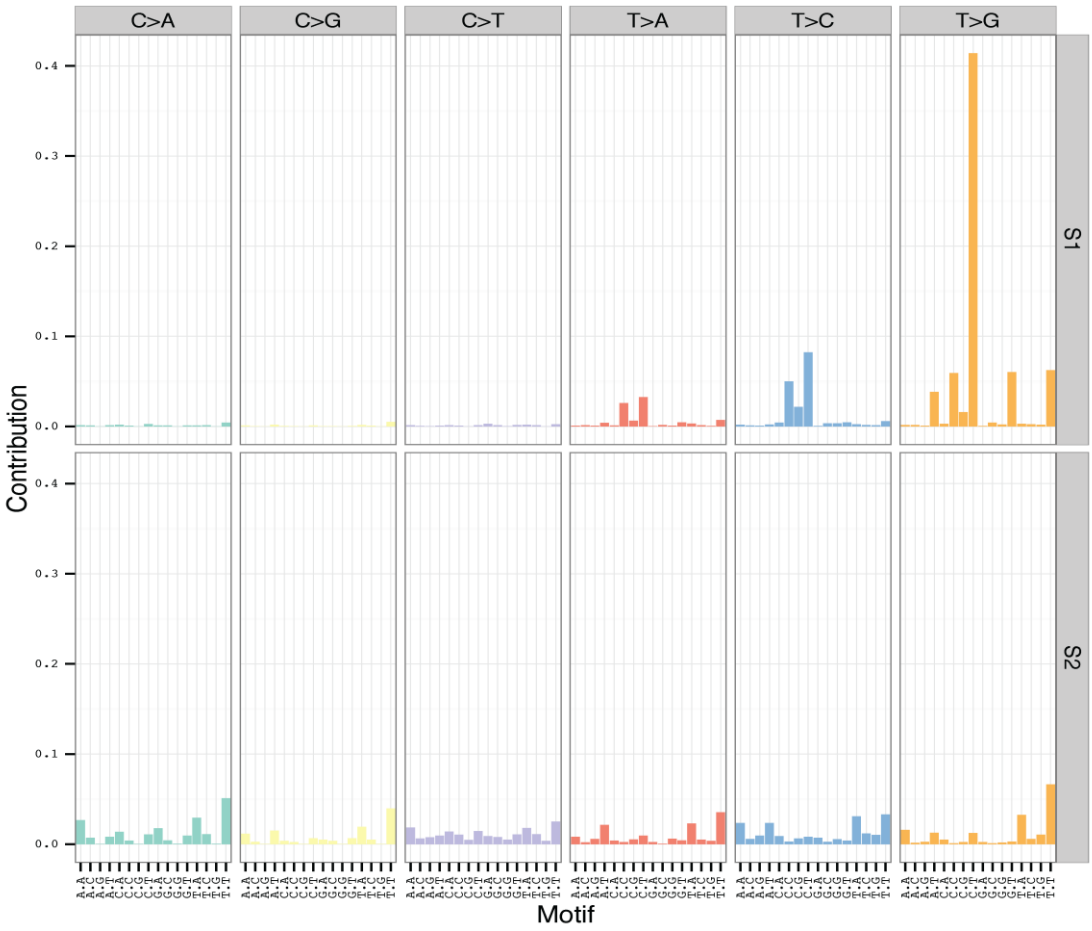


Figure 4.3 I Characterisation of mutation signatures in C1MET and P1. T>G mutations in C1MET occur in the context of a C[T]T motif.

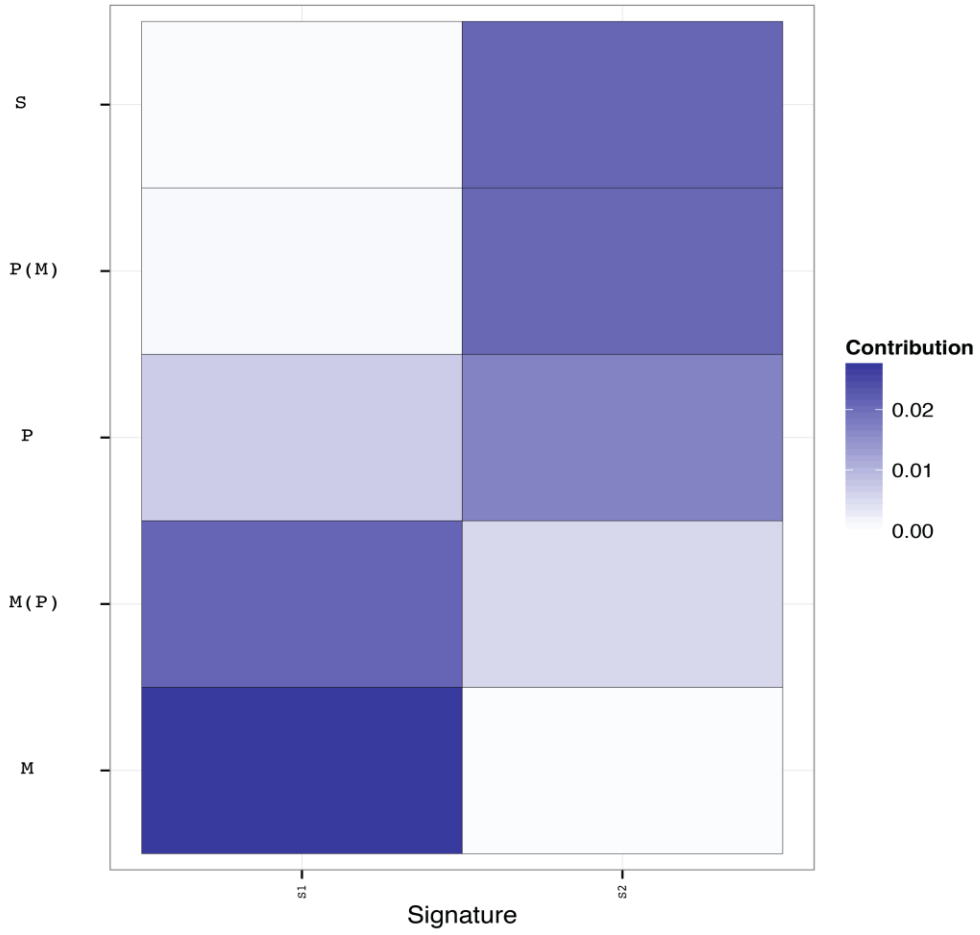


Figure 4.4 I Relative contribution of mutation signatures in C1MET and P1. P-mutations exclusive to primary, P(M)-all mutations in primary, M-mutations exclusive to brain metastasis, M(P)-all mutations in the brain metastasis, S-mutations shared by the primary and metastasis. The S1 signature where T>G mutations occur in a C[T]T motif predominates in mutations found within the metastasis.

Signature S2 resembles the signature identified in the majority of the sequenced brain and liver metastases of colorectal origin that were described in chapter 3. This signature does not closely resemble any of the signatures generated by Alexandrov et al. To analyse the signatures in greater depth, mutations were categorised into the following groups: all mutations in the primary, mutations exclusive to the primary, all mutations in the metastasis, mutations exclusive to the metastasis and mutations shared amongst the primary tumour and the metastasis. Subgroup analysis of somatic mutations in the metastasis and primary confirmed that signature S1

predominated in mutations that were found in the metastasis (Figure 4.4). The change in mutation signature from the primary tumour to the brain metastasis may represent a response to environmental exposure.

4.8 Evaluation of copy number alterations

OncoSNP-SEQ was utilised to infer copy number information directly from C1MET and P1 using NGS data (Figure 4.5). For the purpose of this analysis deviations from the mean copy number of the tumour were considered as regions of CNA. Analysis of CNA in the primary and C1MET confirmed that 76% of regions of CNA in the primary tumour were shared with the metastasis. In contrast, only 10% of the regions of CNA in the brain metastasis overlapped with the primary tumour (Figure 4.5).

Large-scale shared regions of CNA included deletions in 8p and 17p. Deletions in 17p co-localised with the tumour suppressor gene *TP53*. Changes observed selectively in the metastasis included: deletions in 4q, 5q, and 10q, and gains of chromosomes 7p and 8q. The proportion of the genome in the primary tumour that deviated from the average copy number was 3%, whereas it was 18% in the metastasis. A genome wide screen for focal gains/losses (<2MB) in C1MET and P1 to identify putative tumour suppressor genes or oncogenes was negative.

Ding et al undertook WGS of a basal like breast primary tumour and a paired brain metastasis with a time interval of 8 months between the tumours. The patient had received neoadjuvant chemotherapy and radiotherapy upon presentation. The primary tumour was biopsied prior to the initiation of treatment. They reported that 96% of CNA in the primary were found in the brain metastasis, whereas 80% of the CNA in the metastasis overlapped the primary tumour. In contrast my findings demonstrate evolution of CNA in the metastasis as only 10% of CNA regions in the metastasis overlapped the primary tumour.

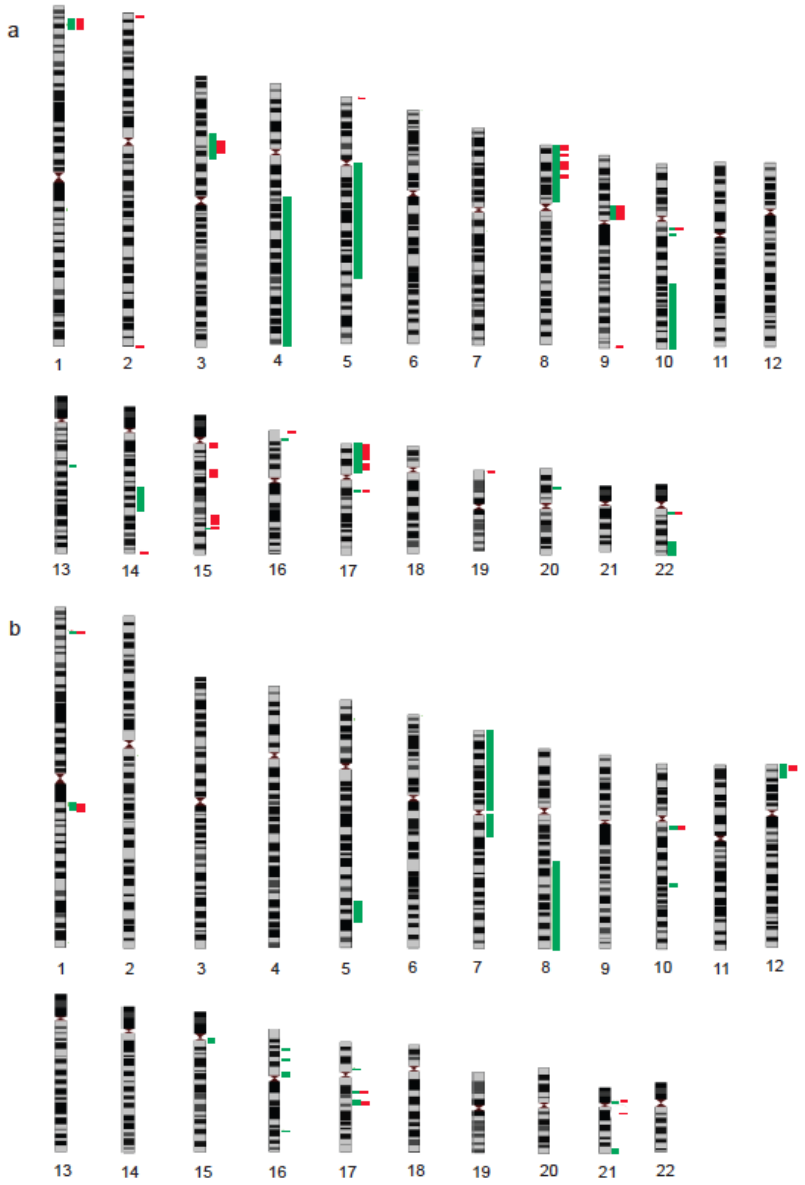


Figure 4.5 I Regions of gains and deletions in C1MET and P1 Deviations from the average copy number of the tumour are depicted. a) regions of deletions b) regions of increased copy number. C1MET and P1 are represented in green and red respectively.

4.9 Screen for drivers of primary colorectal cancer

The degree of concordance between drivers of primary colorectal cancer was evaluated in C1MET and P1. Candidate drivers of primary colorectal cancer were identified by running IntOGen and Mutsig (FDR <0.05) using somatic mutation data from a large cohort of primary tumours obtained from the TCGA cBioPortal.

Following this, putative mutations were filtered to exclude potentially benign or moderate functional effects (SIFT <0.2 PolyPhen 2 >0.8). Mutations in the 'classical' drivers of colorectal cancer were identified in the oncogene *KRAS* and the tumour suppressor gene *TP53* in both the primary and metastasis. The *KRAS* mutation (G12A) is well documented in the literature resulting in constitutive activation of the RAS GTPase (Ellis and Clark, 2000, Smith et al., 2010). The *TP53* mutation (del p175-179 GGTGGGGGCAGC) produced and in frame deletion which is not present in the *TP53* mutation database (Petitjean et al., 2007). No driver mutations were found exclusively in the primary, whereas nonsynonymous mutations in *PIK3R1* and *TCFL2* were present exclusively in the metastasis (Table 4.2). The mutations in *PIK3R1* and *TCFL2* encompassed regions of LOH. Importantly, the aforementioned driver mutations were identified in the metastasis following downsampling analysis.

4.10 Characterisation of clonal architecture

Pyclone was used to reconstruct the clonal architecture of the primary tumour and the metastasis. C1MET demonstrated increased clonal complexity in comparison to P1. There were 35 clonal populations in the metastasis, whereas the primary was clonally simple with 4 clonal groups. To appreciate the hierarchal position of driver mutations in the clonal structure of the tumours, driver mutations were plotted alongside randomly selected genes (Figures 4.6 and 4.7).

Gene	Present in metastasis	Present in primary	Location of mutation	Type of mutation	LOH at locus of mutation	Copy Number at locus of mutation	PolyPhen 2 Score	SIFT Score	Mutation frequency in gene TCGA cBioPortal %	Mutation frequency in gene at that loci TCGA cBioPortal
PIK3R1	Yes	No	exon9:c.1040 T>G:p.L347R	Nonsynonymous	Yes	1	1	0	6% (12/212)	0
TCF7L2	Yes	No	exon13:c.1337 A>Cp.K446T	Nonsynonymous	Yes	1	0.96	0	12% (29/212)	0
TP53	Yes	Yes	exon5:c.524_535del:p.175_179del,	In-frame deletion	Yes	1	--------	--------	52% (112/212)	0
KRAS	Yes	Yes	exon2:c.35 G>C:p.G12A	Nonsynonymous	No	3	0.95	0	42% (89/212)	1

Table 4.2 I Summary of colorectal driver mutations identified in P1 and C1MET

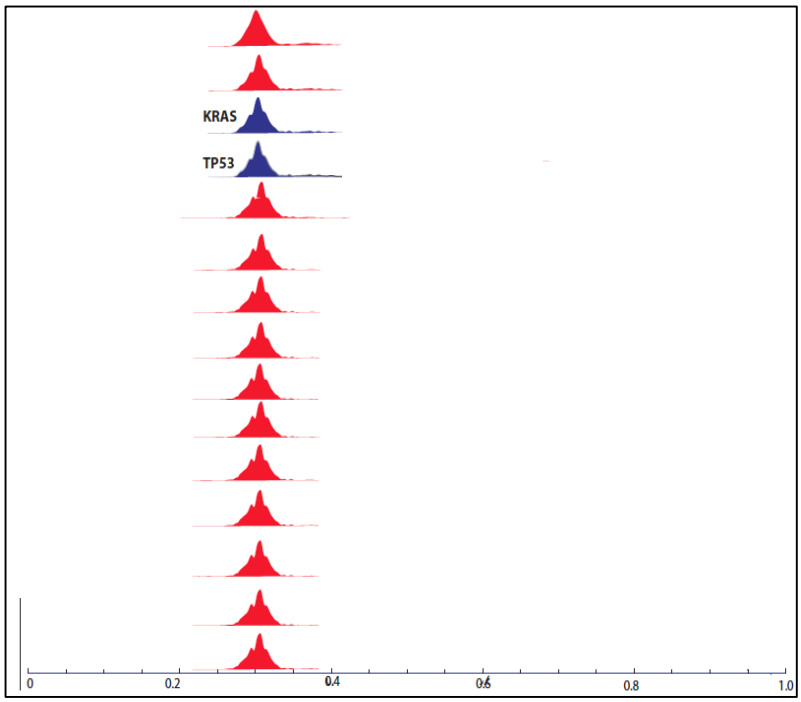


Figure 4.6 I Clonal architecture represented in P1. The clonal architecture in the primary is clonally 'simple'. *TP53* and *KRAS* were identified in P1 as drivers of primary colorectal cancer. Somatic mutations in randomly selected genes are depicted (driver mutations are highlighed in blue).

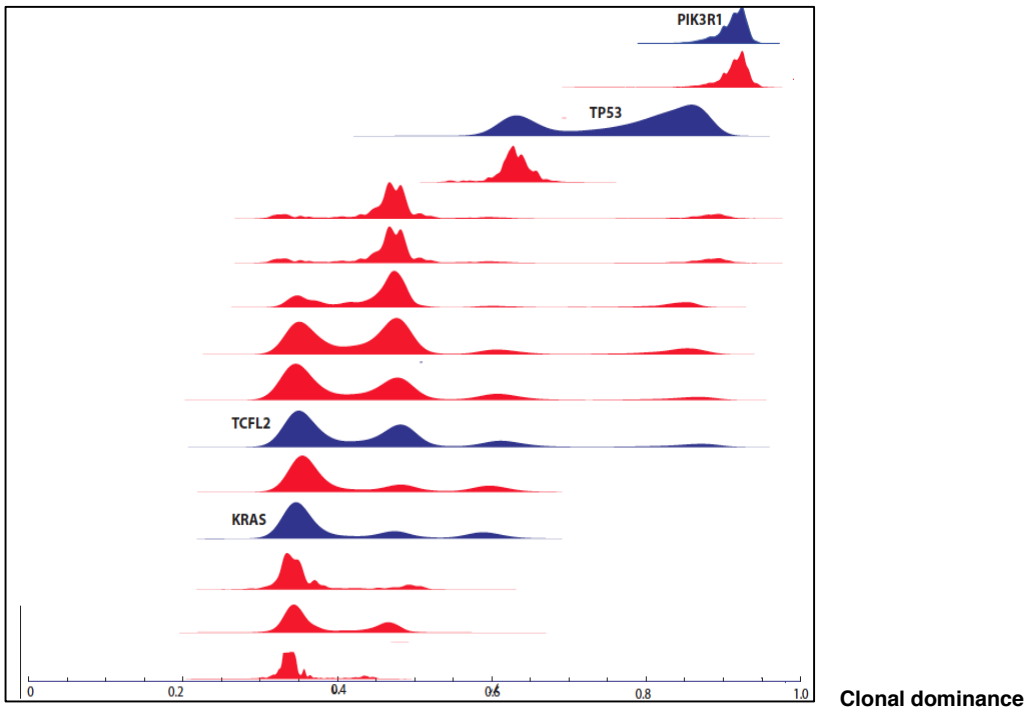


Figure 4.7 I Clonal architecture represented in C1MET. Driver mutations associated with primary colorectal cancer are depicted in blue in the clonal architecture of the brain metastasis C1MET. *PIK3R1* and *TP53* are located in major clones, whereas *KRAS* and *TCF7L2* are located in subclones

Driver mutations have been depicted in blue. Interestingly, of the identified colorectal cancer driver mutations *PIK3R1* and *TP53* were found in major clones. The mutations in *PIK3R1* and *TP53* in the metastasis encompass regions of LOH with VAF of 88% and 78% respectively. In contrast, mutations in *TCF7L2* and *KRAS* in the metastasis were located in subclonal groups. The VAFs in the *TCF7L2* and *KRAS* were 33% and 35% in the metastasis.

The role of the PIK3 signalling pathways has been studied in colorectal cancer (Fuchs and Ogino, 2013). Class 1A PI3Ks are composed of heterodimer proteins, which are composed of a p110 catalytic subunit and a p85 regulatory subunit. PIK3R1 encodes for an isoform of this regulatory subunit. *PIK3R1* has been reported as a tumour suppressor in hepatocellular carcinomas (Taniguchi et al., 2006). In my analysis the nonsynonymous mutation L347R in *PIK3R1* encompassed a region where 1 allele was deleted, which would be in keeping with a tumour suppressor role. The *PIK3R1* mutation was absent in the primary tumour.

A recent study by Lin et al reported that *PIK3R1* negatively regulated EMT in renal cell cancers (Lin et al., 2015). The expression of mRNA of *PIK3R1* negatively correlated with the stage of the tumour, with reduced expression at advanced stages of disease. Moreover, analysis of renal cancer cell lines revealed that a haploid knockout of *PIK3R1* with subsequent down regulation of the gene resulted in increased migratory capacity of the cells.

4.11 Screen for structural variants

Analysis of somatic intermutation distances confirmed the absence of kataegis events in C1MET and P1. A genome wide screen for chromosomal aberrations using copy number data derived from OncoSNP-SEQ did not identify chromothripsis or chromoplexy alterations in both tumours. BreakDancer was utilised with stringent filtering for read depth and quality, targeting breakpoint regions, with the aim to

identify interchromosomal/intrachromosomal translocations. No confident calls were identified in this process after manual inspection of candidate regions on IGV.

4.12 Technical validation of sequencing results

A custom cancer gene panel of 82 genes using the Ion Torrent PGM was used to validate sequencing results. All mutations in the primary and metastasis on the screened list of mutations were identified with 100% concordance [n=8 mutations] (Table 4.3).

Tumour	Gene	Mutation	Type of mutation	Validated
P1	*TP53*	exon5:c.524_535del:p.175_179 del, GGTGGGGGCAGC	In frame deletion	Yes
P1	*KRAS*	exon2:c.35 G>C:p.G12A	Nonsynonymous	Yes
P1	*GNAS*	exon1:c635 C>T p P212L	Nonsynonymous	Yes
C1MET	*ATM*	exon18:c2705 A>GpK902R	Nonsynonymous	Yes
C1MET	*KRAS*	exon2:c.35 G>C:p.G12A	Nonsynonymous	Yes
C1MET	*MTOR*	exon43:c5972 C>TpA1991V	Nonsynonymous	Yes
C1MET	*SMO*	exon1:c286 T>G pS96A	Nonsynonymous	Yes
C1MET	*TP53*	exon5:c.524_535del:p.175_179 del, GGTGGGGGCAGC	In frame deletion	Yes

Table 4.3l Summary of validated mutations in C1MET and P1

Mutations were confirmed through visual inspection on IGV and the automatic torrent server output. *TP53* and *KRAS* mutations were confirmed to be both in the primary and metastasis, whereas *GNAS* P212L existed exclusively in the primary.

4.13 Comparative analysis of heterogeneity using multi-region analysis of primary colorectal adenocarcinomas

Analysis of the mutation burdens and VAF of C1MET and P1 suggested that the primary tumour continued to evolve following detachment of the metastatic founder cell. To appreciate the degree of heterogeneity between the primary and C1MET I extended my analysis to characterise spatial heterogeneity within 2 primary colorectal adenocarcinomas to serve as a comparison. Using similar sequencing parameters, I characterised the degree of heterogeneity that existed within 2 primary colorectal adenocarcinomas, MR1 [stage T4b N1 M0] and MR2 [stage T3 N0 M0]. Multiple biopsies from each tumour [MR1 6 biopsies, MR2 7 biopsies] were taken and underwent WGS using the Illumina HiSeq 2000 platform. The tumours were obtained following surgery from chemotherapy naive patients. Pathogenic somatic driver mutations in colorectal cancer were represented in all biopsied regions in MR1 in *APC* [Q13030X], whereas *APC* [R284X] was represented in all regions in MR2. Each region sequenced in the multi-region analysis was used as a reference in turn to assess the degree of spatial heterogeneity. The overall mutation burdens in MR1 and MR2 ranged from 20,633 to 22,541 [median 21,617] and 11,168-12,222 [median 11,243] respectively. SNV mutations form a reliable measure for comparison of mutation burdens. The number of SNV mutations in MR1 and MR2 ranged from 16,777-18,430 [median 17,615] and 7913-10066 [median 9,279] respectively (Figures 4.8-4.9).

Multiregion sequencing undertaken in metastatic renal cell carcinoma concluded that a single biopsy was not representative of the genomic landscape of the tumour. They reported that 63-69% of somatic mutations were not identifiable in every sequenced region of the tumour (Gerlinger et al., 2012). In comparison multiregion sequencing of 11 localised lung adenocarcinomas revealed that 76% of mutations were ubiquitous in regions sampled (Zhang et al., 2014).

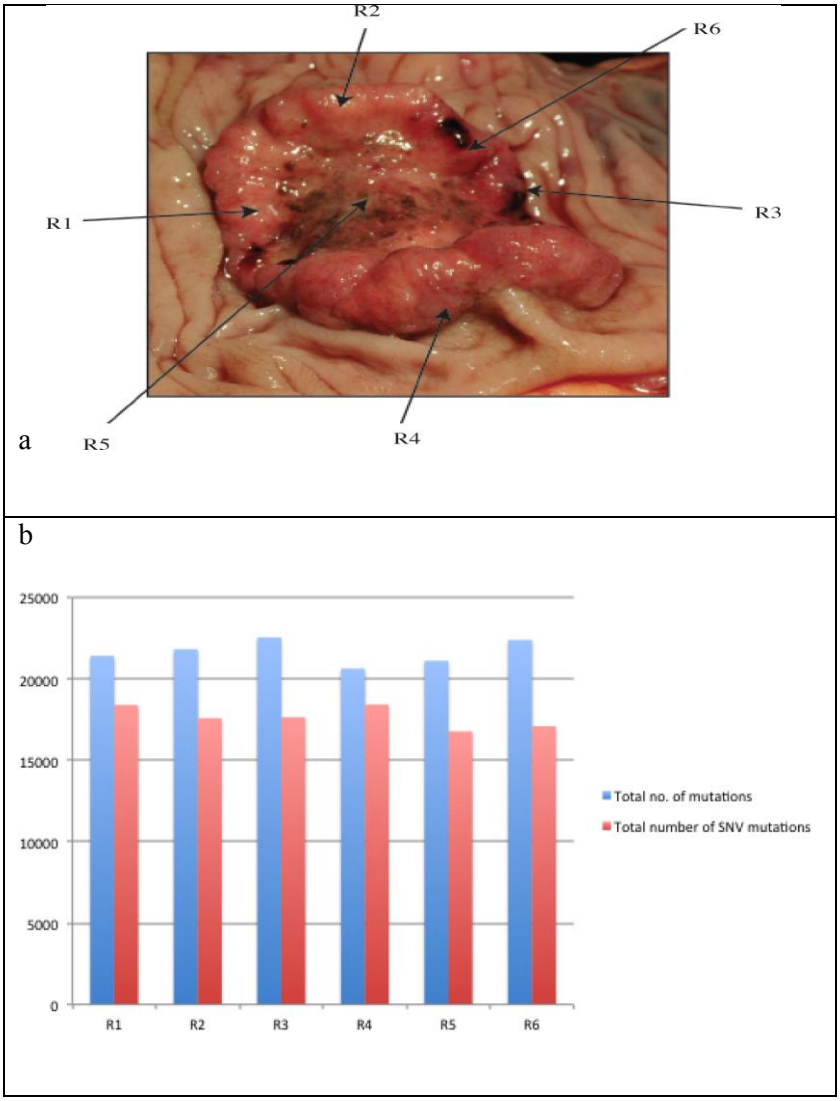


Figure 4.8 l Mutation burdens in multiregion analysis in primary colorectal tumour MR1. a) Photographic representation of the biopsied sites. R1-leading edge 1, R2-leading edge 2, R3-leading edge 3, R4-leading edge 4, R5-necrotic centre, R6-leading edge 5. b) Mutation burdens in MRI are depicted. The number of SNV mutations across the genome are comparable.

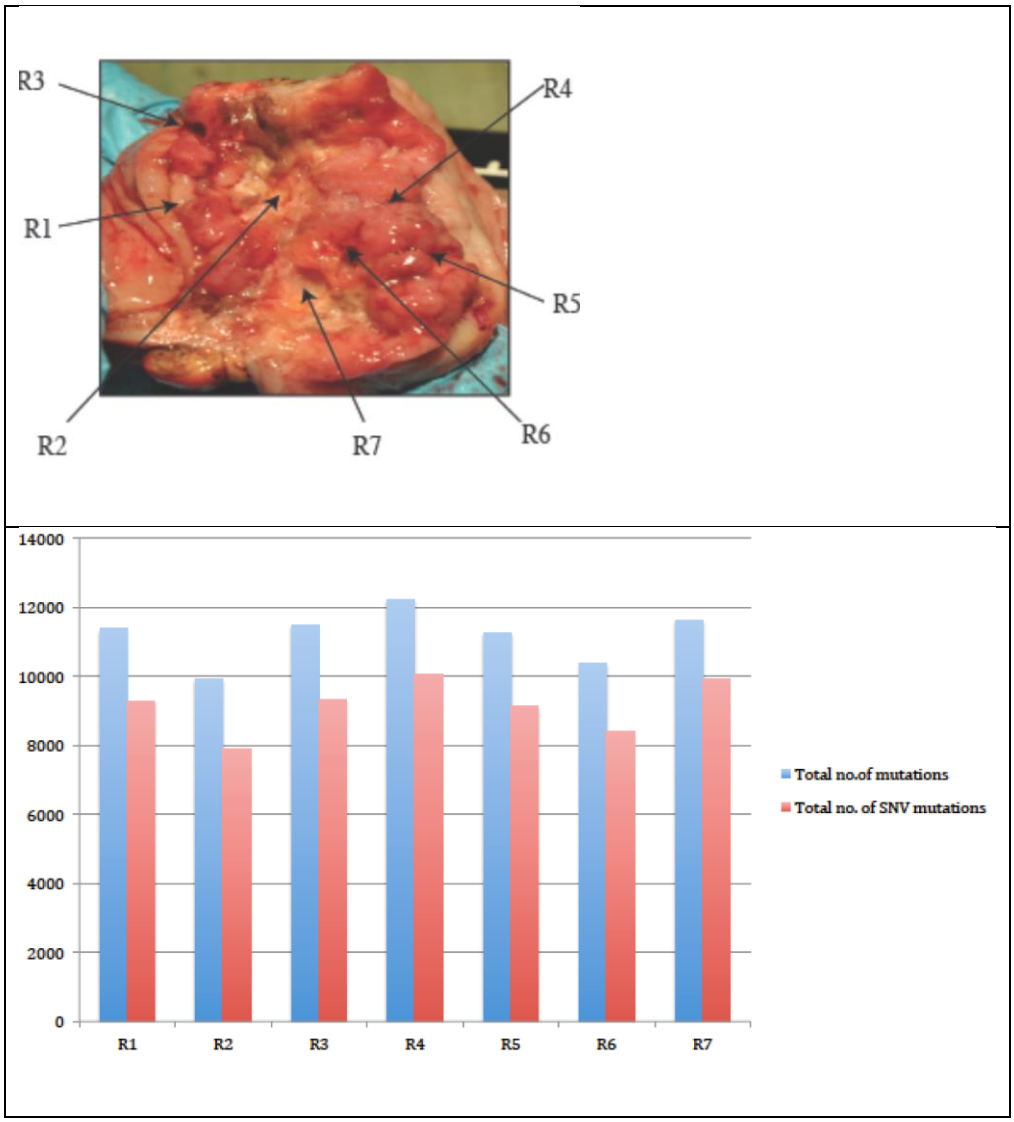


Figure 4.9 l Mutation burdens in multi-region analysis in the primary colorectal tumour MR2. a) Photographic representation of the biopsied sites. R1-leading edge 1, R2-necrotice centre 1, R3 leading edge 2, R4 leading edge 3, R5 leading edge 4, R6 central edge, R7 necrotic centre 2. b) Mutation burdens in MR2 are depicted. The number of SNV mutations across the genome are comparable.

MR1		Percentage of overlap of SNV based on reference					
		R1	R2	R3	R4	R5	R6
Reference	R1	-----	89	78	81	77	78
	R2	93	-----	79	82	78	79
	R3	81	78	-----	83	79	85
	R4	81	79	80	-----	83	82
	R5	85	82	83	91	-----	85
	R6	83	81	88	90	83	-----

Table 4.4 I Spatial heterogeneity in MR1. The proportion of SNV mutations overlapping each region.

MR2		Percentage of overlap of SNV based on reference						
		R1	R2	R3	R4	R5	R6	R7
Reference	R1	-----	81	93	88	87	82	90
	R2	95	-----	94	90	89	84	93
	R3	92	80	-----	80	86	80	89
	R4	81	76	80	-----	85	80	83
	R5	89	76	88	94	-----	88	92
	R6	91	79	90	96	96	-----	83
	R7	88	78	88	89	88	83	-----

Table 4.5 I Spatial heterogeneity in MR2. The proportion of SNV mutations overlapping each region.

In my analysis I report a greater degree of concordance in multiregion sequencing of the primary colorectal cancers. The proportion of SNV mutations exclusive to each region ranged from 2-9% [median 5] in MR1 and 2-4% [median 3] in MR2. In MR1 the percentage of SNVs shared amongst the biopsied regions ranged from 77-93%, whereas it was 76-96% in MR2 (Tables 4.4-4.5). In comparison, the proportion of SNVs in P1 represented in C1MET was 96%. In view of these findings, the degree to which the primary is represented in the brain metastasis emulates the degree of spatial heterogeneity that exists within the chemotherapy naive primary colorectal adenocarcinomas analysed.

4.14 Discussion

4.14.1 Comparison of sequencing studies of primary and metastatic tumours

To date NGS of CCBM have not been published in the literature. Several studies have reported their findings from matched primary/metastasis comparisons. Shah et al sequenced a matched primary and brain metastasis from an oestrogen receptor positive lobular breast carcinoma (Shah et al., 2009). They reported 32 nonsynonymous somatic mutations in the brain metastasis. Comparison with the primary tumour demonstrated that 11 (34%) were detected in the primary tumour, which was resected 9 years earlier. Ding et al sequenced DNA derived from peripheral blood, a basal-like breast primary tumour, the brain metastasis and a xenograft derived from the original primary tumour (Ding et al., 2010). The bulk of mutations were shared between the primary tumour and the metastasis (n=20), with only 2 *de novo* mutations arising exclusively in the metastasis. The time interval between the primary tumour and the development of the brain metastasis was 8 months. More recently, Xi et al published WGS findings from 2 matched primary colorectal carcinomas and corresponding liver metastases, which were resected synchronously from chemotherapy-naive patients (Xie et al., 2014). They reported

that in one pair 25% of mutations in the primary were not represented in the metastasis whereas 95% of mutations in the second pair were shared with metastasis. In both matched pairs, the primary and metastasis continued to develop independent mutations.

In my study, although the majority of mutations in P1 were found within the metastasis a large number of *de novo* mutations in C1MET evolved in short time frame. The limited number of mutations exclusively found in the primary may have resulted from surgical resection of the primary tumour occurring shortly after detachment of the metastatic founder cell. Following dissemination, the metastatic founder cell underwent a change in mutation signature, accompanied with an increase in mutation burden and complexity of clonal architecture.

4.14.2 Parallel and linear progression models

Historically the parallel and linear progression models have been proposed to describe the process of metastasis (Klein, 2009). It is likely that these two models may co-exist to some degree and metastatic development is an evolutionary process. From my findings it is evident that the mutations in the primary are clearly represented in the metastasis. The metastasis has continued to evolve rapidly with the acquisition of new colorectal cancer driver mutations in *P1K3R1* and *TCFL2*. From the data I presented there is evidence that the primary has undergone limited independent evolution following the detachment of the metastatic founder cell.

In summary, I have reported the first comprehensive characterisation of a matched CCBM and primary tumour. There is significant heterogeneity between the primary tumour and brain metastasis with regard to mutation signatures, driver mutations and clonal complexity. This has important therapeutic implications, as these patients present at an advanced stage and unlike other sites of metastasis, biopsy of the metastasis to guide treatment decisions may be inappropriate or impractical.

Advances in this field may require genomic characterisation of a larger cohort matched primary tumours and brain metastases to identify potential druggable targets. Importantly, I have demonstrated that reliable data can be obtained from archived FFPE tumours from WGS that can be technically validated.

Chapter 5

Analysis of lung adenocarcinoma brain metastases genomes

5.1 Introduction

Lung cancer is the leading cause of cancer related deaths worldwide (Siegel et al., 2014). Lung cancer can be broadly subdivided in to non-small cell lung cancer (NSCLC) and small-cell lung cancer (SCLC), where the former constitutes ~85% of cases (Ettinger et al., 2013). Adenocarcinoma of the lung forms the predominant histological subtype of NSCLC (~60%). Tobacco smoking has been implicated as a causal agent in all subtypes of lung cancer, with adenocarcinoma of the lung the most common subtype in individuals who have never smoked (Herbst et al., 2008).

Patients with lung cancer often present at an advanced stage, and have a poor prognosis with a 5 year survival across all stages of disease of 15% (Parkin et al., 2005). This has been ascribed to patients often presenting with advanced disease and limited therapeutic options. Patients with lung cancer present initially with brain metastases in 10-25% of cases, with 40% of patients developing brain metastases throughout the course of their disease (Hanibuchi et al., 2014).

In recent years, there has been increased efforts to elucidate the genetic profile of primary lung adenocarcinoma genomes. To gain insight into the biological mechanisms underpinning tumourigenesis and to characterise heterogeneity two approaches have been adopted. Sequencing strategies employed included analysing large cohorts of tumours or multiregional sampling from individual tumours (Imielinski et al., 2012, de Bruin et al., 2014). To date, characterisation of the genomes of patients with lung adenocarcinoma brain metastases (LABM) in the literature have

been sparse, in part due to human tissue availability. An appreciation of the heterogeneity that exists between LABM and primary tumours will be informative in developing future therapies. In this chapter, I have reported my findings from the characterisation of 7 LABM genomes including: mutation burden and spectra/signatures, copy number aberrations, structural variants and clonal architecture of the brain metastases. Data from TCGA from primary lung adenocarcinomas served as a comparison data set.

5.2 Methods

5.2.1 WGS of LABM genomes

DNA from 7 LABM and paired buffy coat lymphocytes were sequenced. Bioinformatic processing and analysis is outlined in Chapter 2 (sections 2.11 and 2.14).

5.3 Comparison Datasets

5.3.1 TCGA primary lung adenocarcinoma tumours

Somatic mutations from primary lung adenocarcinoma tumours were obtained from TCGA (section 2.2.1). The mutation spectra and driver genes in primary tumours were characterised in a cohort of 300 tumours. The sequenced brain metastasis were genotyped using SNP arrays to obtain copy number information and primary lung adenocarcinomas from TCGA were used as a comparison data set (section 2.16).

5.4 Patient Demographics

Clinical information pertaining to the patients who had their brain metastases sequenced is provided (Table 5.1). Of the 7 patients who had their brain metastasis sequenced, 5 patients presented with a brain metastasis at primary presentation.

Lung adenocarcinoma	Age of patient	Gender	Histology of primary tumour	Time interval between resection of brain metastasis (years)	Treatments received by patient following presentation
L1MET	65	F	Adenocarcinoma	0	Presented with brain metastasis initially
L2MET	74	M	Adenocarcinoma	0	Presented with brain metastasis initially
L3MET	66	M	Adenocarcinoma	0	Presented with brain metastasis initially
L4MET	58	M	Adenocarcinoma	1	Presented with primary lung tumour in 2009 Treatment for primary tumour unavailable.
L6MET	55	F	Adenocarcinoma	0	Presented with brain metastasis initially
L7MET	67	M	Adenocarcinoma	0	Presented with brain metastasis initially
L8MET	48	M	Adenocarcinoma	3	Presented with primary lung tumour in 2008 Had 6 cycles of cisplatin and pemetrexed in 2008. Right middle lobectomy 2009 and radiotherapy. Developed a brain metastasis in

Table 5.1 I Summary of clinical information for patients who had their lung adenocarcinoma brain metastasis sequenced.

Results

5.5 Mutation burden

5.5.1 Mutation burdens of LABM

The LABM had a median total number of somatic mutations of 51897 [range 7847 to 123806]. Of these mutations there was a median number of 49206 SNVs [range 6524 to 118490]. There was a wide variation in the number of somatic mutations in the 7 LABM, with the total number of mutations greatest in L7MET (123,806) and lowest in L8MET (7,847) (Figure 5.1).

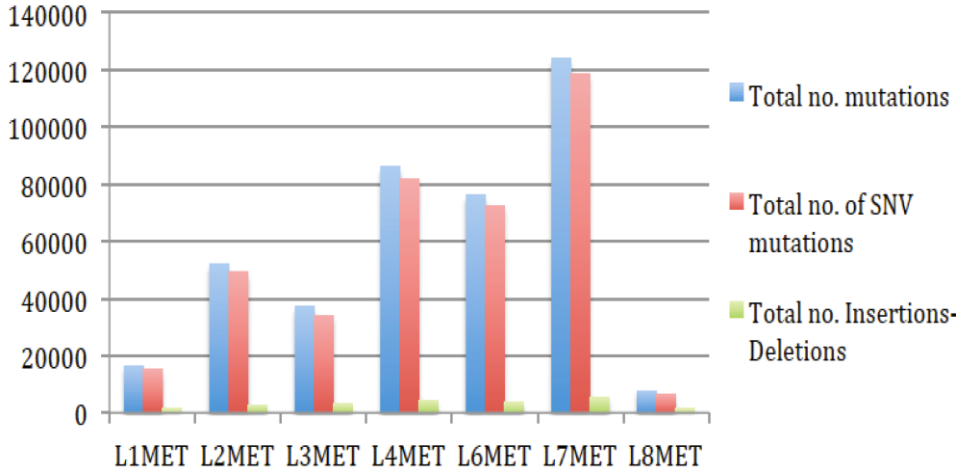


Figure 5.1 l Mutation burdens in LABM in the whole genome. The total number of mutations, SNV mutations and the number of small indels are depicted.

The total number of small indels in the LABM ranged from 1323 to 5316 [median 3058] representing 4 to 16% of the total mutation burdens in the genomes. Although L1MET and L8MET had the lowest number of somatic mutations amongst the LABM, small indels were proportionately higher, 9 and 16% respectively in these 2 metastases in comparison to the other metastases [range 4 to 8%]. Subgroup analysis of small indels identified deletions being more common events in all the LABM forming 71 to 77 % of small indel mutations.

The distribution of SNVs were analysed across the whole genome, categorising mutations based on exonic, intronic and intergenic locations. SNVs located in the exome formed ≤ 1% of the mutation burdens [range 0.6 to 1%]. In comparison, the majority of SNVs were located in intergenic regions in all the LABM [range 66 to 71%]. The proportion of SNVs located in intronic regions ranged from 26 to 32% of the SNV mutation burden. There was an association between the number of SNV mutations in the exome and the number of SNV mutations in intronic (Spearman $\rho=0.98$, $p<0.002$) and intergenic (Spearman $\rho=0.97$ $p<0.002$) regions. The number of mutations in the exome ranged from 2 to 32 per 10^6 (Figure 5.2).

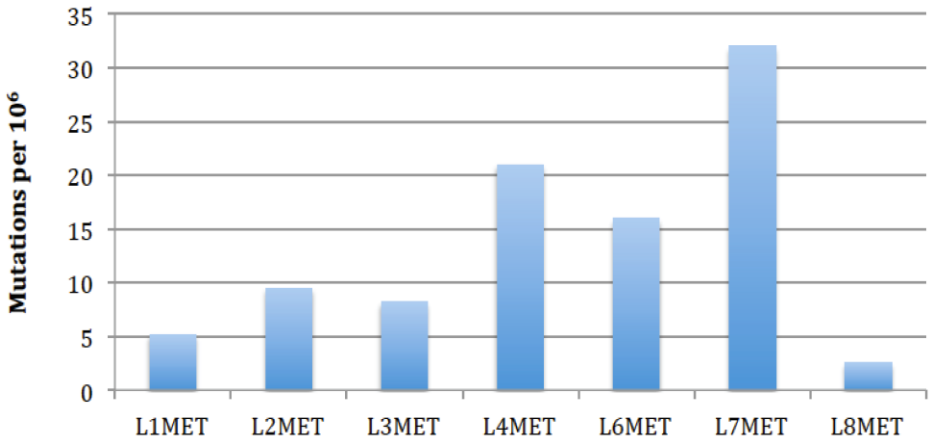


Figure 5.2 I Number of mutations in exome in LABM. Mutations are represented per megabase.

In the exome, synonymous mutations formed 18-28% of mutations. Sequencing data were filtered to determine the number of potentially functional mutations in protein-coding regions (nonsynonymous, frameshift, splice-site and nonsense mutations). A median of 219 [range 57 to 691] were potentially functional mutations within protein-coding regions. Of the potentially functional mutations, nonsynonymous mutations formed the most common subtype (66 to 78% of functional mutations) (Figure 5.3). The number of insertions-deletions, within coding regions ranged from 7 to 33 events

per tumour [median 14]. The number of nonsense mutations detected ranged from 5 to 48 events per tumour [median 12]. No splice-site mutations were detected.

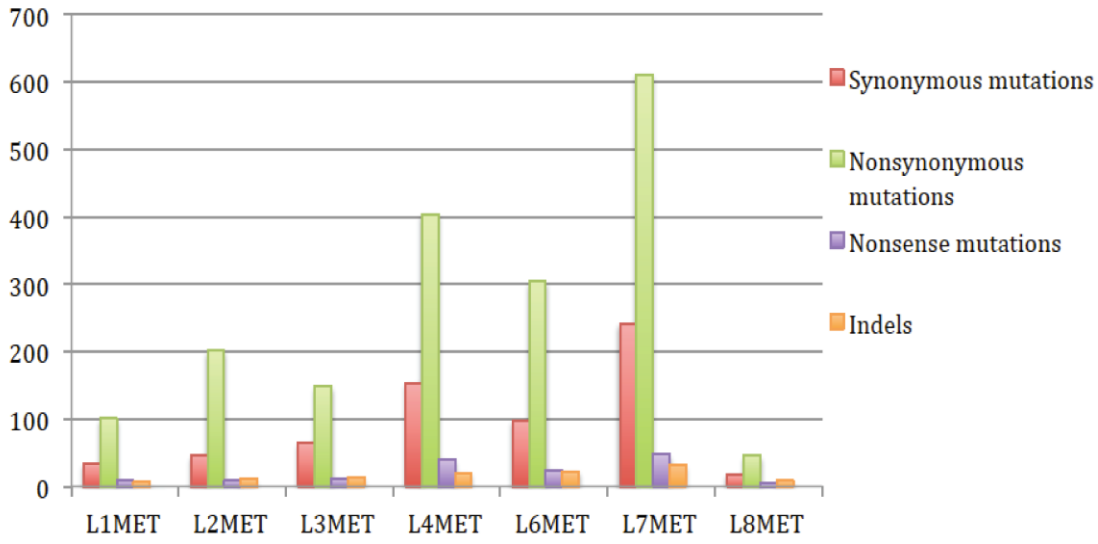


Figure 5.3 I Distribution of types of protein-coding mutations in LABM.

The mean nonsynonymous:synonymous mutation ratio was 2.9:1. The transition/transversion ratios genome wide were calculated and ranged from 0.4 to 1 in the LABM.

Of note, despite a modest sample size, the mutation burdens in L1MET and L8MET were the lowest amongst the brain metastases, 7847 and 16454 somatic mutations respectively. Both these tumours have pathogenic *EGFR* mutations [L1MET exon21, L858R, L8MET exon19:c236-2250del GAATTAAGAGAAGCA] resulting in constitutive activity of the tyrosine kinase receptor (Santos et al., 2011). This finding is consistent with published findings in primary tumours documenting the correlation between the presence of activating *EGFR* mutations and lower mutation frequencies (Ding et al., 2008, Imielinski et al., 2012). Ding et al postulated that the correlation between the presence of a low mutation frequency and *EGFR* mutation status may be related to reduced selection pressures in these tumours as these tumours may have strong tumour promoting activities through oncogene addiction (Ding et al.,

2008). Interestingly, L8MET which harboured a pathogenic 15 base non-frame shift deletion, had the largest proportion of small indels represented in the overall mutation burdens (16%) amongst all LABM.

Of the brain metastases sequenced L4MET and L8MET were derived from patients that initially presented with a primary tumour and developed a brain metastasis after an interval time period. The remaining sequenced brain metastases were derived from patients who had developed brain metastases at their initial presentation. This latter group of patients had their primary tumour resected prior to exposure to chemotherapeutic agents. Although the sample size is limited the mutation burden in these groups of patients were comparable.

5.6 Analysis of mutation spectra

Mutational footprints in lung cancer may be related to either ageing, the activity of APOBEC cytidine deaminases or smoking exposure and the time since the cessation of smoking (de Bruin et al., 2014, Zhang et al., 2014). The mutation spectra of the 6 classes of bases substitutions in the LABM is illustrated in Figure 5.4. In general, C>A/G>T mutations formed the most common type of base substitution. This pattern of base substitution is characteristically associated with tobacco smoking (Pfeifer et al., 2002, Liu et al., 2012c). Polycyclic aromatic hydrocarbons in tobacco smoke form bulky DNA adducts with guanine, which are removed by nucleotide excision base repair mechanisms. Zhang et al undertook multi-region sequencing of primary lung adenocarcinomas, demonstrating that the mutation spectra was reflective of the quantity of smoking exposure and the time following the cessation of smoking (Zhang et al., 2014). C>T/G>A base substitutions predominated in non-smokers. Moreover, the mutation spectra in former smokers who had ceased smoking 20 years ago also demonstrated a C>T/G>A predominant mutation profile, demonstrating a relative decrease in smoking related mutations.

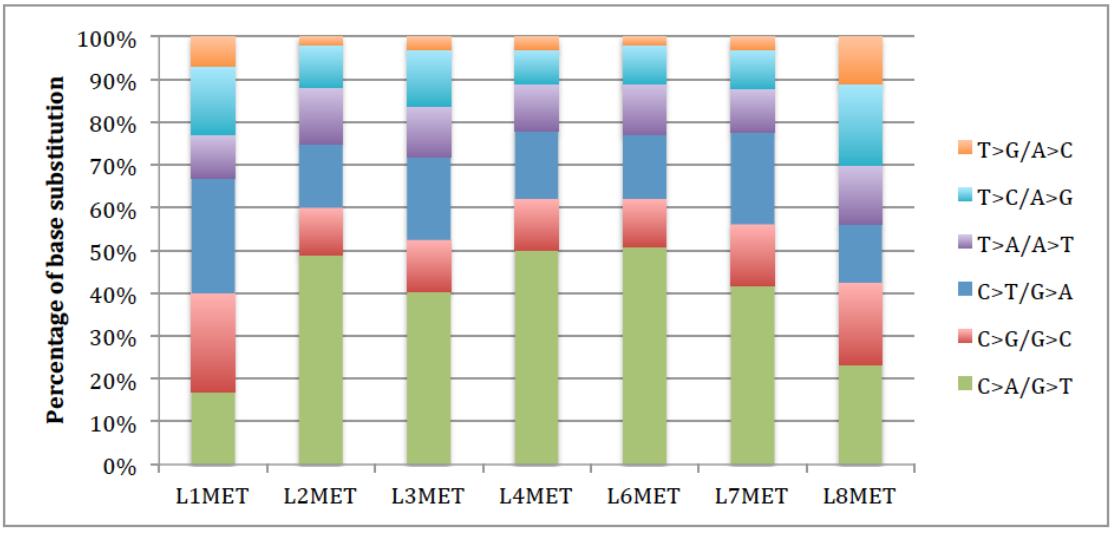


Figure 5.4 I Distribution of the 6 classes of base substitutions in the LABM. With the exception of L1MET and L8MET, C>A/G>T base substitutions predominate.

Unlike the other LABM the mutation spectra of L1MET and L8MET did not exhibit a C>A/G>T base substitution predominance. L1MET and L8MET are known to harbour activating *EFGR* mutations. Activating *EGFR* mutations are more commonly found in light or non-smokers, women and Asians (Dogan et al., 2012). Clinical information regarding the smoking history of the patients in this study was unavailable to reaffirm these findings. L1MET and L8MET were obtained from female and male patients respectively of Caucasian ethnicity.

5.7 Analysis of mutation signatures

Tumours are exposed to numerous biological processes which have their own contribution to the mutation footprint. Analysing somatic mutations in sequence context and deciphering the overriding mutational process can be informative. Using somatic signature, somatic mutations were analysed with regards to the context in which the mutation event occurred taking into account the 5' and 3' base flanking the mutated base (Figure 5.5). Analysis of the LABM generated 3 mutation signatures S1-S3.

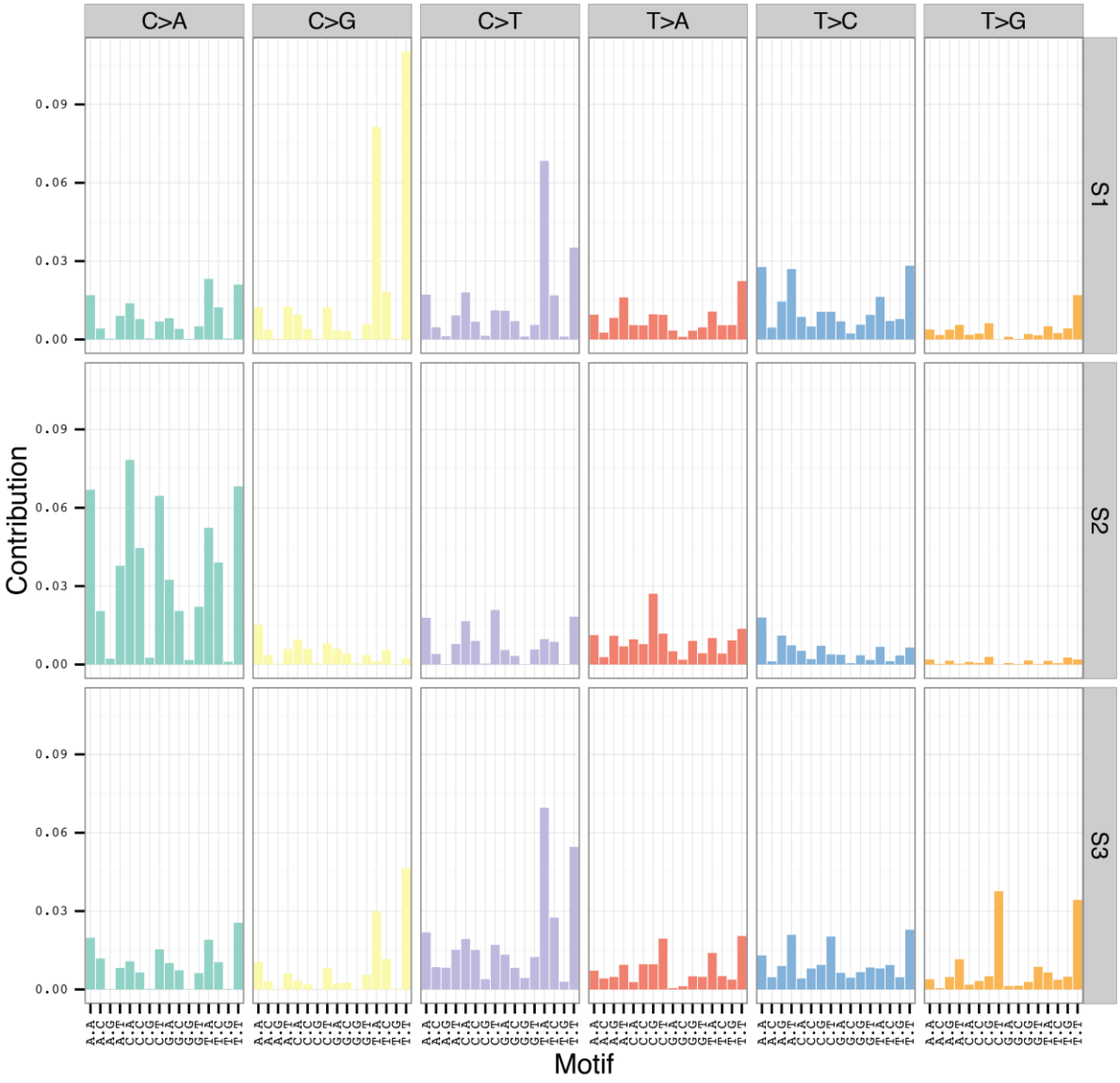


Figure 5.5 I Mutation signatures in LABM. 3 mutation signatures were identified of which S1 and S3 are likely to represent the same biological process.

Alexandrov et al undertook a pan-cancer analysis of mutation signatures in a variety of cancers. Signature S2 in Figure 5.5 was observed in the brain metastases L2MET, L3MET, L4MET, L6MET and L7MET. Signature S2 corresponds to signature [4] in the analysis undertaken by Alexandrov et al which is characterised by an increased

proportion of G>T/C>A mutations across the genome. They demonstrated a statistically significant association between signature [4] and smoking history.

Notably signatures S1 and S3 share similarities in appearance and are likely to represent the same biological process in action. These signatures resemble signature [2] in the pan-cancer analysis by Alexandrov et al. This signature has been associated with APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) cytidine deaminases and is typically characterised by C>G and C>T mutations in the context of a T[C]X motif (where X denotes an adenine or thymine base). Signature [2] is prevalent in numerous histological subtypes of cancers, including lung adenocarcinomas, and was present in 14.4% of the cancers analysed by Alexandrov et al. Although the APOBEC family consists of several enzymes, Burns et al have demonstrated that expression of the APOBEC3B enzyme in particular correlated with the typical mutation spectra and localised clusters of hypermutations termed kataegis (Burns et al., 2013). They proposed that APOBEC3B is responsible for inducing cytosine to uracil substitutions, following which the majority are repaired with fidelity by base excision repair pathways. However it has been proposed that the uracil base may serve to act as a template for the insertion of an adenine in DNA synthesis resulting in a cytosine to thymine change (guanine to adenine substitution on the opposite strand). Additionally, they proposed that uracil DNA glycosylase may also convert uracil into abasic sites. Abasic sites are thought to be non-instructional and when recognised by DNA polymerase and may lead to the insertion of any base opposite.

5.8 Identification of putative drivers of LABM

To identify putative drivers of LABM, IntOGen a program for identifying driver mutations was utilised (FDR <0.05). The pan-cancer gene *TP53* was the only gene identified through this process. I extended my analysis to review truncating mutations

in the same gene in 3 or more LABM. No candidate genes fulfilled these criteria. Although candidate driver mutations of LABM were not identified using these two approaches, this may reflect the modest sample size and insufficient power to detect driver mutations.

5.9 Screen for primary lung adenocarcinoma driver mutations

I sought to determine whether driver mutations in primary lung adenocarcinomas were retained in the brain metastases. Previous studies in brain metastases have primarily focused on screening for targeted mutations in *EGFR* and *KRAS*. A list of candidate driver genes in primary lung adenocarcinomas was generated by running IntOGen and MutSig on a larger cohort of primary tumours obtained from TCGA cBioPortal (FDR <0.05). Copy number and LOH information was integrated with the mutations identified. Somatic mutations were filtered to exclude base substitutions with potentially benign or moderate functional effects (Polyphen 2 > 0.8 and SIFT score < 0.2) after IntOGen/MutSig analysis.

Using this approach, oncogenes and tumour suppressors that have historically been associated with primary lung adenocarcinoma were identified in the LABM (Table 5.2). Mutations were observed in oncogenes *KRAS* and *EGFR* with concurrent gains in copy number at those loci. Activating *EGFR* mutations were identified in L1MET and L8MET, a nonsynonymous mutation in exon 21 (L858R) and an in frame deletion in exon 19 (c236-2250del GAATTAAGAGAAGCA] respectively. Mutations in tumour suppressor genes identified included *TP53* (n=5), *PTEN* (n=1), *STK11* (n=1), KEAP1 (n=1) and *RB1* (n=1). A recent comprehensive analysis of primary lung adenocarcinomas undertaken by the TCGA reported the aforementioned genes as candidate primary lung adenocarcinoma driver mutations (Cancer Genome Atlas Research, 2014). Of the 230 primary lung adenocarcinomas analysed by the TCGA, *TP53* and *KRAS* were the most commonly mutated, 46% and 33% respectively. In

their analysis *KRAS* and *EGFR* mutations were mutually exclusive events which parallels existing reports in the literature (Schmid et al., 2009, van Zandwijk et al., 2007). The mutual exclusivity of *EGFR* and *KRAS* mutations is reflected the LABM genomes. Activating mutations in EGFR in lung adenocarcinomas confers sensitivity to tyrosine kinase inhibitors, which results from a conformational change in the tyrosine kinase domain of the EGFR. The conformational change results in constitutive kinase activity and activation of the downstream signalling cascade. Gefitinib and Erlotinib are tyrosine kinase inhibitors that have been developed to reversibly bind to the ATP-binding site of the EGFR kinase domain and inhibit the activity of the receptor. The tyrosine kinase inhibitors demonstrate a higher binding affinity to EFGR with an activating mutation compared to the wild-type receptor. *KRAS* functions downstream of the EGFR signalling pathway and is associated with a poor response to EGFR-directed tyrosine kinase directed therapy (van Zandwijk et al., 2007, Massarelli et al., 2007). Multi-region analysis of lung adenocarcinomas demonstrated that 40% of 'trunkal' driver mutations in primary lung adenocarcinomas were characterised by G>T/C>A mutations (de Bruin et al., 2014). Table 5.2 demonstrates G>T/C>A nonsynonymous mutations in *TP53, KRAS* and *PTEN* in the LABM.

Gene	Sample	Location of mutation	Type of mutation	LOH at locus of mutation	Copy Number at locus of mutation	PolyPhen2 Score	SIFT Score	Mutation frequency in gene TCGA cBioPortal %	Mutation frequency in gene at that loci -TCGA cBioPortal
EGFR	L1MET	exon21:c.2573 T>G:p.L858R	Nonsynonymous	Yes	7	1	0	13% (24/182)	8
	L8MET	exon19:c.2236_2250 del.746_750del GAATTAAGAGAAGCA	In-frame deletion	No	4	--------	--------		0
TP53	L2MET	exon8:c.840 T>G:p.R280S	Nonsynonymous	Yes	1	1	0	50% (93/182)	0
	L3MET	exon8:c.796 C>Ap.G266X	Nonsense	Yes	2	1	0		0
	L4MET	exon7:c.742 G>Ap.R248W	Nonsynonymous	No	3	1	1		1
	L7MET	exon8:c.818 C>Ap.R273L	Nonsynonymous	Yes	2	1	0		4
	L8MET	exon8:c.818 C>Tp.R273H	Nonsynonymous	Yes	3	1	0.01		2
KRAS	L6MET	exon2:c.34 C>Ap.G12C	Nonsynonymous	Yes	6	1	0.01	30% (54/182)	19
PTEN	L3MET	exon5:c.277 C>A:p.H93N	Nonsynonymous	Yes	1	1	0	2% (4/182)	0
RB1	L3MET	exon14:c.1333 p.R445X	Nonsense	Yes	1	--------	--------	3% (6/182)	0
STK11	L4MET	exon2:c.334 C>T:pQ112X	Nonsense	Yes	1	--------	--------	15% (27/182)	0
KEAP1	L7MET	exon3:c.1255 G>A:p.G419R	Nonsynonymous	Yes	2	1	0	12% (22/182)	0

Table 5.2 l Summary of identified primary drivers in LABM

5.10 Clonal architecture of LABM

Pyclone was utilised to infer the structure of clonal populations in LABM genomes. The number of clonal populations reported in each brain metastasis is listed in Table 5.3 and varied in complexity ranging from 4-39 [median 23]. Generally the complexity of the clonal populations increased with the number of somatic SNVs. Of note, L1MET and L8MET brain metastasis which had activating *EGFR* mutations had the least complex clonal structures. Of all the LABM, L6MET had the most complex clonal structure, though SNV mutation burden was higher in other brain metastases.

Lung adenocarcinoma brain metastasis	Number of clones	Number of SNV mutations
L1MET	7	15077
L2MET	13	49206
L3MET	23	34082
L4MET	25	81794
L6MET	39	72636
L7MET	37	118490
L8MET	4	6524

Table 5.3 I Summary of clonal populations in the LABM

The clonal architecture of the metastases are depicted in Figures 5.6-5.12. Although driver mutations for LABM were not identified in my analysis, the hierarchical position of existing primary lung adenocarcinomas driver mutations in the framework of a brain metastasis genome is of interest. In figures 5.6-5.12 candidate primary lung adenocarcinoma driver genes are highlighted in blue and plotted alongside randomly selected genes to represent their position in the context of the inferred clonal architecture.

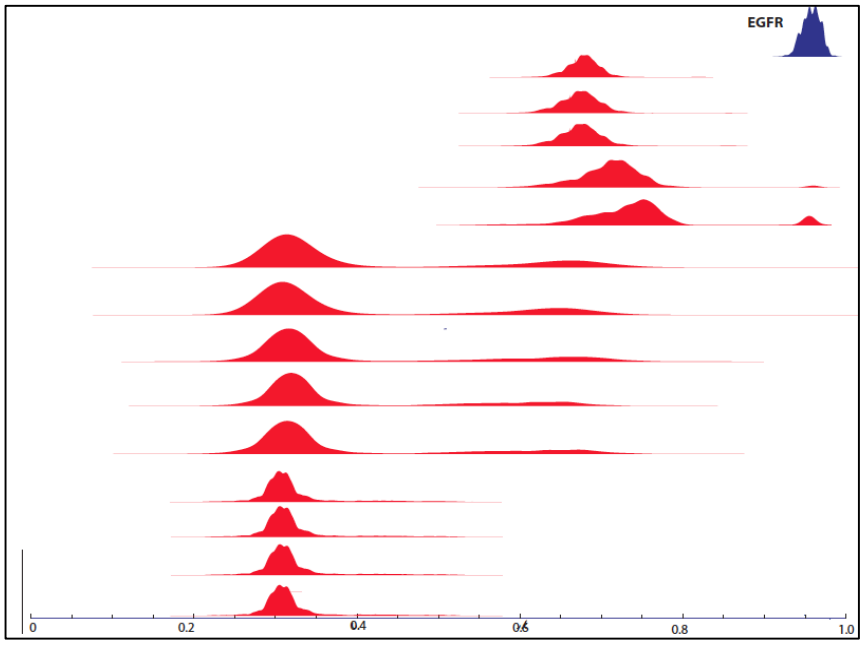


Figure 5.6 I Clonal architecture in brain metastasis L1MET. Somatic mutations in randomly selected genes are represented (driver mutations are highlighted in blue). *EGFR* is a major clone in L1MET. The gene occurs in region of amplification (copy number 7) with concurrent LOH.

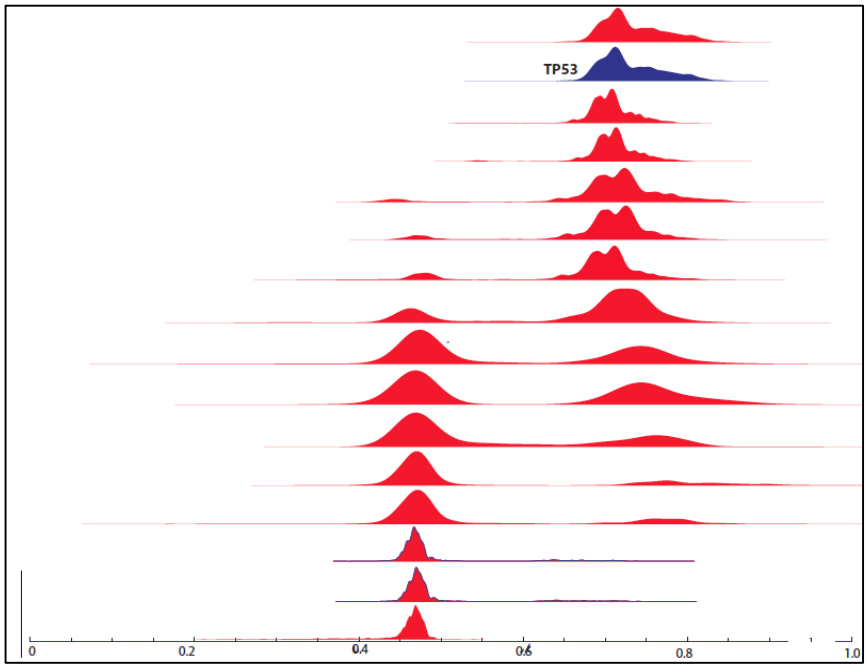


Figure 5.7 I Clonal architecture in brain metastasis L2MET. *TP53* is located in a major clone in L2MET in a region of LOH.

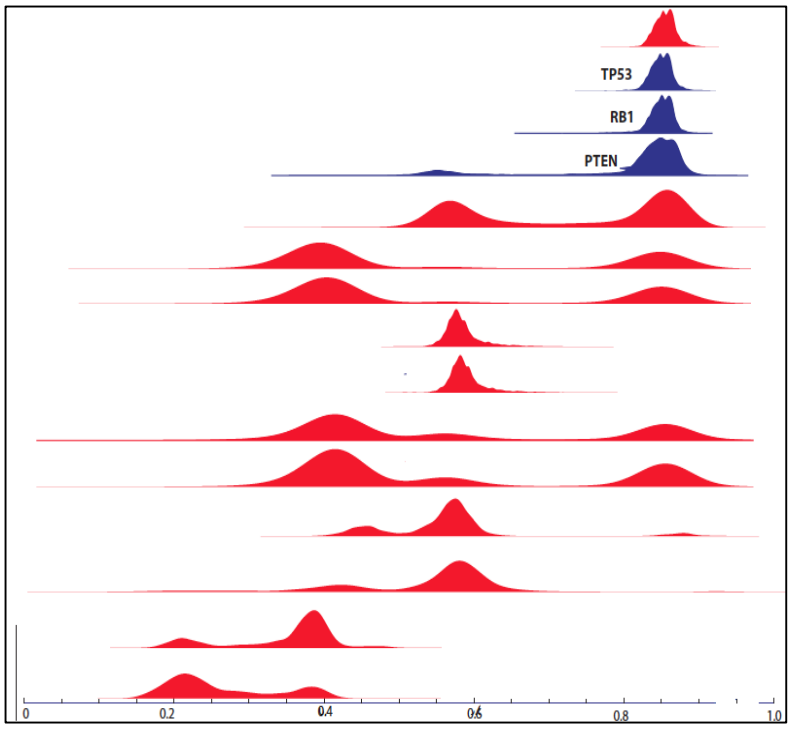


Figure 5.8 l Clonal architecture in brain metastasis L3MET. Tumour suppressors *TP53, RB1* and *PTEN* are located in a major clone in L3MET. *TP53, RB1* and *PTEN* are found in regions of LOH

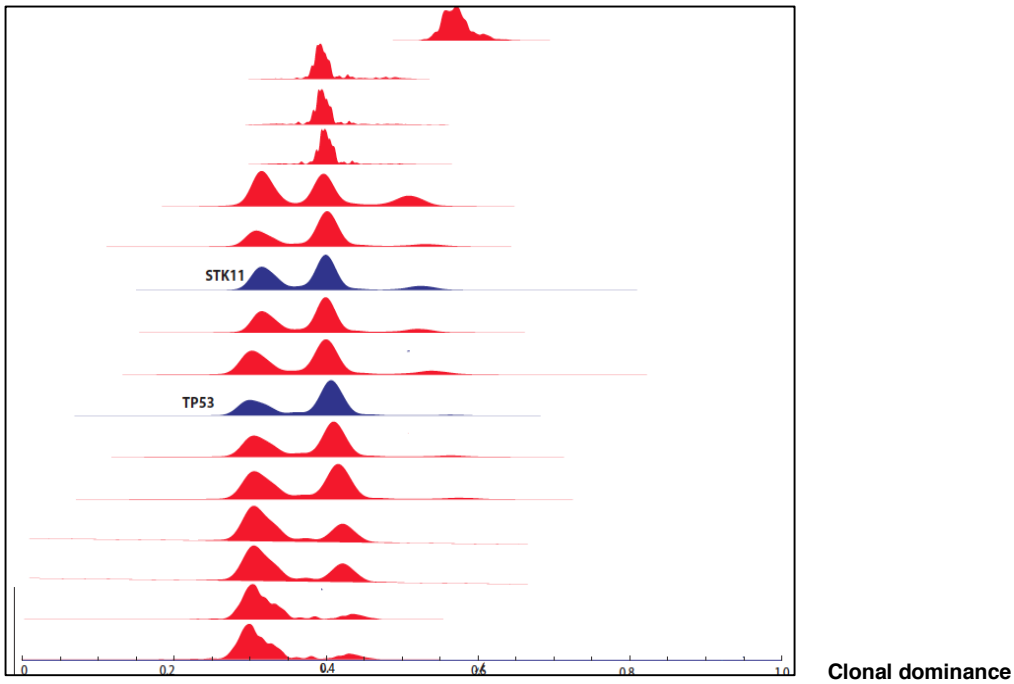


Figure 5.9 l Clonal architecture in brain metastasis L4MET. *STK11* and *TP53* are present in subclones in the clonal structure

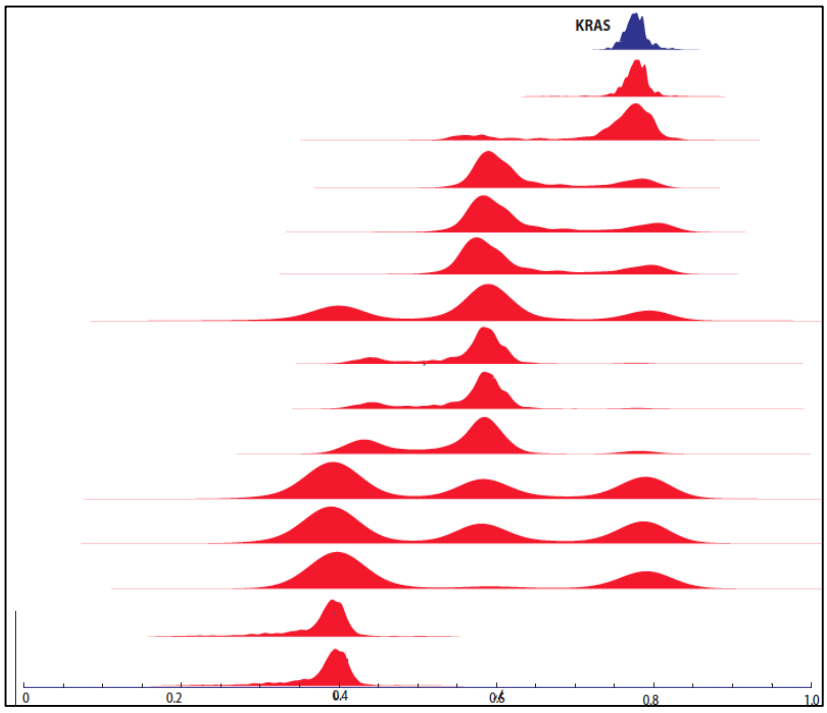


Figure 5.10 | Clonal architecture in brain metastasis L6MET. KRAS (G12C) is located in a dominant clone in L5MET. The mutation overlies a region of amplification, copy number 6.

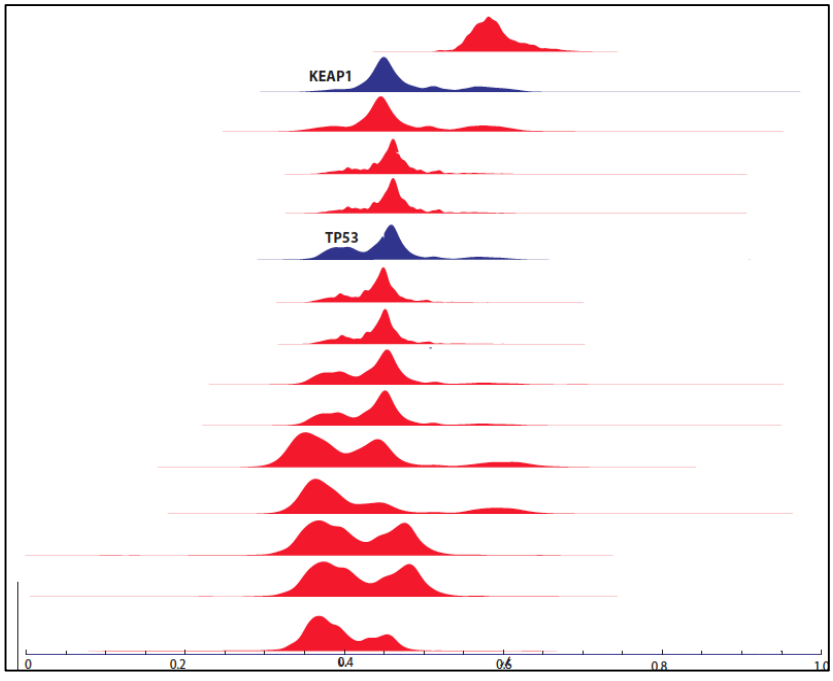


Figure 5.11 | Clonal architecture in brain metastasis L7MET. Driver mutations *KEAP1* and *TP53* are located in major subclones.

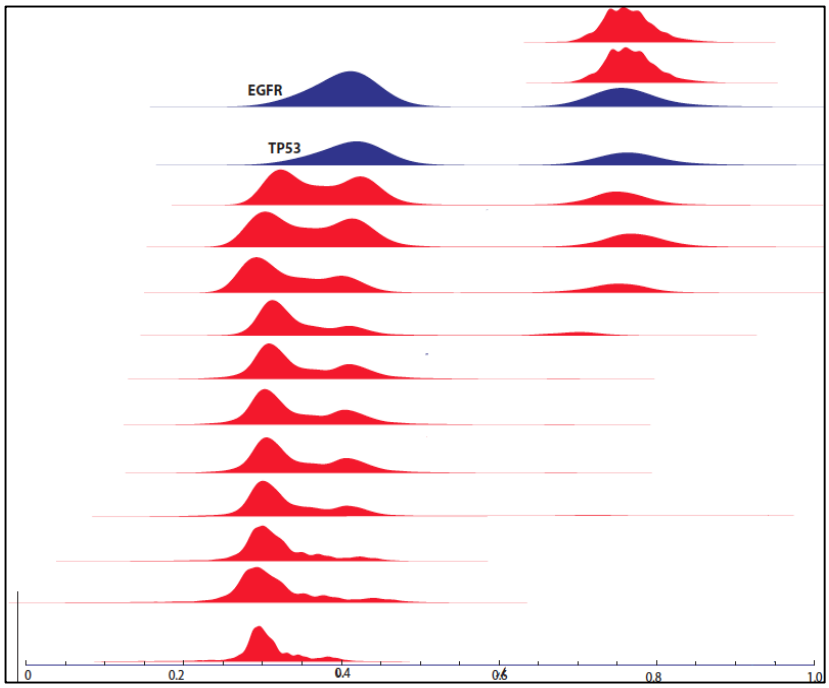


Figure 5.12 l Clonal architecture in brain metastasis L8MET. *EGFR* and *TP53* are located in major subclones in L8MET.

Generally somatic mutations in drivers of primary lung adenocarcinomas mapped to major clones. *EGFR*, *KRAS*, *PTEN* and *RB1* were often located in major clones. To endeavour to understand the relative position of driver genes, VAF, copy number and LOH information were taken into account. The median SNV VAF of the LABM ranged from 18 to 66 [median 41]. *EGFR* and *KRAS* were located in major clones, which may relate to the concurrent occurrence of these mutations in regions of high copy gains with preferential amplification of the allele with the somatic mutation. In L1MET and L6MET *EGFR* and *KRAS* mutations were found in regions of LOH with amplifications with copy numbers of 7 and 6 respectively at these loci. Tumour suppressor genes were frequently found within regions of LOH. In L7MET *KEAP* and *TP53* mutations were found in subclonal populations overlying regions of copy neutral LOH. Reconstruction of the origins of 'trunkal' and 'branched' mutations in a subset of primary lung adenocarcinomas demonstrated that amongst others, mutations in *EGFR, TP53, RB1* were located in the founding clones of the tumour.

Mutations in these genes in the LABM were frequently found in the major clonal groups. *KRAS* formed a dominant clone in L6MET which is also the most complex clonal structure amongst the LABM.

5.11 Copy number aberrations in the LABM

5.11.1 Evaluation of copy number alterations in LABM

Tumour	Number of CNA	Proportion of genome deviating from mean copy number	Chromosome-scale deletion	Chromosome-scale gain
L1MET	43	42	3p, 8p, 9q, 13q, 15q, 22q	7p, 18p, 3q
L2MET	55	40	chr19, 21q, 22q	chr20, 5p ,6p, 8p,
L3MET	15	17	chr 10, 3p,17p,13q 15q	----------
L4MET	64	48	chr19, 6q 9q 21q	5p,10p.
L6MET	52	41	8p,4q.5q,11q,15q,21q	----------
L7MET	48	27	16p,	5p, 10p, 3q
L8MET	35	48	8p,19p, 13q	5p, 7p, 1q, 11q, 21q, 22q

Table 5.4 l Summary of copy number aberrations in individual LABM

Table 5.4 summarises copy number aberrations analysed in the LABM. Deviations from the average copy number were identified in each tumour defined as a CNA event. The brain metastases had a median number of 48 CNA (range 15-64). The proportion of the genome that deviated from the average copy number ranged from 17 to 48 [median =41]. Although there was trend for number of CNA events to be related to the proportion of the genome deviating from the mean copy number this failed to reach statistical significance (Spearman Rho =0.679, p=0.06)

Chromosome-scale aberrations were identified in each tumour (arbitrarily defined as >90% involvement of the arm). Recurrent chromosome-scale changes were categorised as those present in 3 or more LABM in the same arm. Using these criteria gains were only identified in 5p (n=4, L2MET, L4MET, L7MET, L8MET). In comparison, chromosome-scale deletions were more frequent events. They were present in 8p (n=3, L1MET, L6MET, L8MET), 19p (n=3, L2MET, L4MET, L8MET), 13q (n=3, L1MET, L3MET, L8MET) 15q (n=3 L1MET, L3MET, L6MET), and 21q (n=3 L2MET, L4MET, L6MET). Whole-chromosome gains were observed in L2MET in chromosome 20 involving both p and q arms, whereas whole-chromosome deletions were observed in chromosome 19 in L2MET and L4MET and in chromosome 10 in L3MET. Chromosome-scale alterations did not co-localise with the somatic mutations identified through sequencing.

5.11.2 Identification of recurrent focal copy number aberrations

Recurrent focal copy number aberrations (< 2MB in size) were searched for in genome-wide screen to identify candidate oncogenes or tumour suppressor genes. A focal amplification was observed in L3MET and L4MET encompassing the gene DLGAP1 (Figure 5.13). DLGAP1 encodes disk large associated protein. The role of this protein is unclear, though it is found in synaptic junctions (Yao et al., 2002). This gene is reportedly amplified in ~2% of primary adenocarcinoma (TCGA cBioPortal).

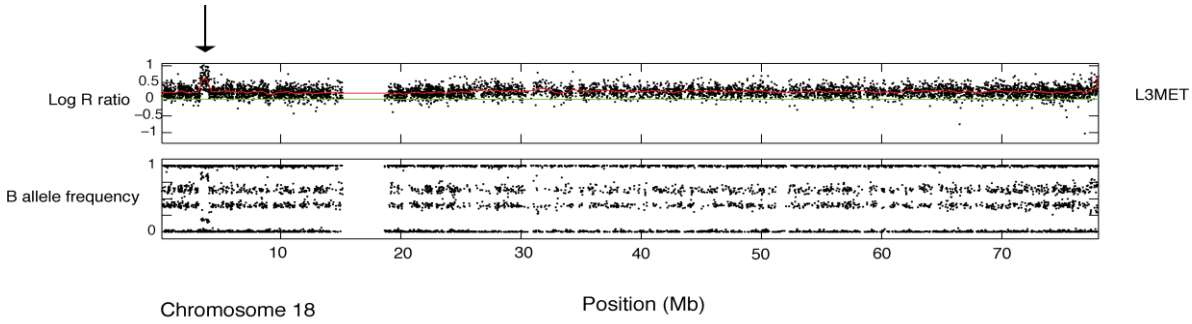


Figure 5.13 I Example of focal amplification in DLGAP1 in L3MET in chromosome 18 from SNP arrays

5.11.3 Concordance of copy number information from SNP array and sequencing data.

OncoSNP-SEQ a software program that infers copy number and LOH information directly from sequencing data was utilised to validate findings from SNP arrays. Focal amplifications in L3MET and L4MET in DLAGPA1 were verified (Figure 5.14).

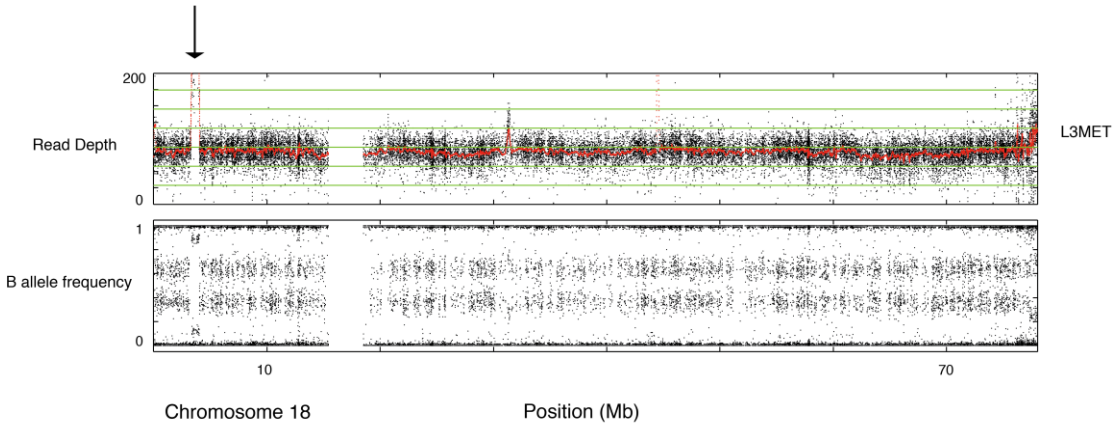


Figure 5.14 l Focal amplification in DLGAP1 in L3MET detected in OncoSNP-SEQ. Copy number profiles have been inferred directly from sequencing data.

The degree of concordance between copy number calls generated using SNP array information and that directly inferred from sequencing data were evaluated. Using the SNP arrays calls generated by OncoSNP as a reference the degrees concordance of calls were evaluated with calls generated by OncoSNP-SEQ. The degree of concordance ranged from 68 to 94% [median 74%]. Cazier et al evaluated the degree of concordance between the two methods of calling copy number information achieving a median of ~76% between the two methods in a subset of primary bladder cancers.

5.12 Structural Variation

Several approaches were employed to analyse structural variants. A genome-wide screen of chromosomal aberrations identified chromothripsis-like changes in L6MET

restricted to chromosome 14q arm (Figure 5.15). Chromothripsis-like alterations were defined as 10 ≥ alterations in copy number and/or LOH transitions involving ≤3 arms. The TCGA identified chromothripsis in 6% of primary tumours when they screened 93 primary lung adenocarcinomas using low-pass whole genome sequencing (Cancer Genome Atlas Research, 2014).

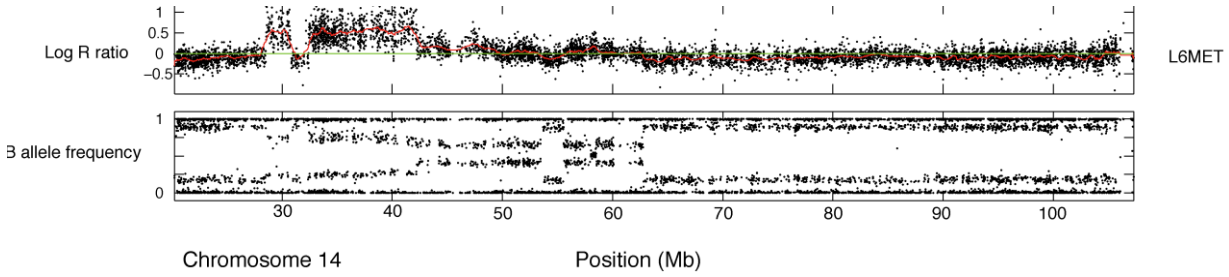


Figure 5.15 I Chromothripsis-like alteration. L6MET q arm chromosome14

BreakDancer was utilised using restrictive filtering for read depth and quality to search for interchromosomal and intrachromsomal translocations. Searches focused on breakpoint regions to generate a list of candidate regions. This systematic approach failed to identify candidate loci. The search was narrowed further to identify genomic rearrangements in *EML4-ALK*, *ROS1* and *RET*. Structural variants in the aforementioned genes were not identified, though this may reflect the low frequency of these variants in primary tumours (Bos et al., 2013, Thunnissen et al., 2012). Intermutation distances of somatic variants were analysed in chromosomal order to identify localised clusters of hypermutations. This screen confirmed the absence of kataegis events in all the LABM including L1MET and L8MET which possess an APOBEC related mutation signature.

5.13 Technical validation of sequencing data

To validate the somatic variants identified through the sequencing analysis pipeline, an alternative sequencing platform was utilised. A custom panel of 82 cancer related genes was used to validate somatic variants using the Ion Torrent PGM.

Tumour	Gene	Mutation	Type of mutation	Validated
L1MET	*BRCA2*	exon18:c.G8022C:p.K2674N	Nonsynonymous	Yes
L1MET	*EGFR*	exon21:c.T2573G:p.L858R	Nonsynonymous	Yes
L2MET	*MET*	exon21:c.C4115A:p.A1372D	Nonsynonymous	Yes
L2MET	*TP53*			Yes
L3MET	*PTEN*	exon5:c.C277A:p.H93N	Nonsynonymous	Yes
L3MET	*TP53*			Yes
L3MET	*TTLL5*	exon20:c.C1955G:p.T652S	Nonsynonymous	Insufficient coverage
L3MET	*RB1*	exon14:c.C1333T:p.R445X	Nonsense	Insufficient coverage
L4MET	*IDH2*	exon1:c.T28C:p.S10P	Nonsynonymous	Yes
L4MET	*STK11*	exon2:c.C334T:p.Q112X	Nonsense	Yes
L4MET	*TP53*			Yes
L5MET	*CDKN2A*	exon2:c.377delC:p.P126fs	Frameshift deletion	Yes
L6MET	*ATRX*	exon9:c.G3137C:p.S1046T	Nonsynonymous	Yes
L6MET	*KRAS*	exon2:c.G34T:p.G12C	Nonsynonymous	Yes
L6MET	*TLR4*	exon4:c.C1418A:p.P473Q	Nonsynonymous	Yes
L6MET	*TNKS2*	exon19:c.A2515G:p.S839G	Nonsynonymous	Yes
L7MET	*CDH1*	exon16:c.G2452C:p.A818P	Nonsynonymous	Yes
L7MET	*ERBB4*	exon28:c.G3824T:p.R1275L	Nonsynonymous	Yes
L7MET	*GNAS*	exon1:c.G181T:p.V61F	Nonsynonymous	Yes
L7MET	*JAK3*	exon6:c.G727T:p.E243X	Nonsense	Yes
L7MET	*KDR*	exon4:c.C464T:p.S155L	Nonsynonymous	Yes
L7MET	*PDGFRA*	exon10:c.G1455T:p.E485D	Nonsynonymous	Yes
L7MET	*SMO*	exon5:c.G1039A:p.G347S	Nonsynonymous	Yes
L7MET	*TET2*	exon3:c.392delG:p.R131fs	Frameshift deletion	Insufficient coverage
L7MET	*TP53*			Yes
L8MET	*EGFR*	exon19:c.2236_2250del:p.746_750del	Nonframeshift deletion	Yes
L8MET	*HDAC1*	exon6:c.G544A:p.G182S	Nonsynonymous	Yes
L8MET	*TP53*			Yes
L1MET	*BRCA2*	exon18:c.G8022C:p.K2674N	Nonsynonymous	Yes

Table 5.5 l Summary of mutations screened in technical validation. The mutations screened using the Ion Torrent cancer panel are represented.

Somatic mutations were identified in all LABM. In total 27 somatic variants were screened, of which 3 variants were non-informative due to insufficient coverage. The remaining 24 variants were validated with 100% concordance (including 2 indels).

5.14 Comparison of LABM with primary tumours

To determine the degree of heterogeneity between primary lung adenocarcinomas and LABM I utilised published findings and TCGA data to serve as a comparison.

5.14.1 Comparison of mutation burdens with primary lung adenocarcinoma tumours

Imeilinski et al performed whole exome (n=159) and whole genome (n=24) sequencing on a large cohort of primary lung adenocarcinoma tumours (Imielinski et al., 2012). They reported a wide range in mutation frequencies in protein-coding regions, median 6.8 per 10^6 [range 0.3 to 94.7] and 13.3 per 10^6 [4.5 to 55.3] from whole exome and genome sequencing respectively. The majority of the tumours sequenced (70%) were classified as stage I or II.

More recently, TCGA published their findings from whole exome sequencing of a large cohort of 230 untreated primary lung adenocarcinomas (Cancer Genome Atlas Research, 2014). They reported a median of 5.8 per 10^6, mean 8.9 per 10^6 [range 0.5 to 48]. In comparison, I report a median of 5.2 per 10^6 [range 2.6 to 32 per 10^6] in the sequenced brain metastases. Taken together, these findings indicate that the LABM have a mutation frequency that is comparable to that observed in primary tumours.

5.14.2 Comparison of mutation spectra with primary lung adenocarcinoma tumours

The mutation spectra in primary lung adenocarcinomas were analysed using a subset of tumours from the TCGA (n=300). The TCGA lung adenocarcinoma samples were sequenced using whole exome sequencing. The mutation spectra of the LABM and in the whole genome and exome are depicted (Figure 5.4). The mutation spectra of LABM closely resembled that of primary lung adenocarcinomas with C>A/G>T and C>T/G>A mutations forming the most common base substitutions observed.

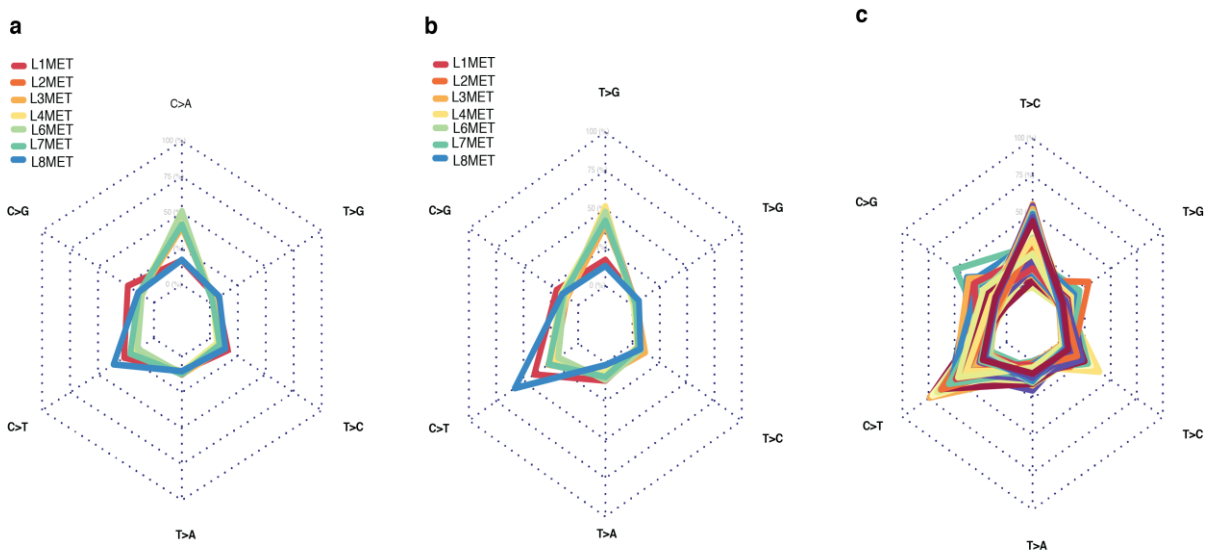


Figure 5.16 I Mutation spectra in primary and metastatic lung adenocarcinomas a) whole genome brain metastases b) exome brain metastases c) TCGA primary tumours

5.14.3 Comparison of copy number alterations in primary lung adenocarcinoma tumours and LABM

Somatic copy number variations in the LABM were compared to data from primary tumours obtained from the TCGA using SNP arrays processed with OncoSNP (section 2.16). The LABM had a median number of 48 CNA (range 15-64). In comparison, the primary lung adenocarcinomas (n=304) had a median number of CNA of 21 (range 0-115). There was a statistically significant difference between the median number of CNA in the lung primaries and LABM ($p<0.01$, Mann-Whitney U test). Of the primary lung adenocarcinomas, 95% of them had CNA with only 11% of primary tumours having >50 CNA.

A comparison of the chromosomal distribution of gains and losses in primary tumours and LABM is depicted in Figure 5.17. Chromosome-scale aberrations were identified in each tumour (arbitrarily defined as >90% involvement of the arm).

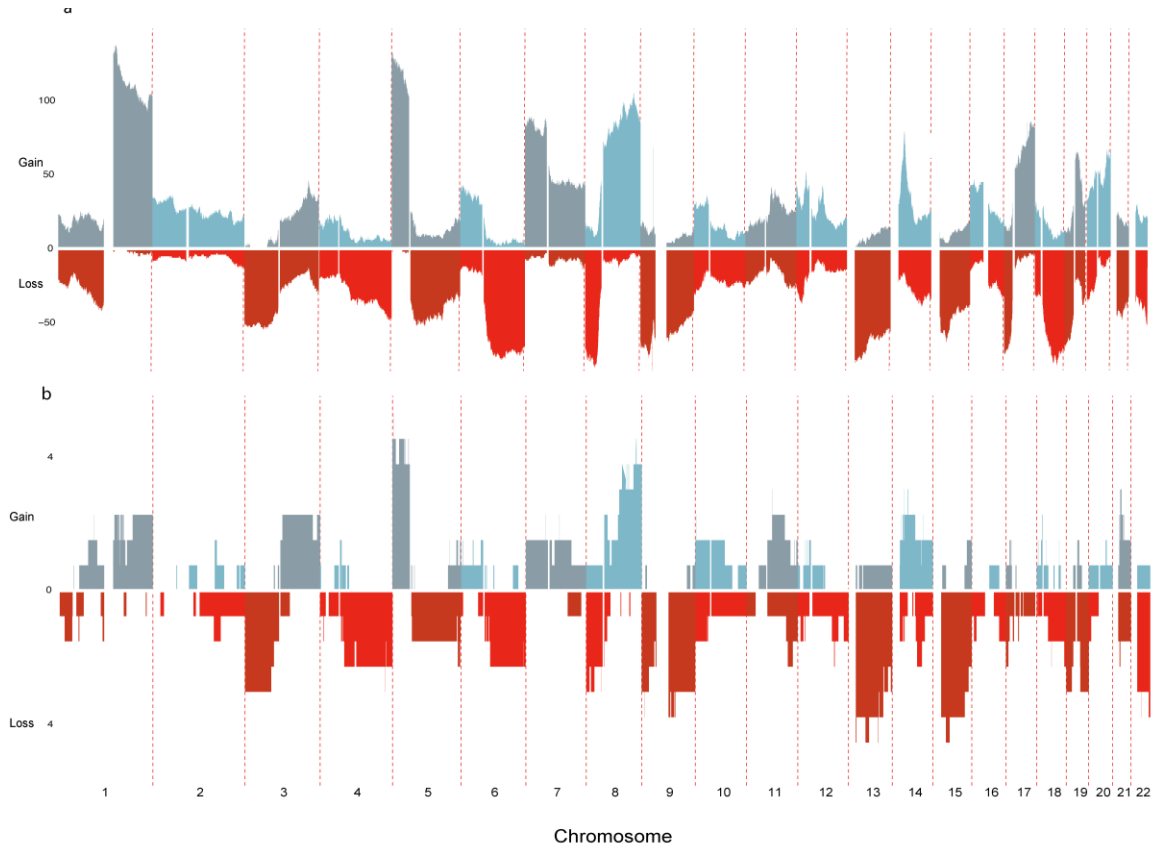


Figure 5.17 l Copy number variations in lung adenocarcinomas a) TCGA primary tumours b) LABM. Copy number deviations relative to the mean copy number are depicted for primary lung adenocarcinomas and brain metastases. Chromosome-scale changes in the primaries involved deletions in 3p, 17p, 19p, 6q, 9q, 13q and gains in 5p, 7p, 16p, 1q, 8q and 17q.

Substantial chromosome-scale changes in the primary tumours were defined as a frequency >10%. Deletions were observed in 3p (n=31), 17p (n-41), 19p (n=42), 6q (n=41) 9q (n=36) 13q n=36. Gains were detected in 5p (n=63), 7p (n=56), 16p (n=31), 1q (n=67), 8q (n=33) and 17q (n=37). To identify statistically significant alterations in the metastases when compared to primary tumours, Fisher exact test was undertaken. Due to the modest sample size, analysis was restricted to arms with a minimum of 40% chromosome-scale changes in the metastases. Deletions in 8p, 15q in the LABM fulfilled these criteria ($p<0.05$).

Analysis of copy number variation in brain metastases of adenocarcinoma are limited in the literature and have often involved unmatched samples and historically involved

the collective analysis of different subtypes of lung cancer (Shiseki et al., 1994, Takahashi et al., 2007). Shiseki et al reported shared allelic losses in 3p, 13q and 17p in the brain metastasis and primary tumours, which were observed in my results. They also reported a higher frequency of losses in 2q, 18q and 22q more frequently the metastases, though these did not reach statistical significance in my analysis.

5.15 Discussion

5.15.1 Mutation burdens and spectra

Although the sample size was modest analysis of the unmatched LABM demonstrated that the mutation burdens in the brain metastases were of comparable magnitude to primary tumours. Interestingly, the mutation frequencies in L1MET and L8MET brain metastases which harbour established activating *EGFR* mutations were the lowest amongst the brain metastases. This parallels observations reported in primary tumours (Imielinski et al., 2012, Ding et al., 2008). *EGFR* is a transmembrane protein with cytoplasmic kinase activity and has a key role in transducing growth signals (Santos et al., 2011). Intracellular signalling leads to increased cellular proliferation, angiogenesis and reduced apoptosis. Activating mutations result in a conformational change in the tyrosine kinase domain of the EGFR, leading to constitutive kinase activity and activation of the downstream signalling cascades. Activating *EGFR* mutations are more common in light/non-smokers, females and of oriental ethnic origin. L1MET and L8MET were in a female Caucasian and male Caucasian patients respectively. Although the smoking history was unavailable, unlike the remaining brain metastases the mutation spectra in these 2 metastases did not have a predominance of C>A/G>T mutations, which correlates with tobacco smoking. The observations in mutation frequencies may be also related to smoking as several authors have reported higher mutation frequencies in smokers

compared to non-smokers in primary tumours (Imielinski et al., 2012, Ding et al., 2008).

5.15.2 Driver mutations and clonal architecture

No driver mutations of LABM were identified in this analysis. Driver mutations in primary lung adenocarcinomas were used to screen the LABM to determine if they were retained in LABM. Historically the studies screening driver mutations have focused on *EGFR* mutations and *KRAS* mutations based on small-scale targeted approaches (Eichler et al., 2010, Munfus-McCray et al., 2011). I identified mutations in *EGFR, TP53*, *and PTEN*. *KEAP* and *STK11* and inferred their position in the clonal architecture of the tumour. Driver mutations were often present in major clones. Interestingly *EGFR* was found to be present in major clones and the brain metastases harbouring (L1MET and L8MET) them had fewer clonal populations and somatic mutations.

5.15.3 Evaluation of copy number alterations

There was statistically significant higher median number CNA in the LABM compared to primary tumours (Man Whitney U test, $p< 0.05$). A comparison of chromosome-scale aberrations in primary tumours and brain metastases was undertaken. This confirmed statistically significant deletions in 8p, 15q in the brain metastases when compared to primary tumours (Fisher exact $p<0.05$).

LABM have previously never been characterised using NGS. In summary, in comparison to the TCGA and published findings, the genomic landscape of mutation burdens and spectra in LABM are comparable to primary tumours. However, there was an increase in CNA in the LABM in comparison to primary tumours. Analysis of structural variants identified chromothripsis-like alterations in one of the 7 LABM

(L6MET), whereas regions of kataegis and rearrangements in *EML4-ALK*, *ROS1* and *RET* were not detected.

Chapter 6

Analysis of Breast Ductal Carcinoma Brain Metastases genomes

6.1 Introduction

Breast cancer forms a collective group of heterogeneous tumours (Polyak, 2011). Breast tumours are a diverse entity and may be stratified based on histopathological evaluation, hormone receptor status and molecular classification (Stingl and Caldas, 2007). Gene expression profiling of breast tumours has led to the subdivision of breast tumours into several categories, chiefly, luminal A, luminal B, basal, HER2 enriched and normal-like breast cancers. The morphological and molecular diversity that exists in breast tumours accounts for the varying prognostic outcomes and the development of tailored therapeutic options.

Following lung cancer, breast cancer constitutes the second most common tumour type that develops brain metastasis (Gil-Gil et al., 2014). It has been estimated that 10-30% of patients with breast cancer develop brain metastases (Weil et al., 2005). Factors that have been associated with an increased likelihood of developing brain metastasis include, tumour grade, nodal status, a negative oestrogen receptor status, HER2 positive disease, triple negative disease (absence of hormonal receptors and non-amplification of HER2) and presentation at a young age (Cheng and Hung, 2007).

TCGA published a large-scale analysis of primary breast tumours using WES. (Koboldt et al., 2012). However, sequencing studies involving brain metastases of primary breast origin are sparse and have been limited to case studies (Shah et al., 2009, Ding et al., 2010). In this chapter, I have reported my findings from the

characterisation of 7 unmatched genomes of breast ductal carcinoma brain metastases (BDCBM). I have utilised data from TCGA from a cohort of primary breast tumours to serve as a comparison data set.

6.2 Methods

6.2.1 WGS of BDCBM genomes

7 BDCBM and paired buffy coat lymphocytes were sequenced The applications and parameters utilised for data processing are presented in section 2.14.

6.3 Comparison Datasets

6.3.1 Somatic mutations from TCGA primary breast tumours

Primary breast tumours from TCGA cBioPortal (section 2.2.1) was used to characterise the mutation spectra and identify candidate driver mutations. The BDCBM were genotyped utilising SNP arrays to elicit information regarding copy number. Raw SNP array data from primary breast tumours obtained from TCGA was used for comparison (section 2.16).

6.4 Patient Demographics

Clinical information where available, for patients who had their breast brain metastases sequenced is reported in Table 6.1.

Tumour	Age of patient (years)	Gender	Histology of primary tumour	Hormone receptor/ HER2 status of primary tumour/met	Time interval between resection of metastasis and primary tumour (Years)	Treatments received by patient
B1MET	66	F	Ductal carcinoma	Primary receptor status unknown Met ER+ Prneg HER2 unknown	1	WLE for left breast cancer and adjuvant chemotherapy 2008.
B2MET	63	F	Ductal carcinoma	Primary ER+ PR+ HER2 neg Met ER+ PR+ HER2 neg	4	Presented with metastatic breast cancer bone involvement in 2005
B3MET	69	F	Ductal carcinoma	Primary ER+ PR+ HER2 neg Met ER+ PR+ HER2 neg	9	**Diagnosis 1:** Metastatic ovarian cancer in 2003: completed 6 cycles of carboplatin and paclitaxel. **Diagnosis 2 :** Ductal carcinoma of breast - axillary clearance and lumpectomy in 1996 treated with CMF chemo and radiotherapy followed by tamoxifen. In 2005 had a right breast recurrence **Diagnosis3:**Parathyroid adenoma parathyroidectomy in 2005 **Diagnosis 4:** Pleomorphic sarcoma 2008 received doxorubicin.
B4MET	51	F	Ductal carcinoma	Primary ER+ PR+ HER2 unknown Met ER+ PR+ HER2 +	7	Primary breast tumour in 2003, received adjuvant chemotherapy and radiotherapy Developed bone and liver metastases in 2007.
B5MET	43	F	Ductal carcinoma	Primary ER+ PR + HER2 + Met ER+ PR unknown HER2 +	3	Presented with metastatic lung disease in 2008. Received docetaxel plus herceptin. Developed bone metastasis in 2010 commenced letrozole
B6MET	46	F	Ductal carcinoma	Primary unknown Met ER+ PR - HER2 unknown	3	Adjuvant FEC–T and herceptin and adjuvant radiotherapy.
B7MET	55	F	Ductal carcinoma	Primary ER- PR- HER2 unknown Met ER- PR- HER2 -	8	Mastectomy and adjuvant chemoradiotherapy

Table 6.1 l Summary of clinical information for patients who had their breast brain metastases sequenced.

Results

6.5 Mutation Burdens

6.5.1 Mutation burdens of BDCBM genomes

The total number of mutations in the BDCBM ranged from 4,034 to 63,896 [median 12,157]. Genome wide the number of somatic SNVs ranged from 5,656 to 61,285. In comparison, the number of small indels ranged from 1,034 to 3,676 [median 1,980]. B7MET had the greatest total mutation burden (63,896) with the highest number of SNVs 61,285 (Figure 6.1). Despite having the lowest SNV mutation burden proportionately small indels were greatest in B2MET across the genome (29%, range 4 to 29%).

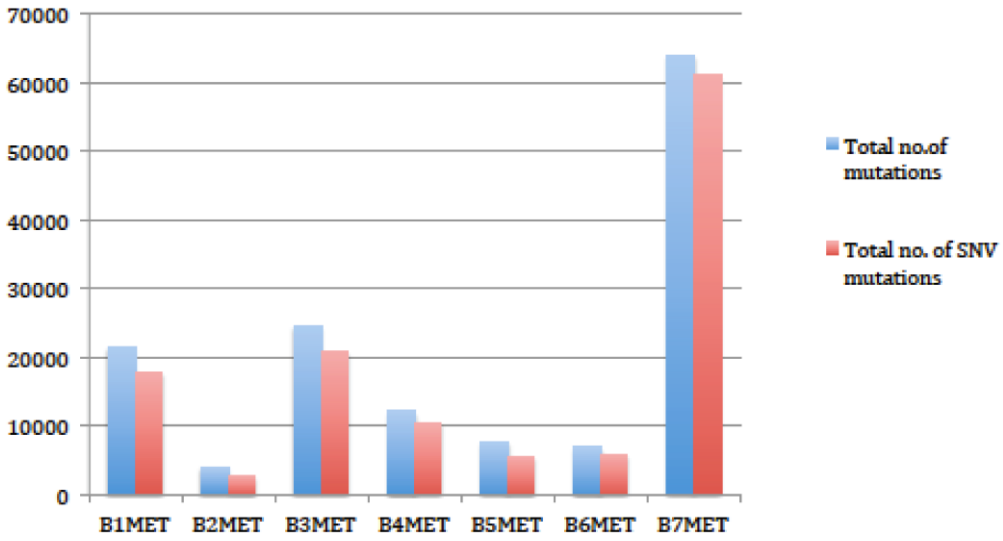


Figure 6.1 l Mutation burdens in BDCBM in the whole genome. The total number of mutations, SNV mutations and the number of small indels are depicted.

SNVs in the genome were categorised into intergenic, intronic and exonic locations. Proportionately the number of SNVs in intergenic regions ranged from 60 to 70%, whereas it ranged from 29 to 39% in intronic regions. The mutation burdens in the exome was ~1% of the overall mutation burden, ranging from 1.3 to 28 per 10^6 (Figure 6.2).

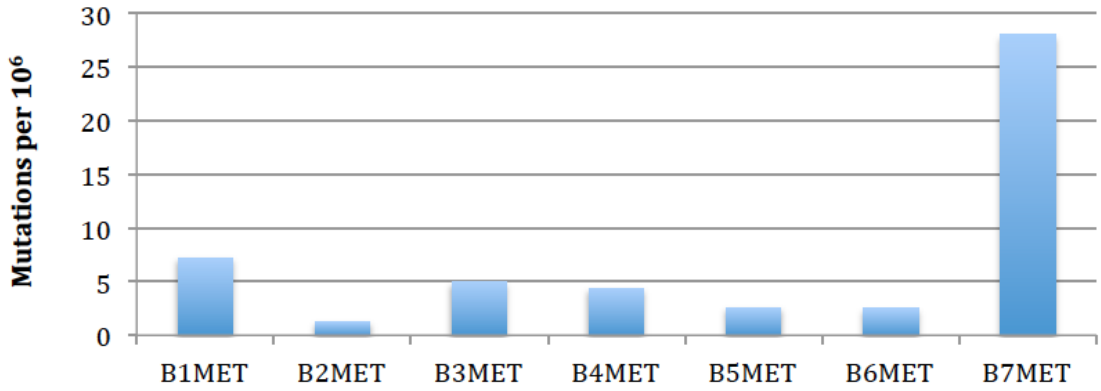


Figure 6.2 I Number of mutations in exome in BDCBM. Mutations are represented per megabase. B7MET has the greatest mutation burden.

Synonymous mutations formed 21 to 36% [median 24%] of the mutation burdens in the exome. In total a median of 74 [range 25 to 616] were potentially functional mutations within protein-coding regions. Proportionately nonsynonymous mutations formed 47 to 65% [median 59%] of the mutation burdens in the exome. The number of nonsense mutations ranged from 0 to 58 events per tumour [median 4] (Figure 6.3). Nonsense mutations were not identified in B2MET. The number of small indels ranged from 4 to 44 [median 14] events per tumour.

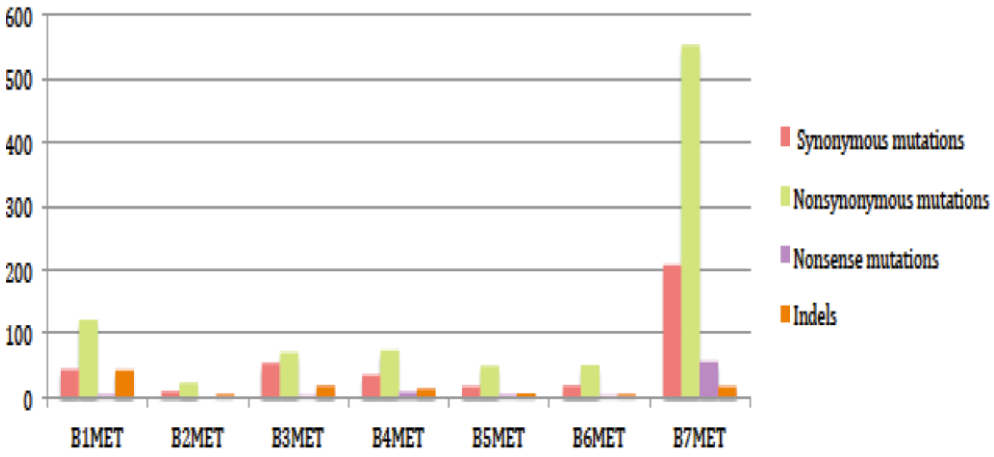


Figure 6.3 I Distribution of types of protein-coding mutations in the BDCBM. B7MET has the greatest mutation burden amongst the BDCBM.

The mean nonsynonymous:synonymous mutation ratio was 2.4:1. The transition/transversion ratios ranged from 0.6 to 1.6. Analysis of the mutation burdens confirmed that the mutation burden in B7MET is greater than in the other BDCBM. Although the time interval between the resection of the primary and presentation of the metastasis was 8 years this was 15 and 7 in B3MET and B4MET respectively which have considerably smaller mutation burdens

6.6 Analysis of mutation spectra

Generally C>T/G>A and C>G/G>C mutations formed the most common type of base substitutions in the BDCBM in the whole genome. The proportion of C>T/G>A mutations in the genome ranged from 21 to 47% [median 41%] whereas C>G/G>C mutations ranged from 6 to 44% [median 22%] (Figure 6.4). C>G/G>C transversions were more common in the brain metastases B4MET, B5MET, B6MET and B7MET.

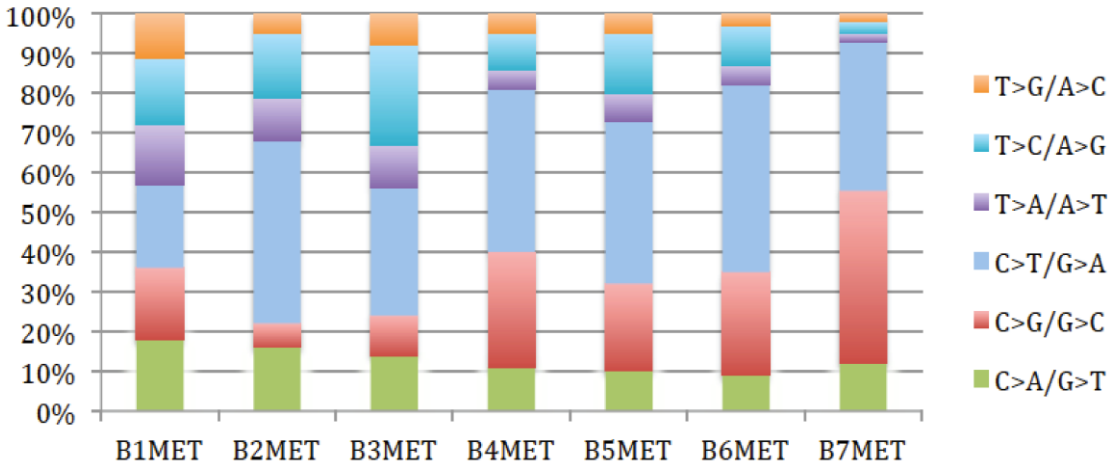


Figure 6.4 l Distribution of the 6 classes of base substitutions in the BDCBM across the genome. C>T/G>A and C>G/G>C mutations predominate.

Analysis of mutations within the exome confirmed that the C>T/G>A [range 28 to 64%] and C>G/G>C [range 3 to 42%] predominated. Sjoblom et al undertook a systematic analysis of the mutation footprints in protein-coding genes in 11 primary breast tumours. They reported that C>T/G>A and C>G/G>C mutations predominated at 35% and 29% of mutations respectively (Sjoblom et al., 2006).

6.7 Analysis of mutation signatures

Base substitutions can be analysed further in sequence context taking into account the 5' and 3' nucleotide immediately flanking the mutated base. Tumours are exposed to several biological processes which leave their own unique footprints. Using somatic signature the relative contribution of mutation footprints were analysed to generate mutation signatures (Figure 6.5).

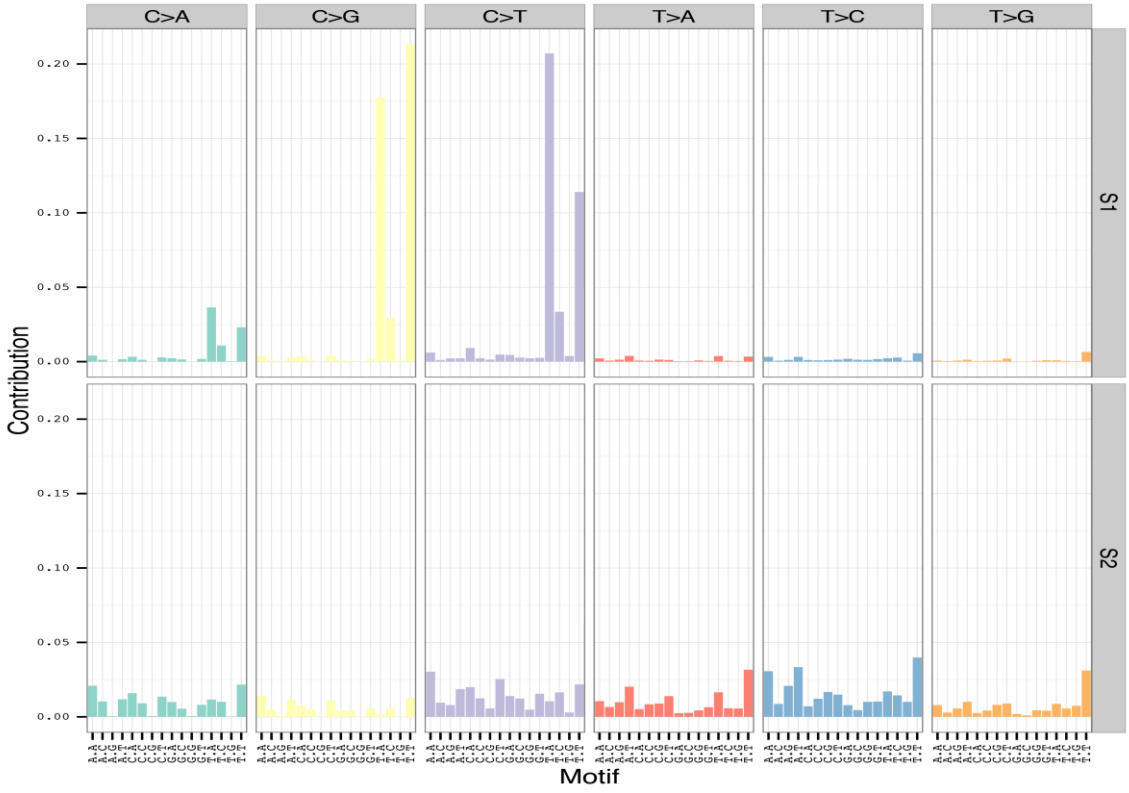


Figure 6.5 I Mutation signatures in BDCBM in sequence context. 2 signatures were generated S1 and S2. The metastases B4MET, B5MET, B6MET and B7MET resembled signature S1.

In my analysis, B4MET, B5MET, B6MET and B7MET corresponded to signature S1 whereas B1MET, B2MET and B3MET resembled signature S2. In signature S1, C>G and C>T base substitutions were overrepresented at T[C]A and T[C]T sites. A study of mutation signatures in 21 breast cancers identified C>T, C>G and C>A base substitutions occurring in the context of a T[C]X motif (where X denotes an adenine or

thymine base) which was termed signature B (Nik-Zainal et al., 2012). Alexandrov et al proceeded to characterise mutation signatures in a large pan-cancer analysis (Alexandrov et al., 2013). The S1 signature has been associated with the APOBEC family of cytidine deaminases and resembled signature [2] in the analysis conducted by Alexandrov et al. Gene expression of APOBEC3B enzyme correlated with this typical mutation footprint (Burns et al., 2013). The APOBEC associated signature has been associated with numerous cancers including breast, lung, cervix, bladder and head and neck cancers (Kuong and Loeb, 2013). Furthermore, this signature has been associated with localised clusters of hypermutation, termed kataegis. Signature S2 in my analysis did not closely resemble the signatures reported by Alexandrov et al (Alexandrov et al., 2013).

6.8 Screen for germline mutations in BDCBM

It has been estimated that breast cancer has a hereditary component in approximately 5-10% of cases (Larsen et al., 2014). Of patients with a familial component, *BRCA1* and *BRCA2* mutations are the most prevalent (Melchor and Benitez, 2013). A screen for germline variants for pathogenic germline *BRCA1* and *BRCA2* was negative. The brain metastases B3MET originated from a patient who had developed numerous different types of cancer during their lifetime including, ductal carcinoma of the breast, ovarian carcinoma, a pleomorphic sarcoma and a parathyroid adenoma. Given the clinical history of the patient, I screened for germline mutations in *TP53*, which has been associated with Li Fraumeni and Li-Fraumeni-like syndromes in 70% and 40% of cases respectively (McBride et al., 2014). These two syndromes have been associated with a predisposition to developing a variety cancers in a lifetime with an early age of onset. Germline mutations in *TP53* were absent in this patient.

6.9 Identification of putative drivers of breast brain metastases

Two approaches were used to endeavour to identify driver mutations in the BDCBM. Firstly, IntOGen was utilised to identify putative drivers in the BDCBM (FDR <0.05). Secondly, the analysis was extended to review genes with ≤3 recurrent truncating mutations. No candidate drivers of brain metastases were identified in either process.

6.10 Screen for primary driver mutations in BDCBM

Driver genes in primary breast tumours were identified by running IntOGen and Mutsig on a large cohort of primary tumours obtained from TCGA cBioPortal. Candidate driver genes were used to screen the BDCBM. Mutations with predicted benign and moderate functional effects were excluded (SIFT <0.2, PolyPhen2 >0.8). Identified driver genes in the BDCBM are reported in Table 6.2. Amongst the BDCBM no driver genes were identified in B1MET and B2MET. However, mutations in *TP53, FOXA1*, *NF1* and *ARID1A* were identified amongst the other BDCBM. TCGA reported a comprehensive analysis of WES of 510 primary breast cancers. They reported a large heterogeneity in the somatic mutations amongst breast cancers. Somatic mutations in only 3 genes occurred in more than 10% of all the breast tumours. These genes included *TP53*, *PIK3CA* and *GATA3*. *PIK3CA* and *GATA3* mutations were absent in BDCBM.

Recently *ARID1A* has been identified as a candidate tumour suppressor gene in several cancers (Wu and Roberts, 2013, Jones et al., 2012). *ARID1A* is a member of the ARID family that has highly conserved domains that can interact with DNA including an AT rich region regulating chromatin remodelling and cell cycle control. Nonsense mutations in *ARID1A* were identified in B3MET [Q1993X] and B7MET [S138X]. Somatic mutations in *TP53* and *ARIDA1* encompassed regions of LOH, which is consistent with inactivation of tumour suppressor genes.

Gene	Sample	Location of mutation	Type of mutation	LOH at locus of mutation	Copy Number at locus of mutation	PolyPhen 2 Score	SIFT Score	Mutation frequency in gene-TCGA cBioPortal %	Mutation frequency in gene at that loci –TCGA cBioPortal
TP53	B5MET	exon6:c.581 A>C:p.L194R	Nonsynonymous	Yes	2	1	0	36% (181/482)	2
	B6MET	exon6:c.584 A>Gp.I95T	Nonsynonymous	Yes	2	1	0		4
NF1	B4MET	exon48c.7203_7204del:p.2401_2402del AC-	Frameshift deletion	No	3	--------	--------	7% (14/182)	0
FOXA1	B4MET	exon2:c.798 C>T:p.F266L	Nonsynonymous	No	3	1	0	2% (4/182)	0
ARID1A	B3MET	exon20:c.5977 C>T:p.Q1993X	Nonsense	Yes	1	--------	--------	3% (26/962)	0
	B7MET	exon1:c.413 C>G:p.S138X	Nonsense	Yes	3	--------	--------	3% (26/962)	0

Table 6.2 I Summary of identified somatic mutations in primary driver genes in the BDCBM

6.11 Clonal architecture of BDCBM

The clonal architecture of the BDCBM were characterised using Pyclone. The number of clonal populations reported in each brain metastases is listed in Table 6.3 and ranges from 2-23 [median 6]. Interestingly, the majority of the BDCBM were clonally simple with less than 10 clonal groups. B1MET had the most complex clonal architecture with 23 clonal groups. Despite, B7MET having the greatest mutation burden amongst the BDCBM the clonal structure of the metastasis consisted of 12 clonal groups. B4MET had the least complex structure with 2 clonal groups. There was no correlation between the complexity of the clonal architecture (measured by the number of clones) and the number of somatic SNVs (Spearman ρ= 0.41 p= 0.24).

Breast brain metastasis	Number of clones	Number of SNVs mutations
B1MET	23	17936
B2MET	4	2884
B3MET	6	20961
B4MET	2	10383
B5MET	10	5656
B6MET	6	5885
B7MET	12	61285

Table 6.3 I Summary of clonal populations in breast brain metastases.

To appreciate the hierarchical position of primary drivers of breast tumours in the BDCBM the clonal structures were analysed. The clonal structures of the metastases are depicted in Figures 6.6-6.12. Driver genes of primary breast tumours are plotted in blue alongside randomly selected genes.

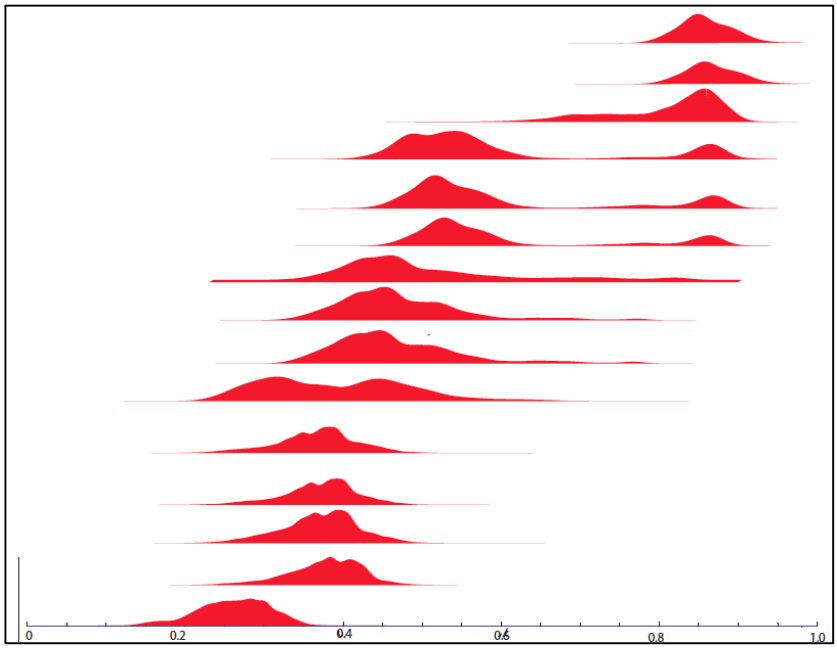


Figure 6.6 I Clonal architecture in B1MET. Somatic mutations randomly selected are depicted to characterise the clonal architecture of B1MET. No putative drivers of primary breast tumours were identified. B1MET had the most complex clonal architecture amongst the BDCBM with 23 clonal groups.

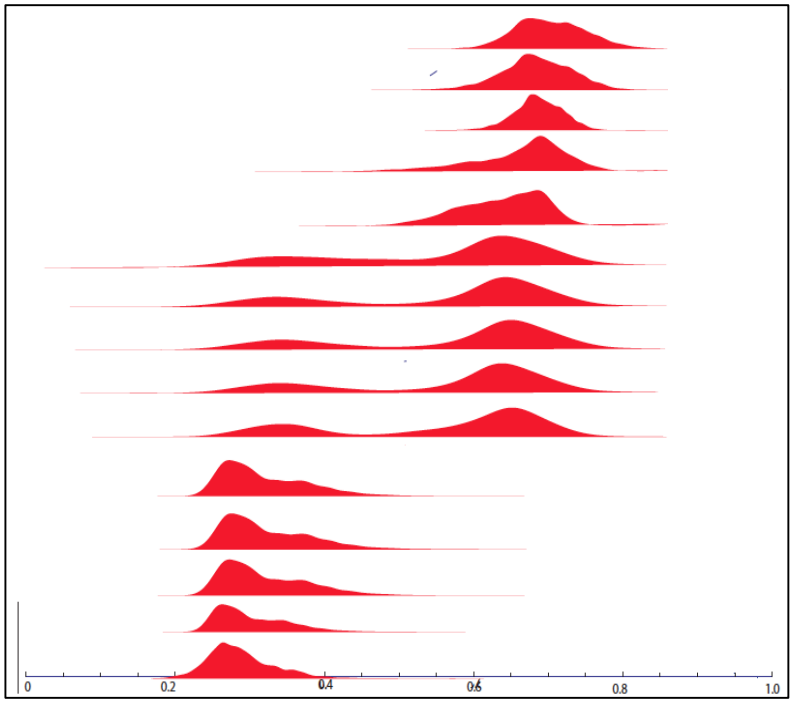


Figure 6.7 I Clonal architecture in B2MET. No putative drivers of primary breast tumours were identified.

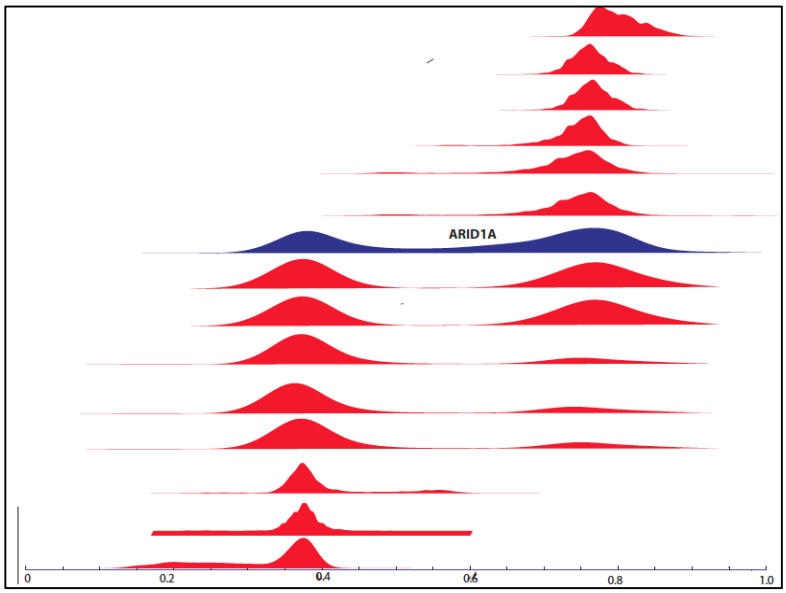


Figure 6.8 l Clonal architecture in B3MET. *ARID1A* a candidate driver in primary breast tumours is located in a dominant clone. Primary breast driver mutations are highlighted in blue.

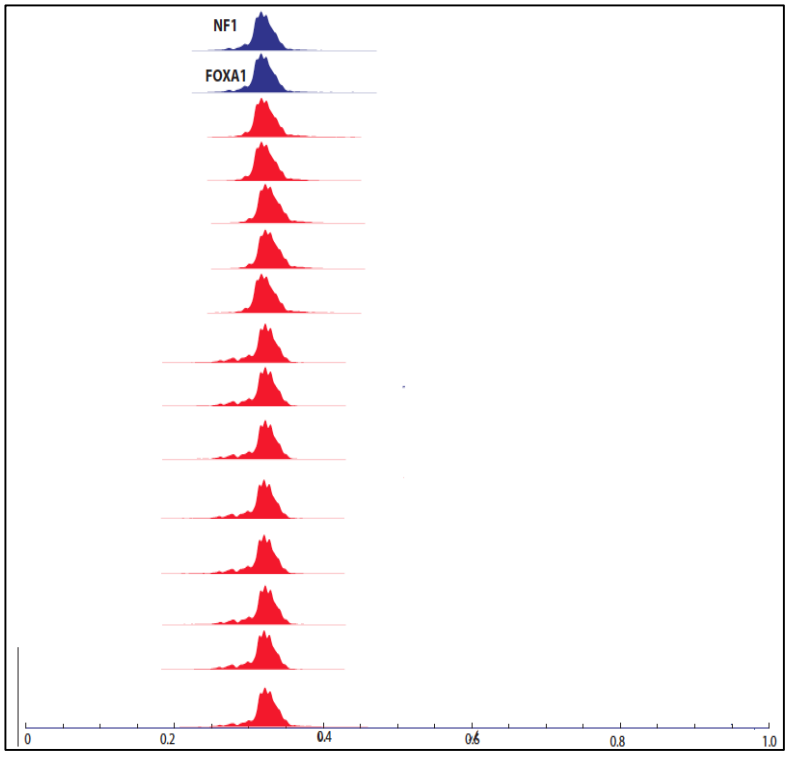


Figure 6.9l Clonal architecture in B4MET. Candidate drivers in primary breast tumours *NF1* and *FOXA1* were identified in B4MET which is clonally simple.

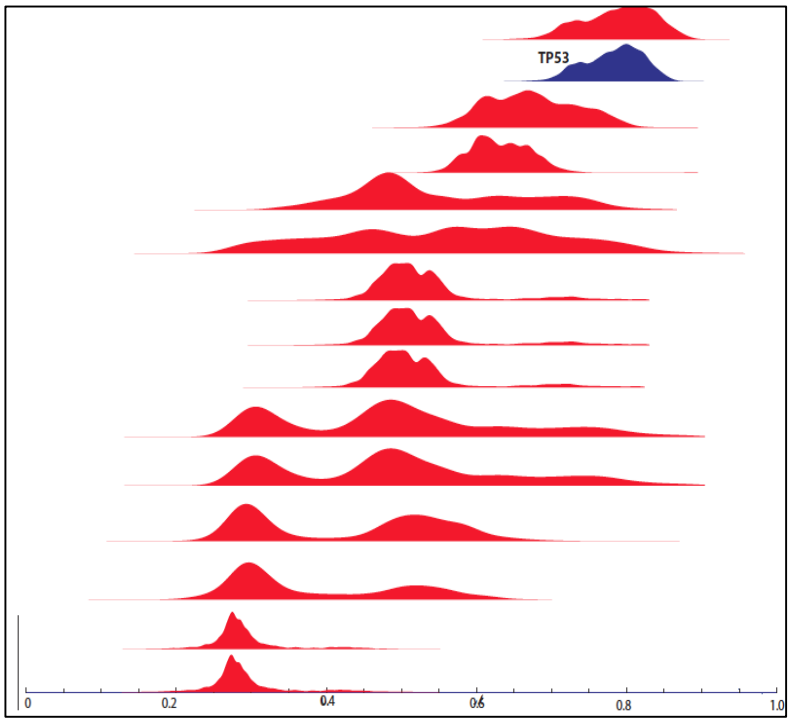


Figure 6.10l Clonal architecture in B5MET. *TP53* is located in a major clone in B5MET encompassing a region of LOH.

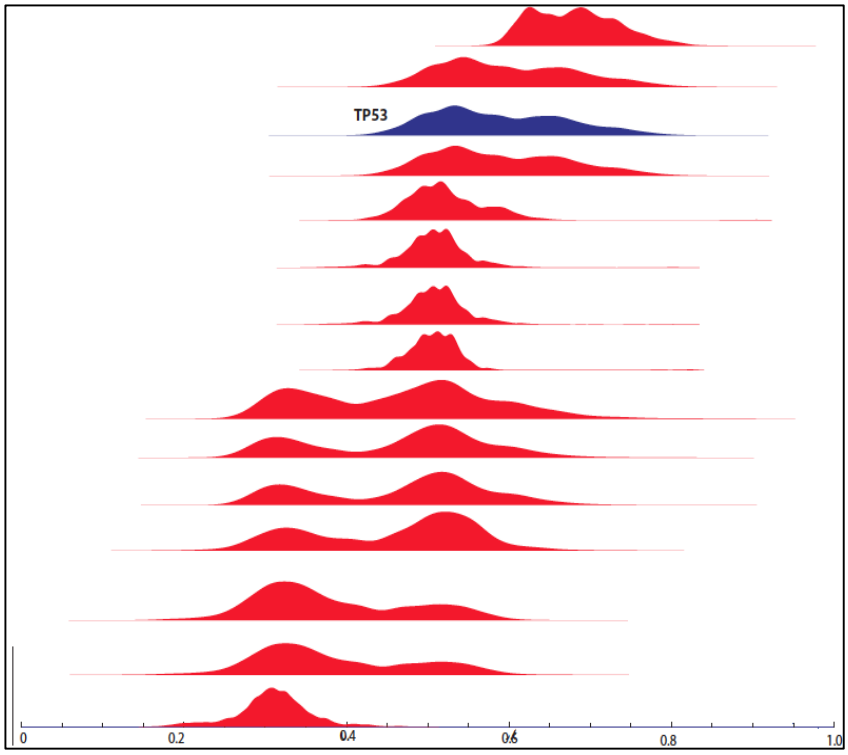


Figure 6.11l Clonal architecture in B6MET. *TP53* is located in a major clonal group in B6MET encompassing a region of LOH.

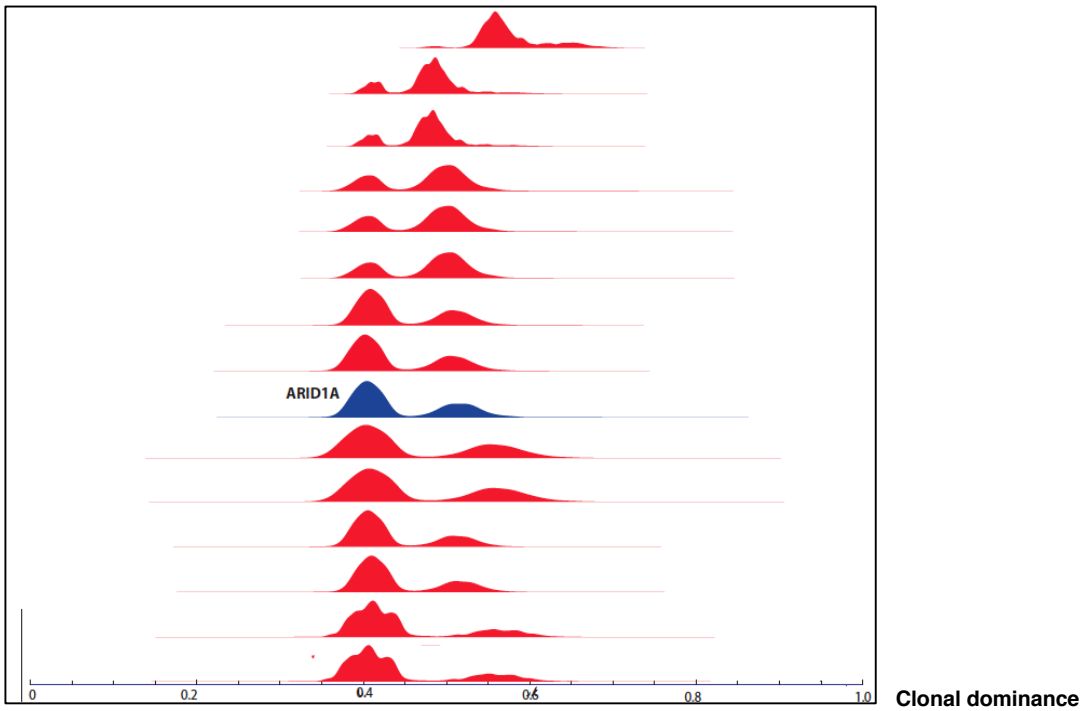


Figure 6.12l Clonal architecture in B7MET. *ARID1A* a candidate driver in primary breast tumours is located in a subclone.

No candidate drivers of primary breast tumours were identified in B1MET and B2MET. Generally, the driver genes of primary breast tumours were located in major clonal groups. Somatic mutations in *NF1* and *FOXA1* mutations were present in B4MET which is clonally simple. *TP53* was located in a major clonal groups in B5MET and B6MET. These metastases originated from primary tumours that are oestrogen receptor and *HER2* positive. Shah et al undertook WGS on a cohort of triple negative primary breast tumours (Shah et al., 2009). They characterised the clonal architecture of these tumours and identified the position of candidate driver mutations within this framework. *TP53* was frequently found in major clonal groups in primary triple negative breast tumours.

6.12 Evaluation of copy number alterations in BDCBM

SNP arrays were undertaken on the BDCBM to characterise copy number aberrations. Table 6.3 summarises characteristics of copy number aberrations amongst the BDCBM. Deviations from the average copy number were identified in each tumour, defined as a

CNA. The BDCBM had a median number of 52 CNA [range 15 to 111]. The proportion of the genome deviating from the mean copy number ranged from 7 to 45% [median 36%].

Tumour	Number of CNA	Proportion of genome deviating from mean copy number	Chromosome-scale deletions	Chromosome-scale gains
B1MET	111	33	---------	19p
B2MET	56	36	chr7, chr10, chr18 17p, 22q	11p,16p, 1q
B3MET	15	7	---------	---------
B4MET	48	36	1p, 19p, 16q, 22q	5p, 6p 20p
B5MET	94	45	chr4, 17p, 9q, 13q, 18q	16p, 20p, 12q
B6MET	42	39	3p,17p 6q 7q	---------
B7MET	70	28	4p,17p, 13q	---------

Table 6.3 l Summary of copy number aberrations characteristics in BDCBM.

Chromosome scale aberrations were identified (defined as >90% involvement of the chromosome arm). Recurrent large-scale aberrations were categorised as those involving 3 or more brain metastases. Deletions fulfilling this in criteria include 17p (B2MET, B5MET, B6MET, B7MET). Whole-chromosome deletions were identified in chromosome 7 (B2MET) and chromosome 4 (B5MET). No BDCBM demonstrated whole-chromosome amplifications.

6.13 Identification of recurrent focal copy number aberrations

A genome wide screen of recurrent focal copy number aberrations (<2MB) detected amplifications in *ERBB2* in B4MET, B5MET, and B6MET. *ERBB2* is also known as

HER2 and encodes a proto-oncogene. Examples of amplification in *ERRB2* are presented in Figure 6.13. *HER2* encodes a tyrosine kinase receptor which belongs to the HER family, of which the EGFR is a member (HER1). Palmieri et al screened archived resected brain metastases and confirmed *HER2* overexpression in 36% of cases (Palmieri et al., 2007). Overexpression of HER2 in primary breast tumours has been reported in 26% of primary tumours (Press et al., 2005). Using a murine model, Palmieri et al overexpressed HER2 in a human breast cancer cell line that selectively metastasises to the brain (Palmieri et al., 2007). Tumour cells overexpressing HER2 and the control transfectants formed comparable numbers of brain metastases. However mice with HER2 overexpression had metastatic outgrowths that were 3 fold greater in size.

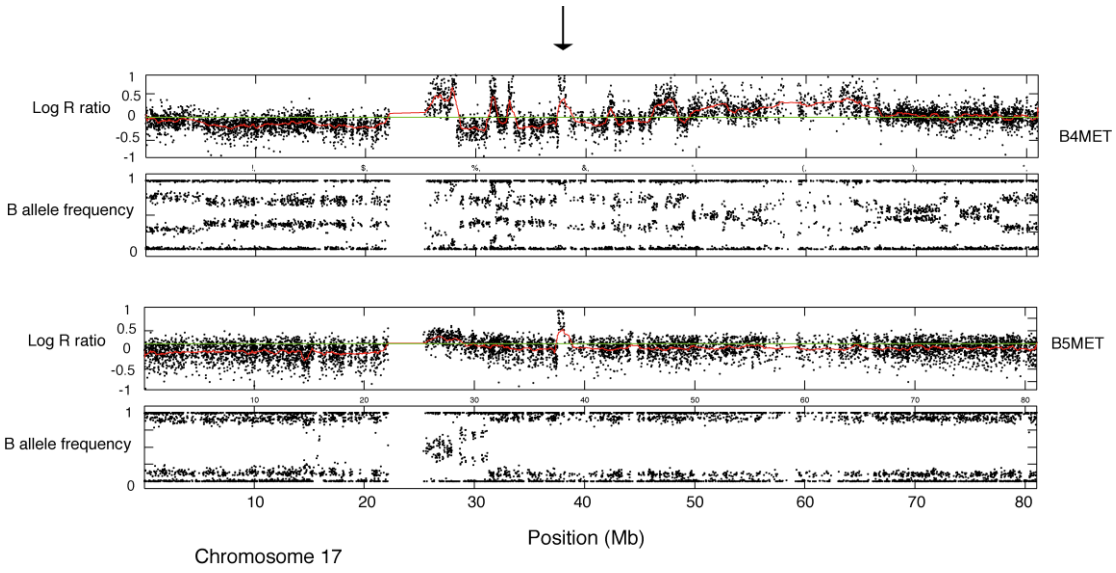


Figure 6.13 I Examples of focal amplifications in HER2 in B4MET and B5MET

6.14 Concordance of copy number information from SNP array and sequencing data

Using OncoSNP-Seq sequencing data was directly utilised to validate findings from SNP arrays. Amplifications in HER2 identified in the SNP arrays were verified. (Figure 6.14).

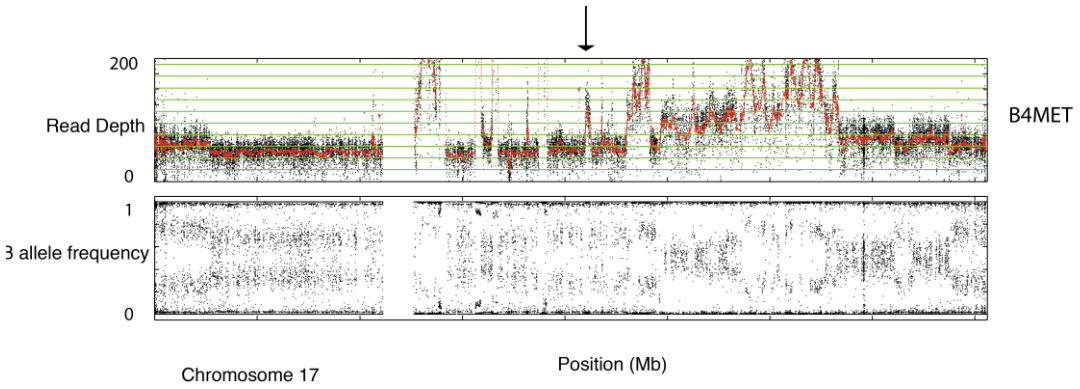


Figure 6.14 I Example of focal amplifications in *HER2* in B5MET identified with OncoSNP-SEQ.

The degree of concordance between copy number calls generated using SNP array information and that directly inferred from sequencing data were assessed. For the purpose of this analysis SNP arrays calls generated by OncoSNP were used a reference. The degree of concordance ranged from 66 to 83% [median 76 %].

6.15 Structural Variation

6.15.1 chromothripsis/chromoplexy

The BDCBM were screened using SNP arrays to detect chromothripsis-like/chromoplexy alterations (Figure 6.15). Chromothripsis and chromoplexy are phenomena which have been identified in recent years resulting in the breakage and fusion of genomic segments of the genome (Zhang et al., 2013). They can result in widespread genomic changes in a narrow time window, and provide new mechanisms to propagate tumour development distinct to the archetypal model of tumour evolution where mutations gradually accumulate through successive rounds of clonal expansion. Chromothripsis typically affects a few chromosomes, with up to hundreds of breakpoint regions. In comparison, chromoplexy typically affects multiple chromosomes ≥5, with fewer genomic rearrangements. Chromothripsis-like alterations (defined as 10 ≥ oscillating alterations in copy number and/or LOH transitions restricted to ≤3 arms) were observed in three of the BDCBM.

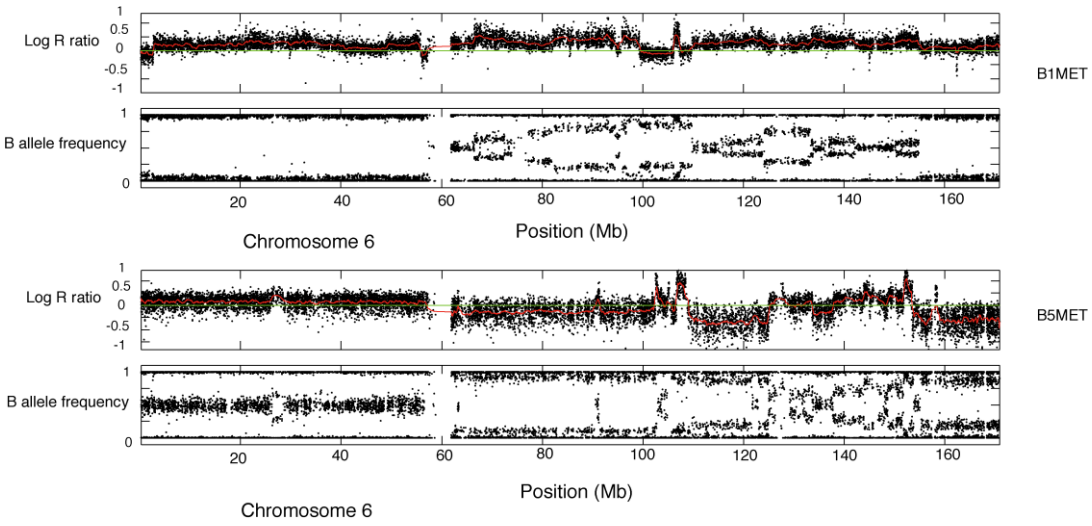


Figure 6.15 I Examples of chromothripsis-like aberrations B1MET and B5MET

In B2MET (chromosome 14q) and B4MET (chromosome 17q) chromothripsis-like alterations were identified restricted to one arm. B5MET demonstrated alterations involving 3 arms (1p, 6q, 11q). In comparison, genomic aberrations consistent with chromoplexy were observed in B1MET (1p, 6q, 7q, 8q, 13q).

BreakDancer was utilised with the aim to identify interchromosomal/intrachromosomal translocations. Candidates were filtered for read depth and quality. The screen was narrowed to focus on breakpoint regions. Through this process reads supporting interchromosomal/intrachromosomal translocations were not identified. However, it is conceivable that reads supporting these translocations may have been filtered out in the mapping process.

6.15.2 Identification of regions of kataegis

Localised regional clusters of mutations termed kataegis were first described in breast cancers and have been associated with genomic rearrangements (Nik-Zainal et al., 2012). I sought to screen for regions of kataegis in the BDCBM using a genome wide screen calculating the intermutation distance for somatic SNVs (Figure 6.16).

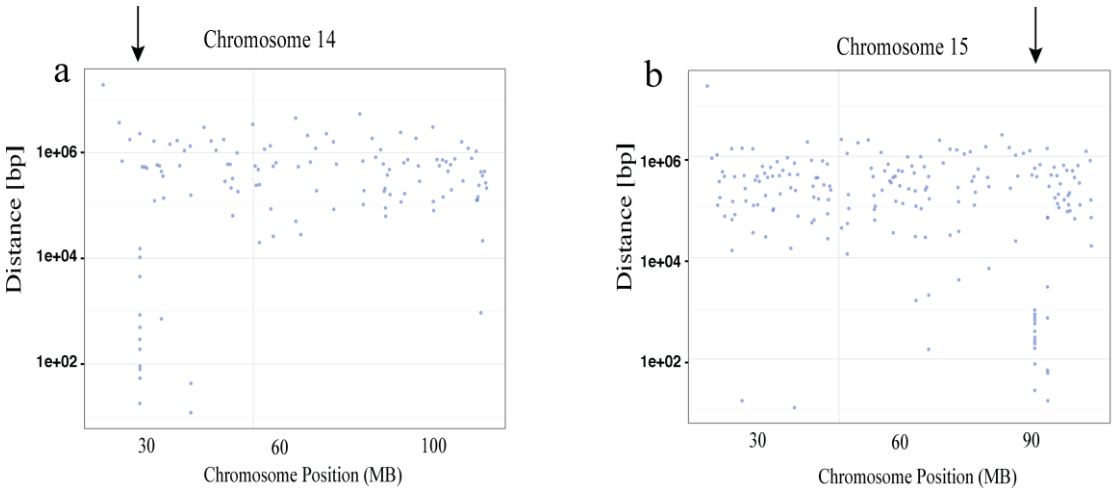


Figure 6.16 l Regions of kataegis identified in chromosome 14 and 15 in B6MET

B6MET contained 2 regions of kataegis located in chromosome 14 and 15 respectively. A cluster of C>T and C>G mutations were found at ~27MB (13 mutations within a 1MB region) in chromosome 14. In chromosome 15 a cluster of generally G>A mutations was located in a region <2MB (10 mutations) at ~90MB. Extension of the analysis to surrounding areas confirmed the absence of the phenomenon of class switching. Class switching occurs when microclusters of mutations of one type i.e. C>T are followed by a switch in the class of mutation such as G>A for several megabases (Nik-Zainal et al., 2012). Structural rearrangements were not identified adjacent to the regions of kataegis. Of note, the mutation signature of B6MET corresponded with the APOBEC signature which has been associated with kataegis (Nik-Zainal et al., 2012, Kuong and Loeb, 2013).

6.16 Technical validation of sequencing data

Somatic variants identified through the sequencing analysis were verified using the Ion Torrent platform (Table 6.4). Validated variants were detected through automated calls on the Ion Torrent server and through manual inspection on IGV. A custom cancer gene panel of 82 genes was utilised. In total 9 variants that were screened were validated with 100% concordance. Of these 2 were frameshift mutations and the remaining were

nonsynonymous. B1MET and B2MET had no somatic mutations that overlapped the custom gene panel.

Tumour	Gene	Mutation	Type of mutation	Validated
B3MET	*BRACA2*	exon 23 c.9090dupA:p.T3030fs	Frameshift insertion	Yes
B4MET	*NF1*	exon 48:c.7203_7204delAC:p.2401_2402 del	Frameshift Deletion	Yes
B5MET	*TP53*	exon6:c.581 A>C:p.L194R	Nonsynonymous	Yes
B6MET	*TP53*	exon6:c.584 A>Gp.I95T	Nonsynonymous	Yes
B7MET	*ABL1*	exon11:c.C2697G:p.H899Q	Nonsynonymous	Yes
B7MET	*ALK*	exon29:c.G4698T:p.M1566I	Nonsynonymous	Yes
B7MET	*FGFR2*	exon10:c.C1367G:p.S456C	Nonsynonymous	Yes
B7MET	*GNAS*	exon10:c.G880C:p.E294Q	Nonsynonymous	Yes
B7MET	*IDH2*	exon8:c.C977G:p.S326C	Nonsynonymous	Yes

Table 6.4 l Summary of somatic variants validated in BDCBM

6.17 Comparison of BDCBM with primary tumours

To determine the degree of heterogeneity between primary breast tumours and BDCBM I utilised published findings and TCGA data to serve as a comparison.

6.17.1 Comparison of mutation burdens with published findings

Brain metastases of a lobular and a basal-like ductal breast cancer subtypes have been characterised using WGS (Shah et al., 2009). Ding et al reported 50 somatic mutations in a basal-like breast cancer brain metastasis with 48 mutations shared with the primary tumour (Ding et al., 2010). Shah et al reported 35 somatic mutations in a lobular breast cancer brain metastasis, of which 19 mutations were undetected in the primary (Shah et al., 2009). In comparison, the mean number of mutations in the BDCBM sequenced in my analysis was 220 [range 23 to 552]. B7MET has a higher mutation frequency in comparison to the

other brain metastases sequenced (61285 vs. [2884-17936] somatic SNV mutations). The number of exonic mutations identified in B7MET was 845. I evaluated the somatic mutation frequency from 1000 primary breast tumours obtained from the TCGA cBioPortal. Of this group 21 primary tumours had a somatic mutation frequency > 700 in protein-coding regions. TCGA comprehensively analysed 510 primary breast tumours of mixed histological subtypes using WES reporting an overall mean number of 60 mutations per tumour in protein-coding regions. Elis et al reported a mutation frequency of 1.62 per 10^6 in endocrine resistant breast tumours that underwent NGS (Ellis et al., 2012). Banerji et al reported a mutation frequency of 1.66 per 10^6 in a cohort of breast tumours of mixed subtypes (Banerji et al., 2012). Taken together, the mutation burdens in the BDCBM are generally comparable in magnitude to primary breast tumours.

6.17.2 Analysis of mutation spectra

Generally C>T/G>A and C>G/G>C mutations formed the most common type of base substitution in the BDCBM. The mutation spectra was characterised in 300 primary breast tumours obtained from the TCGA to characterise the degree of heterogeneity in primary breast tumours (Figure 6.17). The TCGA tumours were processed with WES. In primary breast tumours C>T and C>G mutations predominated with mutations in other base substitutions as occurring less frequently.

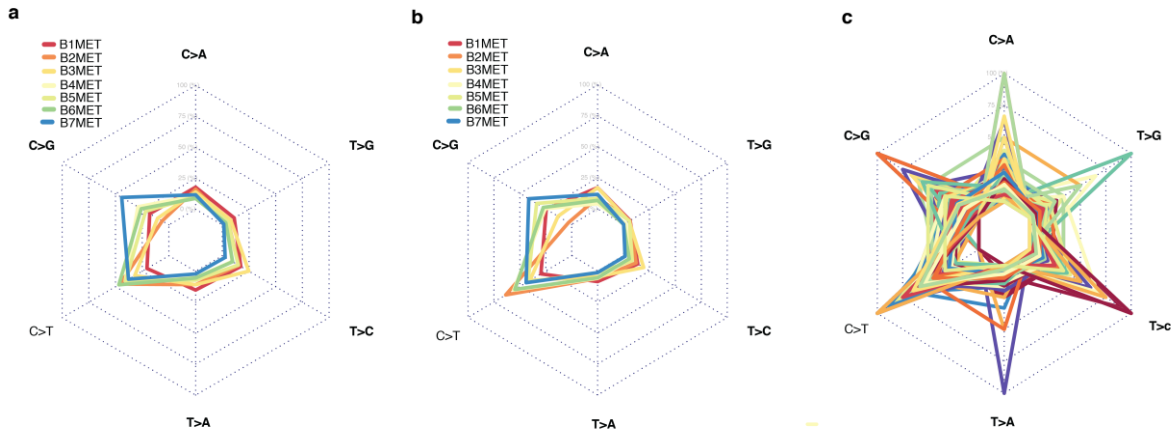


Figure 6.17 I Mutation spectra in primary and metastatic breast adenocarcinomas a) whole genome brain metastases b) exome brain metastases c) TCGA primary tumours.

6.17.3 Evaluation of copy number alterations in BDCBM

SNP arrays were undertaken on the BDCBM which were compared to SNP array data from primary breast tumours obtained from TCGA. Deviations from the average copy number were identified in each tumour, defined as a CNA. Of the primary breast tumours 22% had a CNA > 50. The BDCBM had a statistically significant higher number of CNA compared to primary tumours (Mann Whitney U test, p=0.006). The BDCBM had a median number of 52 CNA [range 15 to 111], whereas the primary breast tumours (n=530) had a median number of CNA of 25 [range 0 to 161]. Although the sample size is modest, these findings suggest that there is increased chromosomal instability in the BDCBM compared to primary breast tumours.

6.18 Discussion

6.18.1 Mutation burdens and signatures

The mutation burdens in the sequenced brain metastases were comparable to that published in primary tumours and brain metastases of breast origin. The mutation signatures of the brain metastases B4MET, B5MET, B6MET and B7MET resembled signature [2] as characterised by Alexandrov et al (Alexandrov et al., 2013). This signature has been associated with the APOBEC family of cytidine deaminases and is characterised by a TCX motif (where X corresponds to either an adenine or thymine and C constitutes the base that is mutated to either a thymine or guanine base) (Roberts et al., 2013). The APOCBEC family of cytidine deaminases serve a role in RNA editing and innate immunity (Smith et al., 2012). The signature has been characterised in several cancers including breast, bladder, cervical, lung and head and neck cancers (Burns et al., 2013). Recent publications have correlated the APOBEC mutation pattern with the presence of kataegis (Roberts et al., 2012, Nik-Zainal et al., 2012). Kataegis events have been associated with genomic rearrangements (Nik-Zainal et al., 2012). Kataegis was identified in B6MET, in chromosomes 14 and 15 which were not associated with genomic rearrangements.

6.18.2 Screen for chromosomal aberrations

A genome wide screen for recurrent focal aberrations identified amplification of *ERRB2* (*HER2*) in 3 of the brain metastases, B4MET, B5MET and B6MET. Traztuzumab a humanised monoclonal antibody, was developed to target the extracellular domain of the HER2. It has been utilised in early stage breast cancer as an adjuvant treatment or in metastatic breast cancer in HER2 amplified disease. The introduction of traztuzumab has led to improved survival rates in HER2 positive disease (Slamon et al., 2001). Despite improved control of visceral disease with traztuzumab, patients receiving traztuzumab developed increasing numbers of brain metastases compared to historical datasets (Stemmler et al., 2006) (Clayton et al., 2004). Pestalozzi et al reported their findings from a large retrospective study of 9524 women who presented early stage breast cancer between 1978-1999 in the pre-traztuzumab era. In patients with HER2 positive primary tumours the 10 year cumulative incidence of CNS relapse at any period was 6.8% versus 3.5% in HER2 negative patients. These findings suggested that HER2 positive breast tumours may have a biological predilection to develop CNS metastases. Studies of the concentrations of traztuzumab in serum and cerebrospinal fluid have confirmed the inefficiency of traztuzumab to penetrate the BBB. A ratio of 420:1 of serum to cerebrospinal fluid levels of traztuzumab have been reported. Moreover, despite radiotherapy which is thought to disrupt the integrity of the BBB the ratio of serum to cerebrospinal fluid level of traztuzumab was 76:1. These observations have led to the concept of the brain acting as a sanctuary site in patients with HER2 positive breast cancer (Lin and Winer, 2007a). The improved control of extracranial disease with traztuzumab may have resulted in increasing numbers of patients developing brain metastases as patients are living longer. Lapatinib, a dual tyrosine kinase inhibitor which targets HER2 and EGFR pathways has demonstrated activity with regards to response rates in patients with brain metastasis when given in combination with capecitabine (Bachelot et al., 2013). There is increasing interest in treating HER2 positive brain metastases with the recently developed antibody-cytotoxin conjugate, TDM1 (traztuzumab and emtansine) (Bartsch et al., 2014).

6.18.3 Structural variants

A comparison of CNA events in the BDCBM and primary tumours confirmed that the number of CNA were significantly higher in the brain metastases. Moreover chromothripsis-like/chromoplexy alterations were identified in 4 brain metastases. Chromothripsis is a process that results in genome shattering in a single episode resulting in genomic rearrangement from fragmentation and reassembly of contiguous chromosomal regions (Stephens et al., 2011). The mechanism underlying the phenomenon of chromothripsis remains unknown. Operational definitions of chromothripsis-like changes have differed amongst studies and have used either whole genome sequencing data or genomic arrays (Cai et al., 2014). Recently Cai et al developed a mathematical algorithm to identify chromothripsis–like changes in a large-scale analysis of genomic arrays from numerous cancer types through a screening process. They reported that chromothripsis-like alterations occurred in 11% of the breast cancers screened (stage of tumours assessed unknown).

In summary, the BDCBM had more CNA in comparison to primary breast tumours. Chromothripsis-like/chromoplexy alterations were identified in 4 tumours. Kataegis was identified in 1 breast brain metastasis which correlated with the APOBEC mutation signature.

Chapter 7

Amalgamated analysis of brain metastases genomes

7.1 Introduction

Tumours with a propensity to develop brain metastases include breast, lung, renal, melanoma and colorectal tumours. Genomes of brain metastases characterised by NGS are limited in the literature (Shah et al., 2009, Ding et al., 2010). In this chapter, I have amalgamated information regarding mutation burdens, copy number aberrations and clonal clusters from sequenced colorectal, breast and lung adenocarcinoma brain metastases, to determine whether the genomes originating from different histological subtypes share characteristic features.

7.2 Methods

7.2.1 Sequencing of brain metastases

In total 4 CCBM, 7 LABM and 7 BDCBM underwent WGS using the Illumina HiSeq 2000 platform, Chapter 2 (sections 2.11 and 2.14)

7.2.2 Evaluation of non-protein coding regions from sequencing data

Non-coding regions of the genome were interrogated by adopting two approaches (section 2.14.12). The first approach involved screening mutations genome wide to identify recurrent mutations in transcription factor binding sites (Encode database). The second approach utilised the microRNA database (http://www.mirbase.org) to identify recurrent mutations overlapping regions encoding for microRNA.

7.2.3 Copy number analysis of brain metastases

7.2.3.1 GISTIC and STAC analysis of brain metastases

GISTIC and STAC algorithms were utilised to identify shared regions of copy number aberrations of statistical significance amongst the entire cohort of brain metastases in this study (sections 2.14.10 and 2.14.11). Paired tumour and normal BAM files were processed to obtain copy number information directly from sequencing data. Analysis was undertaken using the Nexus Biodiscovery tools package.

7.2.3.2 Analysis of FFPE brain metastases using SNP arrays

Genomic DNA was extracted from 40 FFPE brain metastases of lung adenocarcinoma, colorectal and breast origin. DNA was assessed for quality, restored where necessary (Illumina manufacturer's protocol). Samples that passed quality controls were genotyped on the Human Omniexpress 24 platform. Data were subsequently processed using OncoSNP and verified for quality by visual inspection on Genome Studio (Illumina). Technical validation regarding the quality of SNP arrays obtained from FFPE samples involved comparison with SNP arrays undertaken from fresh frozen tissue from the same tumour where available.

7.2.3.3 Cluster analysis of copy number aberrations

SNP arrays from FFPE tumours were processed using OncoSNP to evaluate copy number alterations. Segmental gains and losses for each tumour were identified in relation to the average copy number of the tumour. Using bins of 500,000 bases, each segment of the genome were assigned a copy number state (gain, loss and neutral). Copy number alterations were clustered using consensus clustering, and the optimal number of clusters was identified from a consensus matrix identifying an optimal k value.

Results

7.3 Assessment of relationships between tumour characteristics

The properties of each tumour sequenced with regards to mutation burdens, clonal populations, and number of CNA are summarised in Table 7.1. For the purpose of analysis CNA are defined as deviations from average copy number.

Sample	Number of SNV	Number of indels	Number of clonal clusters	Number of CNA	Time interval between primary tumors and brain metastases (years)	Presence of chromothripsis/ chromoplexy	Arms affected
	Colorectal						
C1MET	63111	6873	35	20	2	No	
C2MET	22930	3281	17	31	3	No	
C3MET	31407	5019	24	33	6	No	
C5MET	10492	2955	7	28	1	Chromothripsis	10q
	Breast						
B1MET	17936	3676	23	111	1	Chromoplexy	1p, 6q, 7q, 8q, 13q
B2MET	2884	1150	4	56	4	Chromothripsis	14q
B3MET	20961	3612	6	15	5	No	
B4MET	10383	1774	2	48	7	Chromothripsis	17q
B5MET	5656	1980	10	94	3	Chromothripsis	1p, 6q, 11q
B6MET	5885	1034	6	42	3	No	
B7MET	61285	2611	12	74	8	No	
	Lung						
L1MET	15077	1377	7	42	0	No	
L2MET	49206	2691	13	40	0	No	
L3MET	34082	3058	23	17	0	No	
L4MET	81794	4395	25	48	1	No	
L6MET	72636	3665	39	41	0	Chromothripsis	14q
L7MET	118490	5316	37	27	0	No	
L8MET	6524	1323	4	48	3	No	

Table 7.1 | Summary of characteristics for each brain metastasis**. A time interval of 0 refers to brain metastases manifesting at the initial presentation of the cancer.**

7.3.1 Mutation Burdens

There was a wide range in overall mutation burdens amongst the brain metastases from 4,034 to 123,806 [median 25,392] mutations. Using SNVs as a measure for comparison the mutation burdens amongst the brain metastases ranged from 2,884 to 118,490 [median 2,1946] (Figure 7.1). L7MET a LABM had the greatest mutation burden, whereas the lowest mutation burden was present in B2MET. Amongst the LABM, L1MET and L8MET had the lowest mutation burdens. These brain metastases harboured activating *EGFR* mutations [L1MET exon21 L858R, L8MET exon19:c236-2250del GAATTAAGAGAAGCA]. Lower mutation frequencies in primary lung adenocarcinomas harbouring activating mutations have been reported (Imielinski et al., 2012, Ding et al., 2008). The mutation burdens varied widely within each tumour type. When all the brain metastases were considered, the mutation burdens from each tumour type overlap each other and are of similar magnitude. There was no significant difference in mutation burdens amongst tumour types (Kruskal Wallis test p=0.07). The number of small indels ranged from 1034 to 6873 [median 3007]. There was a strong relationship between the number of SNVs and small indels present in tumours (Spearman ρ=0.84 p<0.001).

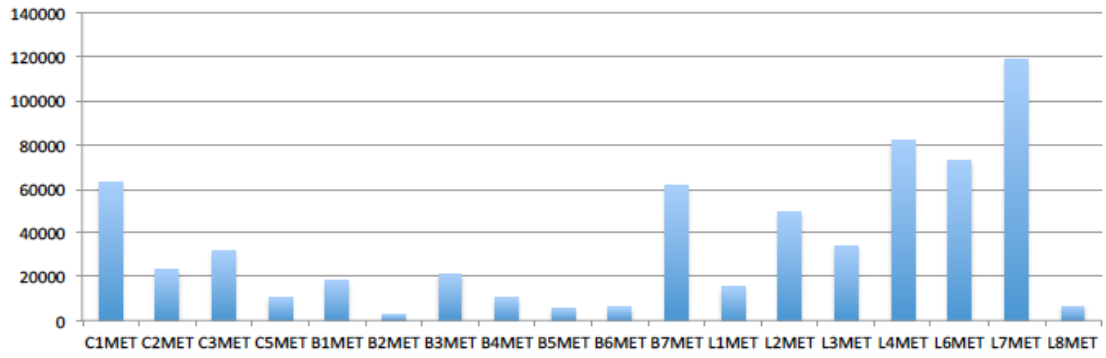


Figure 7.1 I SNV mutation burdens in the brain metastases. The SNV mutation burdens across the genome are depicted. There is a wide range in mutation burdens.

Evaluation of protein coding and non-coding regions confirmed that in all brain metastases the majority of mutations were located predominantly in intergenic locations consisting of 65 to 71% [median 68%] of overall mutation burdens. Mutations located in intergenic regions formed 28 to 34% [median 30%]. The number of mutations found within protein-coding regions formed ~1% of the overall mutation burdens in all brain metastases. Amongst the brain metastases, there was a wide variation in the mutation burdens within the exome which ranged from 1.3 to 32 per 10^6 [median 6.3 per 10^6] (Figure 7.2). There was a strong relationship between the mutation burdens within the exome and the number of mutations located in intergenic and intronic regions (Spearman ρ=0.96 p<0.001 and ρ=0.95 p<0.001 respectively).

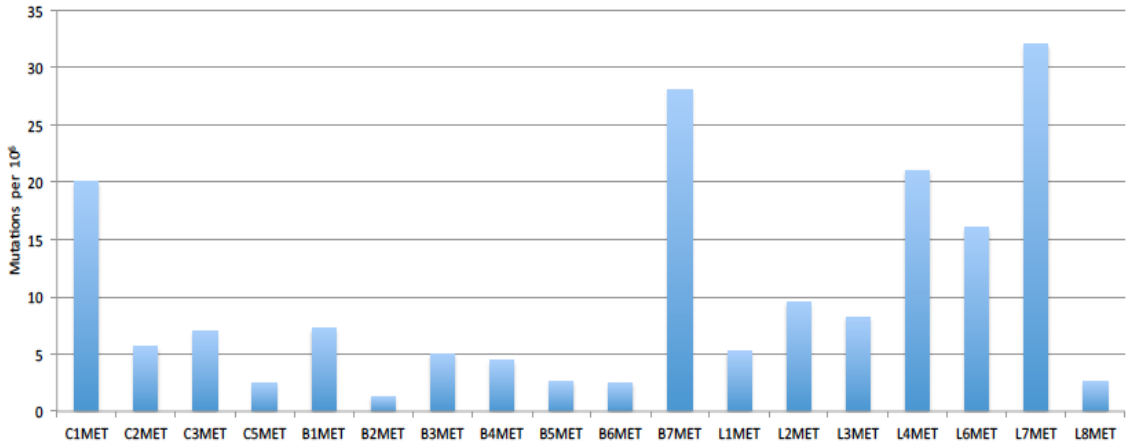


Figure 7.2 I Mutations in the exome in the brain metastases. Mutation burdens are reported per megabase.

The number of synonymous mutations ranged from 10 to 242 [median 48]. Mutations in protein-coding regions were filtered by function. The number of nonsynonymous mutations ranged from 23 to 611 [median 111.5] events per tumour. The number of nonsense mutations ranged form 0 to 56 [median 7.5] events per tumour. Nonsense mutations were absent in the tumour B2MET, whereas the number of nonsense mutations was greatest in B7MET. The number of small indels in the exome ranged from 4 to 44 [median 14] events per tumour.

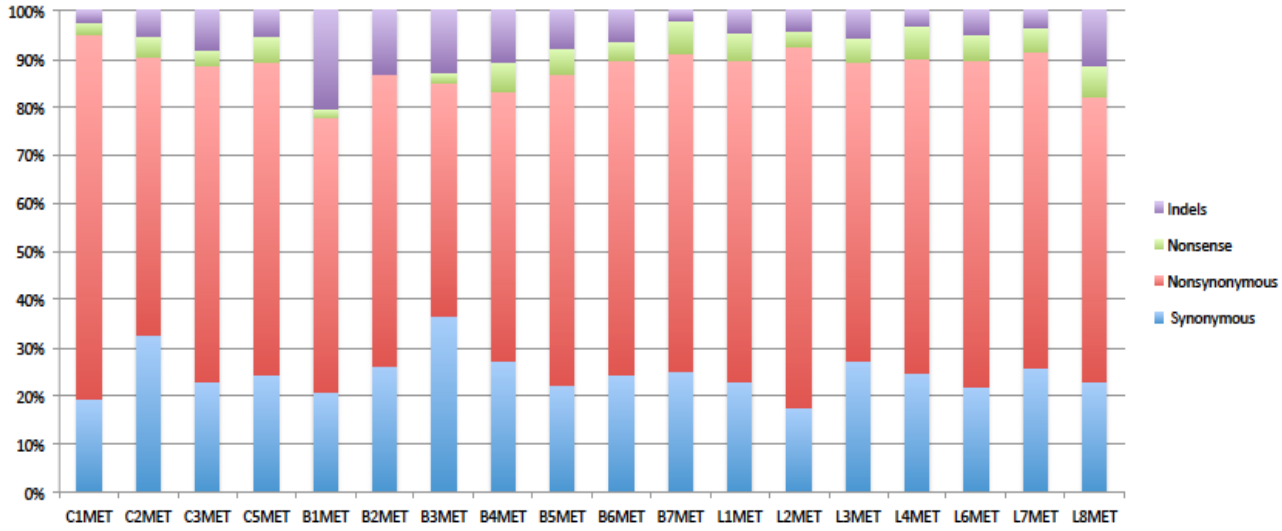


Figure 7.3 I Mutations in protein-coding regions represented as a proportion of the mutation burden in the exome. Mutations were categorised as indels, nonsense, nonsynonymous and synonymous.

Nonsynonymous mutations formed the most common type of mutation within the exome forming 47 to 74% [median 63%]. In contrast indels formed 2 to 20% [median 5.5%] of mutations with the exome. Proportionately, when considering mutations in the exome indels occurred most frequently in B1MET. Interestingly, B1MET exhibited features of chromoplexy which would be associated with widespread genomic rearrangements. Amongst the LABM proportionately indels were commonly found in L8MET (11%) which harboured a pathogenic 15 base non-frame shift deletion (L8MET exon19:c236-2250del GAATTAAGAGAAGCA).

7.3.2 Mutation spectra of brain metastases

The mutation spectra in the brain metastases were characterised using Watson and Crick base pairing (C>A/G>T, C>G/G>C, C>T/G>A, T>A/A>T, T>C/A>G, T>G/A>C). The mutation spectra in the brain metastases resembled that found within primary tumours. The mutation spectra observed in the brain metastases were generally reflective of the imprint of biological processes previously reported in these tumours types. Generally in CCBM, C>T/G>A mutations predominated across the genome and

exome [genome: median 23% range 9 to 30% exome: median 51% range 16 to 57%]. This class of base substitution is known to occur in colorectal cancer at 5'-CpG-3' sites resulting from deamination of 5-methylcytosine at these sites (Sjoblom et al., 2006). Amongst the BDCBM C>T/G>A and C>G/G>C were the most common base substitutions. In the exome C>T/G>A mutations ranged from 28 to 64% and C>G/G>C ranged from 3 to 42%. Evaluation of the mutation footprints in coding regions in primary breast tumours had previously demonstrated that C>T/G>A and C>G/G>C mutations predominated (Sjoblom et al., 2006). In LABM, the mutation spectra of the majority of tumours reflected smoking exposure with a predominance of G>T/C>A base substitutions (exome: range 14 to 51% median 42%). This is typical for the DNA insult that results from tobacco smoking and subsequent nucleotide excision base repair (Pfeifer et al., 2002, Govindan et al., 2012). Amongst all brain metastases the proportion of C>T/G>A mutations increased in the exome [range 16 to 65%, median 41%] compared to the genome [range 9 to 47% median 26%]. This observation may reflect the increased G:C content in the exome and has been reported in bladder tumours (Cazier et al., 2014a).

7.3.3 Mutation signatures

Using somatic signature mutation signatures were characterised in sequence context taking into account the base flanking the 5' and 3' end of the mutated base. For the purpose of this analysis the brain metastases were analysed as a cohort. In this process 4 signatures were generated. Signature S1 was associated with C1MET alone which was characterised by an overrepresentation of T>G mutations in the context of a C[T]T motif. Signature S1 correlates with signature [17] in a pan-cancer analysis of mutation signatures undertaken by Alexandrov et al identified in liver, gastric, oesophageal carcinomas and B cell lymphomas (Alexandrov et al., 2013).

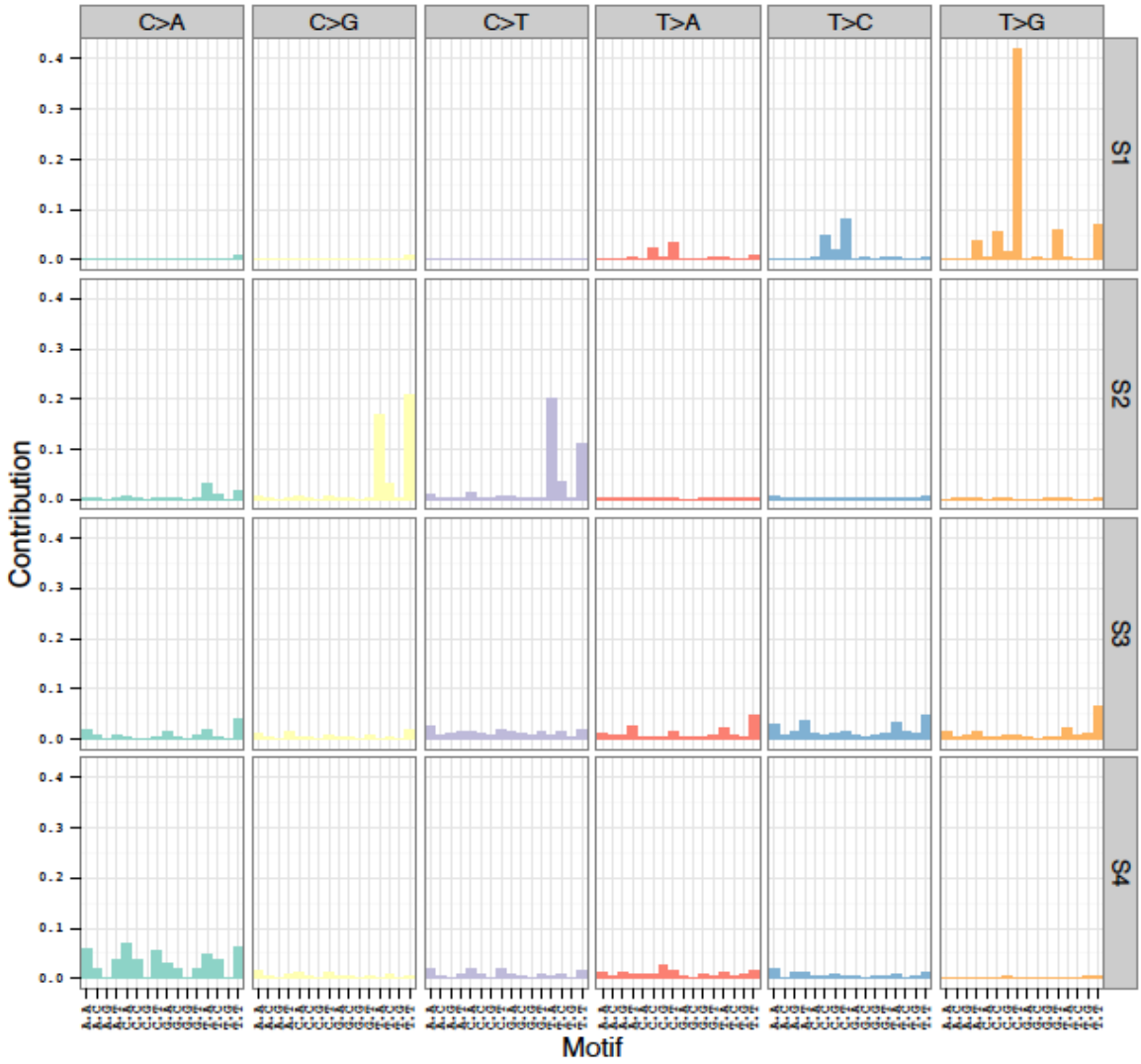


Figure 7.4 I Mutation signatures in brain metastases. 4 mutation signatures were identified.

More recently, this signature has been reported in colorectal cancers (Katainen et al., 2015). The signature has been associated with an increased frequency of mutations within CTCF/cohesin binding sites in DNA using a combination of WGS and ChiP seq technology, with mutations preferentially occurring at A.T sites in 3 mutation hotspots within the CTCF motif. The underlying aetiology for this signature is unknown, though it has been proposed that it may develop in response to environmental exposure (Katainen et al., 2015). Interestingly, the paired primary tumour, P1 did not exhibit signature [17] with C>T/G>A mutations predominating in the mutation spectra of the primary tumour. Signature S2 was identified in the LABM, L1MET and L8MET, and the

DCBM, B4MET, B5MET B6MET and B7MET. This signature corresponded to signature [2] in the pan-cancer analysis by Alexandrov et al and is typically characterised by C>G and C>T mutations in the context of a T[C]X motif (where X denotes an adenine or thymine base). This signature has been associated with the enzyme APOBEC3B and has been identified in numerous histological subtypes of cancers, including lung adenocarcinomas and breast cancers. Signatures S4 is associated with smoking exposure resulting from G>T/C>A mutations. Polycyclic aromatic hydrocarbons form bulky DNA adducts with guanine bases that are subsequently removed by nucleotide excision base repair (Pfeifer et al., 2002). Signature S4 was observed in L2MET, L3MET, L4MET, L5MET, L6MET and L7MET and resembles signature [4] in the pan-cancer analysis (Alexandrov et al., 2013). The remaining brain metastases exhibited signature S3, which does not readily correlate with signatures reported by Alexandrov et al (Alexandrov et al., 2013). Collectively, these findings suggest that mutation signatures are reflective of the biological process in action and are independent of the histological subtype of tumour.

7.3.4 Clonal structures of brain metastases

As described in chapters 3, 5 and 6 the clonal structures of the brain metastases were characterised using Pyclone. The complexity of the clonal structures varied considerably ranging from 4 to 39 [median 12.5]. Interestingly, L6MET had the most complex clonal architecture though the mutation burdens were greater in other brain metastases. The most clonally ‘simple’ structure was observed in B4MET (2 clonal groups). The driver mutations of primary tumours were often though not universally found within major clones in the brain metastases. Broadly, mutations found in major clones were often represented with a high VAF within the tumour and encompassed regions of LOH or copy number gains. Driver mutations in primary tumours were identified in all the brain metastases with the exception of B1MET and B2MET. This finding may be reflective of the characteristic heterogeneity observed in breast tumours.

A comprehensive analysis of sequencing a large cohort of breast tumours undertaken by the TCGA identified mutations were only present in a few genes with a frequency >10%. These genes included *TP53*, *PIK3CA* and *GATA3*. There was an association between the number of SNV mutations in the genome and the number of clonal groups present in the tumour (Spearman ρ=0.84 $p<0.001$).

7.3.5 Copy number aberrations

A measure of copy number aberrations can be considered as a surrogate for chromosomal instability. Amongst the brain metastases CNA varied widely ranging from 15 to 115 [median 41.5]. L3MET had the least number of CNA with 15 in total. B1MET and B5MET had the most number of CNA amongst the brain metastases 115 and 94 respectively. B1MET harboured chromoplexy-like alterations affecting several chromosome arms (1p, 6q, 7q, 8q, 13q), whereas chromothripsis-like alterations were observed in B5MET affecting chromosomes (1p, 6q, 11q). Chromothripsis and chromoplexy can result in the development of chromosomal aberrations in a narrow time point resulting in chromosomal instability (Stephens et al., 2011, Zhang et al., 2013, Baca et al., 2013). Although chromothripsis was identified amongst the other brain metastases the number of CNA in these tumours was lower 28 to 56 [median 44.5]. This may reflect the observation that chromothripsis in the other tumours only affected a single chromosome arm. The proportion of the genomes that deviated from the mean copy number of the tumour ranged from 7% to 48% [median 36%].

Amongst the brain metastases there was no relationship between the number of CNA and the proportion of the genome deviating from the mean copy number (Spearman ρ= -0.25 p=0.31). Genome wide, there was no relationship between the SNV mutation burdens and the number of CNA and or proportion genome deviating from the mean copy number (Spearman ρ= -0.43 p=0.09, ρ= -0.23 p=0.35 respectively).

7.3.6 Chronological relationships

In the cohort of brain metastases sequenced 5 LABM presented with brain metastases at original presentation. The remaining metastases were resected after a time interval, following the occurrence of the primary tumour ranging from 1 to 8 years [median 3.6]. There was no relationship between the time interval between the primary tumour and resection of the brain metastasis and the SNV mutation burdens and CNA (Spearman ρ= -0.41 p= 0.09, ρ= 0.28 p=0.28 respectively).

7.4 Interrogation of non-protein coding regions

Two approaches were adopted to identify recurrent mutations of potential significance in non-protein coding regions. Firstly the Encode database served to provide a list of transcription factor binding sites that were interrogated. Secondly the microRNA database (http://www.mirbase.org) was used to provide a list of coordinates for microRNA loci within the genome. Somatic mutations from all the sequenced brain metastases (n=18) were amalgamated to interrogate the aforementioned databases. Analysis confirmed the absence of recurrent mutations encompassing transcription factor binding sites and regions encoding microRNAs amongst the metastases. For the purpose of this analysis the brain metastases were analysed as a cohort and within each histological subtype of tumour.

7.5 Driver gene discovery in brain metastasis

Although no drivers of brain metastasis were identified amongst the individual tumour types. I proceeded to analyse the brain metastases as a cohort. Somatic variants identified from all the brain metastases were combined and analysed using the IntOGen algorithm (FDR < 0.05). Mutations identified in this process were subsequently filtered to exclude potentially benign or moderate somatic mutations (SIFT <0.2 and PolyPhen2 >0.8). The pan-cancer genes *TP53* and *ARID1A* were identified through this process. *TP53* was the most commonly mutated gene amongst all the brain metastases. In total

10 somatic mutations were found in individual tumours in *TP53* (3 colorectal, 2 breast, 5 lung adenocarcinomas).

Interestingly, the gene *DST* reached statistical significance with somatic mutations present in 5 individual tumours. *DST* mutations were found in C1MET, C3MET, L2MET, L3MET and B7MET. In C1MET a nonsynonymous mutation was present L845F, and a frameshift mutation in C3MET, exon23 c3812 insertion T were detected. L2MET and L3MET harboured the nonsynonymous mutations R5096L and A2247D respectively. Two independent mutations were identified in B7MET a nonsense mutation E1843X and a nonsynonymous mutation I95T. The role of DST is poorly understood, it encodes an adhesion plaque protein which is may be responsible for anchoring neural intermediate filaments to the actin cytoskeleton (Yang et al., 1999, Yang et al., 1996).

I extended my analysis to determine the frequency of mutations in *DST* in primary tumours using TCGA cBioPortal as a reference. The frequency of mutations was 4.7% in colorectal tumours (n= 10/230), 1.2% in invasive breast tumour (n= 6/482) and 15.9% in lung adenocarcinomas (n= 26/182). Mutations in *DST* in TCGA cBioPortal were predominantly missense mutations which occurred at various loci throughout the gene. No recurrent mutation 'hotspots' were identified amongst these mutations within individual tumour types or amongst the tumour types as a cohort. Amplifications in DST were reported in colorectal, breast and lung adenocarcinoma at low frequencies 0%, 1.2% and 1% respectively.

Gene	Sample	Location of mutation	Type of mutation	LOH at locus of mutation	Copy Number at locus of mutation	Poly phen 2 Score	SIFT Score	Mutation frequency in gene TCGA cBioPortal %	Mutation frequency in gene at that loci -TCGA cBioPortal
DST	C1MET	exon17:c2535 C>G:L845F	Nonsynonymous	No	2	0	1	4.7% (n= 10/230)	0/212
	C3MET	exon23:c3812 insertion T	Frame shift Insertion	Yes	2	--------	--------		0/212
	L2MET	exon6:c15287 C>A:R5096L	Nonsynonymous	No	3	0.03	1	15.8% (n= 26/182).	0/212
	L3MET	cxon24:c6740 G>T:A2247D	Nonsynonymous	Yes	2	0.16	0.98		0/212
	B7MET	exon23:c.5527 G>T:p.E1843X	Nonsense	No	2	------	------	1.2% n= 6/482)	0/212
	B7MET	exon6:c.584 A>Gp.I95T	Nonsynonymous	No	2	0.97	0.01		0/212

Table 7.2 I *DST* was identified as putative driver for brain metastases

7.6 Clonal location of *DST*

To determine the relative importance of *DST* within the clonal structures of the brain metastases Pyclone was utilised to reconstruct the clonal architecture of the genomes. *DST* mutations have been plotted in green, amongst the mutations for primary drivers of the each tumour type and a cohort of randomly selected genes in Figures 7.5 to 7.9.

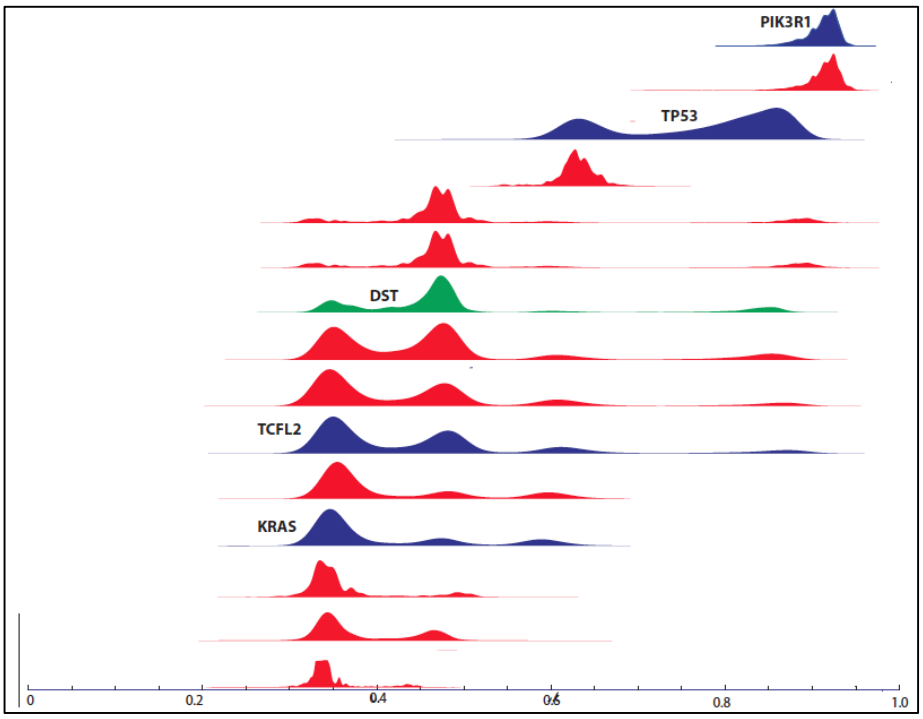


Figure 7.5l Clonal architecture of C1MET. The clonal architecture of C1MET is depicted using randomly selected genes. Primary driver mutations are highlighted in blue and the gene *DST* is illustrated in green. *DST* is located in a subclonal group in the clonal structure of C1MET.

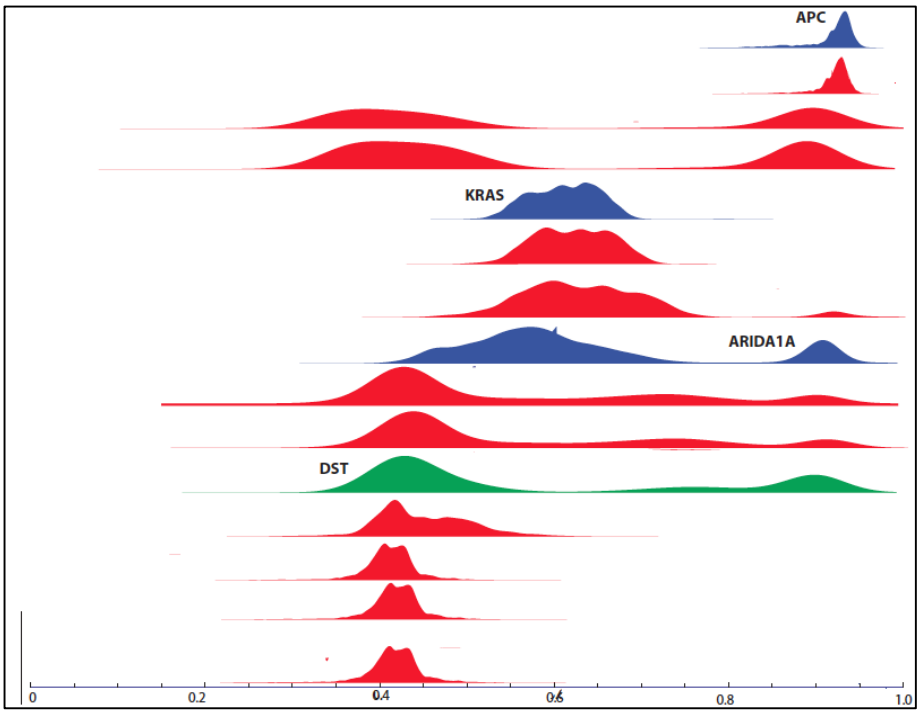


Figure 7.6l Clonal architecture of C3MET. *DST* is located in a predominantly subclonal group. In comparison *APC* is located in a major clonal group.

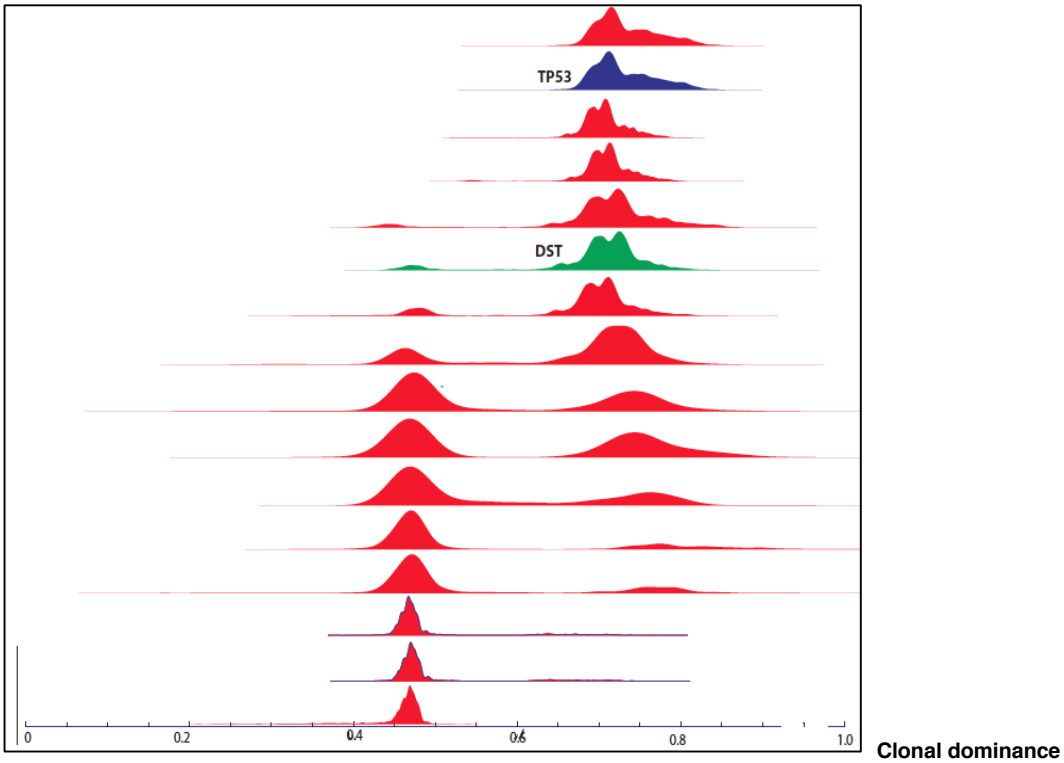


Figure 7.7l Clonal architecture in L2MET. *DST* is located in a major clonal group

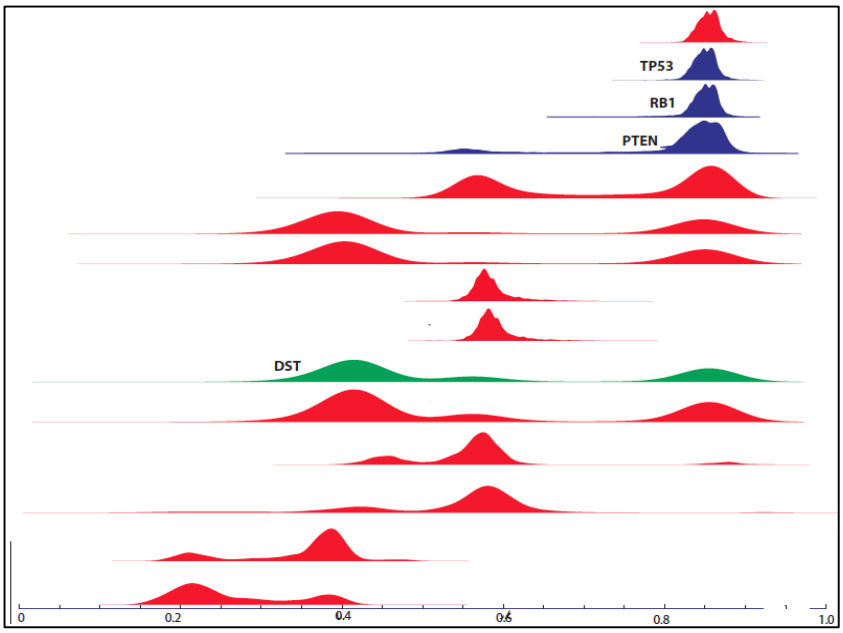


Figure 7.8l Clonal architecture in L3MET.*DST* is located in subclonal groups, whereas *TP53*, *RB1* and *PTEN* are located in major clonal groups.

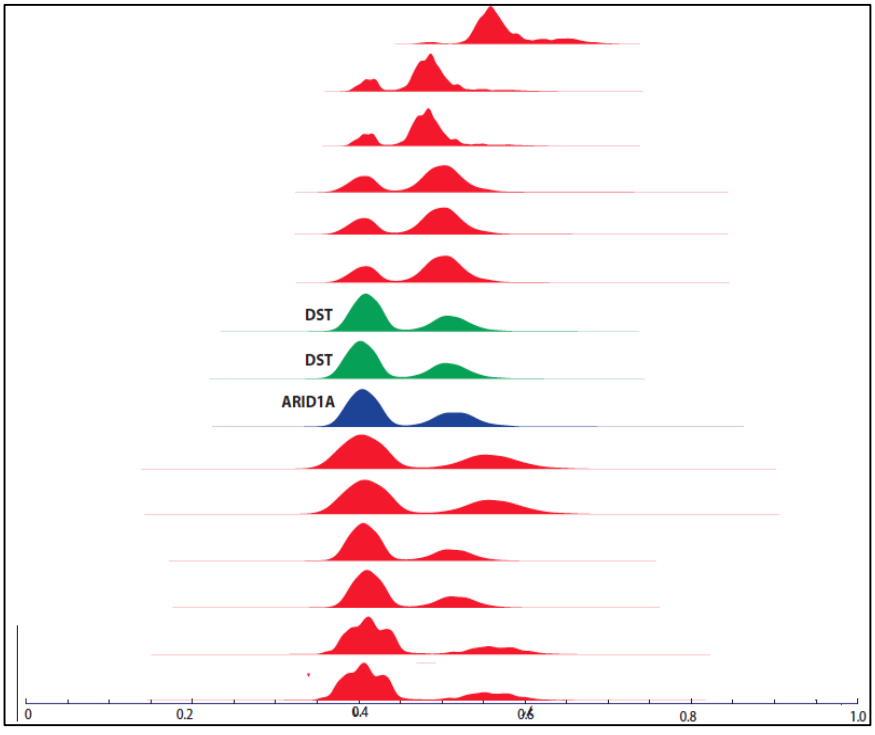


Figure 7.9l Clonal architecture of B7MET. *DST* is located in a subclonal group alongside *ARID1A*

The complexity of the metastases harbouring *DST* ranged from 12 to 35 clonal groups [median 18]. *DST* was predominantly located in subclonal groups. An exception to this is L2MET where *DST* was located in a major clone. The VAF of *DST* ranged from 20 to 50 [median 36]. The median VAF of the metastases with a *DST* mutation ranged from 24 to 42. In L2MET where *DST* was present in a major clone the VAF of *DST* were 36 whereas it was 42 when all mutations in this metastasis were considered. It is evident from Figures 7.5 to 7.9 that drivers of the primary tumours in particular histological subtypes were often located in major clones whereas somatic mutations in *DST* were predominantly located in subclonal groups.

Although *DST* was identified as a putative driver of brain metastases through IntOGen analysis, these mutations are largely located in subclonal groups questioning the role of *DST* in driving brain metastasis. It is conceivable that *DST* has a general role in the metastatic process which is not brain specific, as mutations in *DST* are infrequent in primary tumours. Analysis of sequencing data from metastases to other sites is likely to address this.

7.7 Copy number analysis of brain metastases as a cohort

7.7.1 GISTIC and STAC analysis of brain metastases

To identify shared regions of copy number aberrations of statistical significance GISTIC and STAC algorithms were utilised using WGS data. No regions reached significance using the STAC algorithm following exclusion of regions that overlapped regions of segmental duplications. However, using the GISTIC algorithm amplifications overlying the gene *EGFR* reached statistical significance in the region of chr7:55,210,000-55,215,000 ($p < 0.05$). In total 14 brain metastases (4 colorectal, 6 breast, 4 lung adenocarcinoma) were found to have gains overlying this region. The brain metastases with no copy number gains in this region included B2MET, B6MET, L3MET and L4MET.

Bos et al developed a murine model to analyse brain metastases in breast tumours. Using an iterative process to isolate brain ‘seeking’ tumour cells and subsequently analysing gene

expression in these cells, they identified 3 candidates genes enriched in this process: *COX2* which encodes a prostaglandin synthesising enzyme, *ST6GALNAC5* a sialyltransferase involved in cell in surface interactions and heparin binding EGF (*HBEGF)* which is a ligand for the EGFR. Mice treated with cetuximab a monoclonal antibody that targets the extracellular domain of the EGFR demonstrated reduced metastatic activity. COX2 and epiregulin (EREG) a ligand of the EGFR have been associated with colonisation of metastases to the lungs in breast tumours (Minn et al., 2007). Bos et al proposed that the *EGFR* and *COX2* were associated with mediating metastases to the lung and brain in breast tumours, whereas *ST6GALNAC5* was selective for mediating brain metastasis as it is selectively expressed in the brain microenvironment.

7.8 SNP array analysis of FFPE brain metastases

Chromothripsis was identified in several brain metastases from the discovery set of fresh frozen tumours. To screen for chromothripsis from a larger cohort of tumours, SNP arrays were undertaken on brain metastases derived from FFPE tumours of colorectal, breast and lung adenocarcinoma origin. In total FFPE derived tumour tissue were available from 3 brain metastases from the original discovery set. This enabled a comparison of SNP arrays undertaken on fresh frozen and FFPE tissue derived from the same tumour and the opportunity to verify the reliability of FFPE genotyped SNP arrays. Figure 7.10 demonstrates that the FFPE genotyped SNP array closely resembled the paired fresh frozen SNP array from the same tumour.

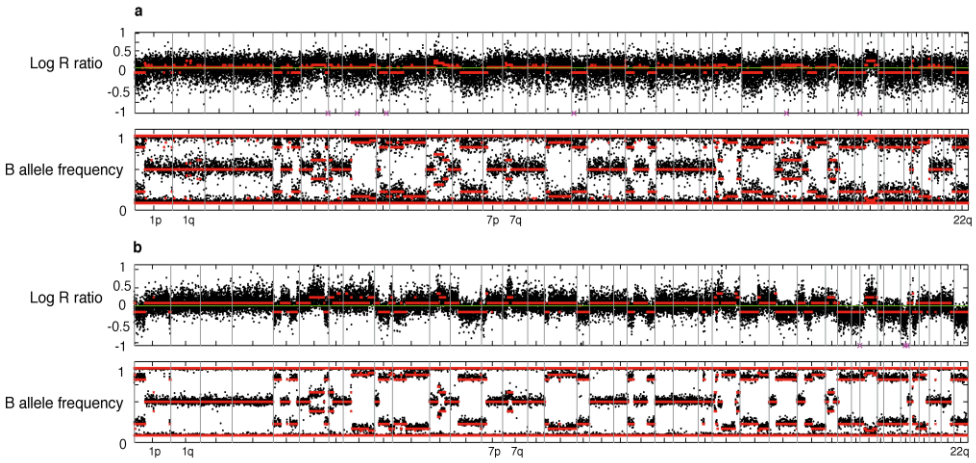


Figure 7.10 | Comparison of SNP array analysis of B3MET. The figure depicts a comparison of SNP arrays performed on the brain metastasis B3MET using fresh frozen and FFPE tissue from the same tumour. A genome view of copy number alterations is presented from OncoSNP analysis. a) FFPE derived tumour b) fresh frozen derived tumour. The FFPE genotyped tumour closely resembles that of the fresh frozen specimen.

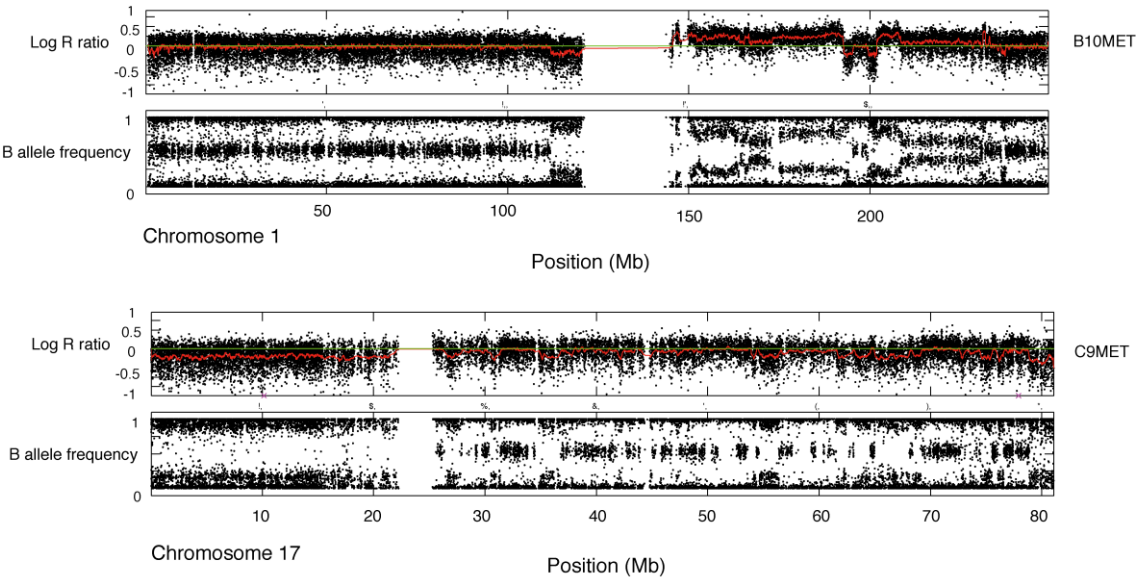


Figure 7.11 | Chromothripsis identified in brain metastases derived from FFPE in B10MET and C9MET.

In total 22 tumours passed quality control assessment 6 colorectal, 9 breast and 7 lung adenocarcinoma breast metastases (section 2.9). Chromothripsis was identified in 10 of the 22 brain metastases (Table 7.3)

Tumour	Number of CNA	Chromosomal aberration	Chromosome arm involved
B3MET	16		
B4MET	62	chromothripsis	17q
B5MET	152	chromothripsis	1p, 6q, 11q
B8MET	8		
B9MET	51	chromothripsis	13q
B10MET	42	chromothripsis	1p
B11MET	48		
B12MET	25	chromothripsis	20q
B13MET	38		
C1MET	13		
C5MET	30	chromothripsis	10q
C6MET	47	chromothripsis	20q
C7MET	68		
C9MET	6	chromothripsis	17q
C10MET	43		
L6MET	45	chromothripsis	14q
L9MET	29		
L10MET	32		
L11MET	45		
L12MET	55		
L13MET	75	chromothripsis	10q
L14MET	36		

Table 7.3 I Summary of chromosomal aberrations in FFPE derived brain metastases

The incidence of chromothripsis in the FFPE derived brain metastases was 55% (n=5/9) in breast tumours, 50% (n=3/6) in the colorectal carcinomas and 29% (n=2/7) in lung adenocarcinomas. Operational definitions of chromothripsis vary in the literature. Chromothripsis has been identified through sequencing data and through the analysis of SNP arrays (Cai et al., 2014, Stephens et al., 2011). The underlying aetiology for this phenomenon is unknown, though it is generally thought to occur in a single event leading to numerous genomic rearrangements which are restricted to a single or a few arms (Zhang et

al., 2013). For the purpose of this analysis I have utilised operational terms based on those devised by Caizier et al defining chromothripsis as 10 ≥ alterations in copy number and LOH in ≤3 arms (Cazier et al., 2014a). Cai et al developed a mathematical algorithm to identify chromothripsis–like changes in a large-scale analysis of tumours of various histological subtypes. They reported that chromothripsis-like alterations occurred in 11% of the breast cancers screened (Cai et al., 2014). The TCGA undertook low pass sequencing on a cohort of primary lung adenocarcinomas estimating the incidence of chromothripsis to be 6% (Cancer Genome Atlas Research, 2014). The incidence of chromothripsis in lung adenocarcinomas was estimated to be 9% by Cai et al (Cai et al., 2014). The incidence of chromothripsis in colorectal tumours is unclear as a large-scale analysis addressing this has not been conducted to date. Taken together, although speculative as the studies pertaining to chromothripsis have used different methods, seemingly the incidence of chromothripsis may be greater in brain metastases in comparison to primary tumours.

7.9 Consensus clustering of brain metastases

Chromosomal aberrations are thought to be characteristic to particular histological subtypes of tumour. To determine if this was replicated in the brain metastases tumours, I undertook consensus clustering using copy number aberrations to determine whether tumours of particular histological subtypes clustered together. Using SNP arrays copy number information was divided in to 500,000bp bins based on deviations from the mean copy number assigning 3 copy number states, gains, losses or neutral. Using consensus clustering the brain metastases formed 7 groups of clusters (Figure 7.12).

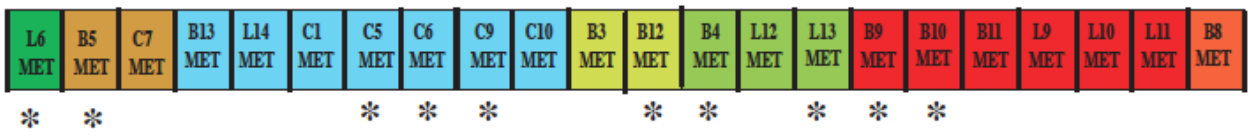


Figure 7.12 I Consensus clustering of copy number aberrations. The FFPE derived brain metastases formed 7 groups upon clustering of copy number aberrations. Colorectal metastases largely clustered together. * denotes the presence of chromothripsis.

From this analysis it is evident that generally colorectal tumours clustered together. C1MET, C5MET, C6MET, C9MET and C10MET were found in the same group. An exception to this was C7MET which clustered with B5MET which had chromothripsis-like changes involving 1p, 6q, 11q. The number of CNA in C7MET and B5MET were higher compared to most other brain metastases, 68 and 152 respectively. B8MET did not cluster with remaining brain metastases presumably due to the low number of CNA in this tumour (n=8). Brain metastases appeared to cluster together based on copy number alterations across the genome, irrespective of the presence of chromothripsis. In support of this L13MET and C5MET were found in different clusters despite displaying chromothripsis-like features in chromosome 10q. Similar observations were observed in B12MET and C6MET which harboured chromothripsis-like alterations in 20q, though were located in different clusters.

7.10 Discussion and overall conclusions

Patients presenting with brain metastases often have a poor prognosis with limited treatment options. Surgery and stereotactic radiotherapy are therapeutic options that may be considered in patients presenting with ≤3 metastases. In patients presenting with several metastases WBRT may be considered. WBRT in combination with surgery and stereotactic radiotherapy improves control from local recurrence. However, the overall survival of patients is not improved with the addition of WBRT, with this treatment modality being associated with morbidity (Kocher et al., 2011, Aoyama et al., 2006).

The number of patients presenting with brain metastases is increasing (Owonikoko et al., 2014). This is attributable to numerous factors. Firstly, with the development of new systemic treatments extracranial disease is increasingly better controlled leading to improved survival rates, leading to more patients presenting with advanced disease. This has been coupled with improved imaging techniques and access to these resources. These factors have resulted in a clinical patient group with an unmet need.

Studies in brain metastases have focused on animal models where brain 'homing' tumour cells are identified in an iterative process or tumour cells are directly implanted (Eichler et al., 2011). Critics have questioned the ability of these models to faithfully replicate the biological processes in action in humans. Many laboratory techniques to analyse genomes have historically relied on fresh frozen tissue to generate reliable data. Fresh frozen brain metastases are limited in availability and obtaining tumour tissue may be impractical or inappropriate in this cohort of patients.

In this DPhil I have characterised the genomic landscape of a cohort of brain metastases originating from colorectal, ductal breast and lung adenocarcinoma genomes. Understanding the degree of heterogeneity between metastatic and primary tumours is likely to inform decisions in clinical treatment and the design of new therapies. To characterise the degree of heterogeneity in the brain metastases compared to primary tumours I utilised data obtained from the TCGA to serve as a comparison.

7.10.1 Mutation Burdens

There was a wide range in mutation burdens amongst the brain metastases across the genome and within the exome, 4,034 to 123,806 [median 25,392] and 1.3 to 32 per 10^6 [median 6.3 per 10^6] respectively. As outlined in chapters 3, 5 and 6 generally the mutation burdens published in primary tumours were of a comparable magnitude. TCGA reported exome sequencing data from a large cohort of primary colorectal cancers with the majority of tumours (84%) having a mutation frequency less than 8.24 per 10^6. Tumours with a mutation frequency greater than 12 per 10^6 were considered as hypermutated (Cancer Genome Atlas, 2012). The TCGA reported a median of 5.8 per 10^6, mean 8.9 per 10^6 [range 0.5 to 48] in primary lung adenocarcinomas (Cancer Genome Atlas Research, 2014). I estimated the median number of mutations as 2 per 10^6 in a cohort of 1000 primary breast tumours from the TCGA.

In comparison, analysis of the paired primary tumour P1 and colorectal brain metastasis C1MET (chapter 4) demonstrated that the bulk of mutations in the primary were retained in the brain metastasis 94%. However C1MET evolved to become hypermutated developing a large number of *de novo* mutations, possibly in response to exposure to therapeutic agents in response to oxidative stress.

7.10.2 Mutation spectra/signatures

Analysis of mutation spectra and signatures identified characteristic changes in the brain metastases reflecting imprints of biological processes in action and were independent of the histological subtypes of tumour (Alexandrov et al., 2013). Signatures identified in this process included a predominance of G>T/C>A mutations associated with smoking, C>G and C>T mutations in the context of a T[C]X motif associated with the APOBEC3B enzyme. Characterisation of the mutation footprints of primary tumours obtained from the TCGA demonstrated that the mutation spectra in the metastases generally resembled unpaired primary tumours. However the comparison of P1 and C1MET confirmed the appearance of a signature in C1MET which is poorly understood (T>G/A>C overrepresented in the context of a C[T]T motif). This was absent in the primary tumour, demonstrating the potential for signatures to evolve in brain metastases.

7.10.3 Driver mutations in brain metastases

Drivers of primary tumours were retained in the brain metastases and were often located in major clones or major subclonal groups. No drivers of brain metastases were found upon analysis of individual tumour types. However a combined analysis of the cohort of metastases identified the gene *DST*. Reconstruction of the clonal architectures of the brain metastases demonstrated that *DST* was often located in subclonal groups questioning the role of this gene as a driver specific to brain metastasis. Interestingly, the comparison of C1MET and P1 demonstrated that the complexity of genome can evolve which is

accompanied by the development of new driver mutations. Although driver mutations in the primary tumour were retained in C1MET, *P13KR1* emerged as a new driver mutation in the metastasis in a dominant clonal group demonstrating heterogeneity between the primary and metastasis. Several clinical studies targeting primary drivers in brain metastases have demonstrated clinical benefit in patients with pathogenic driver mutations. Examples of these include the activating mutations in the *EGFR* gene in lung adenocarcinomas, *BRAF* mutations in melanomas and *HER2* amplification with lapatinib and capecitabine (Welsh et al., 2013, Sutherland et al., 2010, Falchook et al., 2012). Although the brain metastases were unpaired, I have demonstrated that generally, though not universally, primary driver genes were located in major clones in the brain metastases. This supports the emerging clinical view to consider treating selective patients with brain metastases with biological agents (Soffietti et al., 2012). Advances in future treatments for these subgroup of patients will involve developing treatment paradigms to optimise outcomes (Lin, 2014).

7.10.4 Chromosomal aberrations

My analysis has suggested that brain metastases have increased chromosomal instability in comparison to primary tumours. Analysis of CNA from genotyped SNP arrays in the brain metastases demonstrated that CNA were more frequently observed in the brain metastases when compared to unpaired primary tumours. This was observed in lung and breast adenocarcinomas. As the samples size of the CCBM was small there was no statistically significant difference in CNA between CCBM and MSI negative primary colorectal tumours.

Building on the identification of chromothripsis in the discovery set of brain metastases, FFPE derived brain metastases were genotyped to determine the frequency of this aberration in a larger cohort.

Although operational terms to define chromothripsis in the literature differ, chromothripsis-like alterations were frequent amongst the brain metastases and seemingly more common than that reported in primary tumours. It is considered to occur in a genome shattering single

event, rather than the gradual accumulation of genetic aberrations (Zhang et al., 2013). The rapid acquisition of a large number of genomic aberrations would promote tumour proliferative events.

The mechanisms governing chromothripsis are poorly understood. Several models have been proposed to explain this phenomenon. Stephens et al postulated that the affected chromosome is fragmented resulting in loss of some segments interspersed with stochastic joining of other regions (Stephens et al., 2011). An alternative theory proposed that chromothripsis developed from dysfunctional replication forks resulting in cycles of microhomology mediated break induced replication with sequences lying distal being incorporated at dysfunctional replication forks (Liu et al., 2012b).

Zhang et al have recently experimentally provided new insights into the mechanisms underlying chromothripsis utilising a combination of WGS and live cell imaging (Zhang et al., 2015). Chromosomes that fail to correctly attach to the mitotic spindle can physically separate and are partitioned into structures called micronuclei. Zhang et al disrupted the mitotic spindle using the agent nocodazole to generate micronuclei. Using live cell imaging they observed that micronuclei ruptured during the S phase followed by further cell division. DNA replication within micronuclei was reduced and copy number asymmetry developed from the unequal distribution of chromosomes amongst daughter cells. Subsequent sequence analysis confirmed copy number asymmetry and the concentration of genomic rearrangements on micronucleus derived chromosomes. The development of chromothripsis in the brain metastases may result in response to cellular stress generated from environmental exposure to therapies. Understanding the biological processes underlying chromothripsis and quantifying its incidence in metastatic tumours may provide further targets in the armamentarium to target tumours and metastatic development.

In summary, brain metastases have comparable mutation burdens and signatures to primary tumours. Drivers in primary tumours are retained within the metastases and often are

located in major clonal groups. Distinctive attributes found in the brain metastases involve the apparent chromosomal instability as supported by an increase in CNA and frequent occurrence of chromothripsis-like alterations in the brain metastases.

References

ABBOTT, N. J., RONNBACK, L. & HANSSON, E. 2006. Astrocyte-endothelial interactions at the blood-brain barrier. *Nature Reviews Neuroscience,* 7**,** 41-53.

ADZHUBEI, I. A., SCHMIDT, S., PESHKIN, L., RAMENSKY, V. E., GERASIMOVA, A., BORK, P., KONDRASHOV, A. S. & SUNYAEV, S. R. 2010. A method and server for predicting damaging missense mutations. *Nat Methods,* 7**,** 248-9.

ALEXANDROV, L. B., NIK-ZAINAL, S., WEDGE, D. C., APARICIO, S. A. J. R., BEHJATI, S., BIANKIN, A. V., BIGNELL, G. R., BOLLI, N., BORG, A., BORRESEN-DALE, A. L., BOYAULT, 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., IMIELINSK, M., JAGER, N., JONES, D. T. W., 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. J., VALDES-MAS, R., VAN BUUREN, M. M., VAN 'T VEER, L., VINCENT-SALOMON, A., WADDELL, N., YATES, L. R., 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., GENOME, A. P. C., CONSORTIUM, I. B. C., CONSORTIUM, I. M.-S. & PEDBRAIN, I. 2013. Signatures of mutational processes in human cancer. *Nature,* 500**,** 415-+.

AMADO, R. G., WOLF, M., PEETERS, M., VAN CUTSEM, E., SIENA, S., FREEMAN, D. J., JUAN, T., SIKORSKI, R., SUGGS, S., RADINSKY, R., PATTERSON, S. D. & CHANG, D. D. 2008. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *Journal of Clinical Oncology,* 26**,** 1626-1634.

AOYAMA, H., SHIRATO, H., TAGO, M., NAKAGAWA, K., TOYODA, T., HATANO, K., KENJYO, M., OYA, N., HIROTA, S., SHIOURA, H., KUNIEDA, E., INOMATA, T., HAYAKAWA, K., KATOH, N. & KOBASHI, G. 2006. Stereotactic radiosurgery plus whole-brain radiation therapy vs stereotactic radiosurgery alone for treatment of brain metastases: a randomized controlled trial. *JAMA,* 295**,** 2483-91.

ARTALE, S., SARTORE-BIANCHI, A., VERONESE, S. M., GAMBI, V., SARNATARO, C. S., GAMBACORTA, M., LAURICELLA, C. & SIENA, S. 2008. Mutations of KRAS and BRAF in primary and matched metastatic sites of colorectal cancer. *Journal of Clinical Oncology,* 26**,** 4217-4219.

BACA, S. C., PRANDI, D., LAWRENCE, M. S., MOSQUERA, J. M., ROMANEL, A., DRIER, Y., PARK, K., KITABAYASHI, N., MACDONALD, T. Y., GHANDI, M., VAN ALLEN, E., KRYUKOV, G. V., SBONER, A., THEURILLAT, J. P., SOONG, T. D., NICKERSON, E., AUCLAIR, D., TEWARI, A., BELTRAN, H., ONOFRIO, R. C., BOYSEN, G., GUIDUCCI, C., BARBIERI, C. E., CIBULSKIS, K., SIVACHENKO, A., CARTER, S. L., SAKSENA, G., VOET, D., RAMOS, A. H., WINCKLER, W., CIPICCHIO, M., ARDLIE, K., KANTOFF, P. W., BERGER, M. F., GABRIEL, S. B., GOLUB, T. R., MEYERSON, M., LANDER, E. S., ELEMENTO, O., GETZ, G., DEMICHELIS, F., RUBIN, M. A. & GARRAWAY, L. A. 2013. Punctuated evolution of prostate cancer genomes. *Cell,* 153**,** 666-77.

BACHELOT, T., ROMIEU, G., CAMPONE, M., DIERAS, V., CROPET, C., DALENC, F., JIMENEZ, M., LE RHUN, E., PIERGA, J. Y., GONCALVES, A., LEHEURTEUR, M., DOMONT, J., GUTIERREZ, M., CURE, H., FERRERO, J. M. & LABBE-DEVILLIERS, C. 2013. Lapatinib plus capecitabine in patients with previously untreated brain metastases from HER2-positive metastatic breast cancer (LANDSCAPE): a single-group phase 2 study. *Lancet Oncol,* 14**,** 64-71.

BALLABH, P., BRAUN, A. & NEDERGAARD, M. 2004. The blood-brain barrier: an overview - Structure, regulation, and clinical implications. *Neurobiology of Disease,* 16**,** 1-13.

BANERJI, S., CIBULSKIS, K., RANGEL-ESCARENO, C., BROWN, K. K., CARTER, S. L., FREDERICK, A. M., LAWRENCE, M. S., SIVACHENKO, A. Y., SOUGNEZ, C., ZOU, L., CORTES, M. L., FERNANDEZ-LOPEZ, J. C., PENG, S., ARDLIE, K. G., AUCLAIR, D., BAUTISTA-PINA, V., DUKE, F., FRANCIS, J.,

JUNG, J., MAFFUZ-AZIZ, A., ONOFRIO, R. C., PARKIN, M., PHO, N. H., QUINTANAR-JURADO, V., RAMOS, A. H., REBOLLAR-VEGA, R., RODRIGUEZ-CUEVAS, S., ROMERO-CORDOBA, S. L., SCHUMACHER, S. E., STRANSKY, N., THOMPSON, K. M., URIBE-FIGUEROA, L., BASELGA, J., BEROUKHIM, R., POLYAK, K., SGROI, D. C., RICHARDSON, A. L., JIMENEZ-SANCHEZ, G., LANDER, E. S., GABRIEL, S. B., GARRAWAY, L. A., GOLUB, T. R., MELENDEZ-ZAJGLA, J., TOKER, A., GETZ, G., HIDALGO-MIRANDA, A. & MEYERSON, M. 2012. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature,* 486**,** 405-9.

BAO, S., JIANG, R., KWAN, W., WANG, B., MA, X. & SONG, Y. Q. 2011. Evaluation of next-generation sequencing software in mapping and assembly. *J Hum Genet,* 56**,** 406-14.

BARBAS, C. F., 3RD, BURTON, D. R., SCOTT, J. K. & SILVERMAN, G. J. 2007. Quantitation of DNA and RNA. *CSH Protoc,* 2007**,** pdb ip47.

BARNHOLTZ-SLOAN, J. S., SLOAN, A. E., DAVIS, F. G., VIGNEAU, F. D., LAI, P. & SAWAYA, R. E. 2004. Incidence proportions of brain metastases in patients diagnosed (1973 to 2001) in the Metropolitan Detroit Cancer Surveillance System. *J Clin Oncol,* 22**,** 2865-72.

BARTSCH, R., BERGHOFF, A. S. & PREUSSER, M. 2014. Breast cancer brain metastases responding to primary systemic therapy with T-DM1. *Journal of Neuro-Oncology,* 116**,** 205-206.

BASS, A. J., LAWRENCE, M. S., BRACE, L. E., RAMOS, A. H., DRIER, Y., CIBULSKIS, K., SOUGNEZ, C., VOET, D., SAKSENA, G., SIVACHENKO, A., JING, R., PARKIN, M., PUGH, T., VERHAAK, R. G., STRANSKY, N., BOUTIN, A. T., BARRETINA, J., SOLIT, D. B., VAKIANI, E., SHAO, W., MISHINA, Y., WARMUTH, M., JIMENEZ, J., CHIANG, D. Y., SIGNORETTI, S., KAELIN, W. G., SPARDY, N., HAHN, W. C., HOSHIDA, Y., OGINO, S., DEPINHO, R. A., CHIN, L., GARRAWAY, L. A., FUCHS, C. S., BASELGA, J., TABERNERO, J., GABRIEL, S., LANDER, E. S., GETZ, G. & MEYERSON, M. 2011. Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A-TCF7L2 fusion. *Nat Genet,* 43**,** 964-8.

BATLLE, E., SANCHO, E., FRANCI, C., DOMINGUEZ, D., MONFAR, M., BAULIDA, J. & DE HERREROS, A. G. 2000. The transcription factor Snail is a repressor of E-cadherin gene expression in epithelial tumour cells. *Nature Cell Biology,* 2**,** 84-89.

BEROUKHIM, R., GETZ, G., NGHIEMPHU, L., BARRETINA, J., HSUEH, T., LINHART, D., VIVANCO, I., LEE, J. C., HUANG, J. H., ALEXANDER, S., DU, J., KAU, T., THOMAS, R. K., SHAH, K., SOTO, H., PERNER, S., PRENSNER, J., DEBIASI, R. M., DEMICHELIS, F., HATTON, C., RUBIN, M. A., GARRAWAY, L. A., NELSON, S. F., LIAU, L., MISCHEL, P. S., CLOUGHESY, T. F., MEYERSON, M., GOLUB, T. A., LANDER, E. S., MELLINGHOFF, I. K. & SELLERS, W. R. 2007. Assessing the significance of chromosomal aberrations in cancer: Methodology and application to glioma. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 20007-20012.

BOLAND, C. R., THIBODEAU, S. N., HAMILTON, S. R., SIDRANSKY, D., ESHLEMAN, J. R., BURT, R. W., MELTZER, S. J., RODRIGUEZ-BIGAS, M. A., FODDE, R., RANZANI, G. N. & SRIVASTAVA, S. 1998. A National Cancer Institute Workshop on Microsatellite Instability for Cancer Detection and Familial Predisposition: Development of International Criteria for the Determination of Microsatellite Instability in Colorectal Cancer. *Cancer Research,* 58**,** 5248-5257.

BOS, M., GARDIZI, M., SCHILDHAUS, H. U., BUETTNER, R. & WOLF, J. 2013. Activated RET and ROS: two new driver mutations in lung adenocarcinoma. *Transl Lung Cancer Res,* 2**,** 112-21.

BOS, P. D., ZHANG, X. H. F., NADAL, C., SHU, W. P., GOMIS, R. R., NGUYEN, D. X., MINN, A. J., VAN DE VIJVER, M. J., GERALD, W. L., FOEKENS, J. A. & MASSAGUE, J. 2009. Genes that mediate breast cancer metastasis to the brain. *Nature,* 459**,** 1005-U137.

BRANNON, A. R., VAKIANI, E., SYLVESTER, B. E., SCOTT, S. N., MCDERMOTT, G., SHAH, R. H., KANIA, K., VIALE, A., OSCHWALD, D. M., VACIC, V., EMDE, A. K., CERCEK, A., YAEGER, R., KEMENY, N. E., SALTZ, L. B., SHIA, J., D'ANGELICA, M. I., WEISER, M. R., SOLIT, D. B. & BERGER, M. F. 2014. Comparative sequencing analysis reveals high genomic concordance between matched primary and metastatic colorectal cancer lesions. *Genome Biol,* 15**,** 454.

BURNS, M. B., TEMIZ, N. A. & HARRIS, R. S. 2013. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nature Genetics,* 45**,** 977-+.

BUTT, A. Q. & MILLS, K. H. G. 2014. Immunosuppressive networks and checkpoints controlling antitumor immunity and their blockade in the development of cancer immunotherapeutics and vaccines. *Oncogene,* 33**,** 4623-4631.

CAI, H., KUMAR, N., BAGHERI, H. C., VON MERING, C., ROBINSON, M. D. & BAUDIS, M. 2014. Chromothripsis-like patterns are recurring but heterogeneously distributed features in a survey of 22,347 cancer genome screens. *Bmc Genomics,* 15.

CAMERER, E., QAZI, A. A., DUONG, D. N., CORNELISSEN, I., ADVINCULA, R. & COUGHLIN, S. R. 2004. Platelets, protease-activated receptors, and fibrinogen in hematogenous metastasis. *Blood,* 104**,** 397-401.

CANCER GENOME ATLAS, N. 2012. Comprehensive molecular characterization of human colon and rectal cancer. *Nature,* 487**,** 330-7.

CANCER GENOME ATLAS RESEARCH, N. 2014. Comprehensive molecular profiling of lung adenocarcinoma. *Nature,* 511**,** 543-50.

CARDEN, C. P., AGARWAL, R., SARAN, F. & JUDSON, I. R. 2008. Eligibility of patients with brain metastases for phase I trials: time for a rethink? *Lancet Oncol,* 9**,** 1012-7.

CARTER, C. L., ALLEN, C. & HENSON, D. E. 1989. Relation of Tumor Size, Lymph-Node Status, and Survival in 24,740 Breast-Cancer Cases. *Cancer,* 63**,** 181-187.

CASCINO, T. L., LEAVENGOOD, J. M., KEMENY, N. & POSNER, J. B. 1983. Brain metastases from colon cancer. *J Neurooncol,* 1**,** 203-9.

CAZIER, J. B., RAO, S. R., MCLEAN, C. M., WALKER, A. K., WRIGHT, B. J., JAEGER, E. E., KARTSONAKI, C., MARSDEN, L., YAU, C., CAMPS, C., KAISAKI, P., OXFORD-ILLUMINA, W. G. S. C., TAYLOR, J., CATTO, J. W., TOMLINSON, I. P., KILTIE, A. E. & HAMDY, F. C. 2014a. Whole-genome sequencing of bladder cancers reveals somatic CDKN1A mutations and clinicopathological associations with mutation burden. *Nat Commun,* 5**,** 3756.

CAZIER, J. B., RAO, S. R., MCLEAN, C. M., WALKER, A. L., WRIGHT, B. J., JAEGER, E. E., KARTSONAKI, C., MARSDEN, L., YAU, C., CAMPS, C., KAISAKI, P., OXFORD-ILLUMINA, W. G. S. C., TAYLOR, J., CATTO, J. W., TOMLINSON, I. P., KILTIE, A. E. & HAMDY, F. C. 2014b. Whole-genome sequencing of bladder cancers reveals somatic CDKN1A mutations and clinicopathological associations with mutation burden. *Nat Commun,* 5**,** 3756.

CERESOLI, G. L., CAPPUZZO, F., GREGORC, V., BARTOLINI, S., CRINO, L. & VILLA, E. 2004. Gefitinib in patients with brain metastases from non-small-cell lung cancer: a prospective trial. *Ann Oncol,* 15**,** 1042-7.

CHAMBERS, A. F., GROOM, A. C. & MACDONALD, I. C. 2002. Dissemination and growth of cancer cells in metastatic sites. *Nat Rev Cancer,* 2**,** 563-72.

CHAPMAN, P. B., HAUSCHILD, A., ROBERT, C., HAANEN, J. B., ASCIERTO, P., LARKIN, J., DUMMER, R., GARBE, C., TESTORI, A., MAIO, M., HOGG, D., LORIGAN, P., LEBBE, C., JOUARY, T., SCHADENDORF, D., RIBAS, A., O'DAY, S. J., SOSMAN, J. A., KIRKWOOD, J. M., EGGERMONT, A. M. M., DRENO, B., NOLOP, K., LI, J., NELSON, B., HOU, J., LEE, R. J., FLAHERTY, K. T., MCARTHUR, G. A. & GRP, B.-S. 2011. Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation. *New England Journal of Medicine,* 364**,** 2507-2516.

CHEN, K., WALLIS, J. W., MCLELLAN, M. D., LARSON, D. E., KALICKI, J. M., POHL, C. S., MCGRATH, S. D., WENDL, M. C., ZHANG, Q., LOCKE, D. P., SHI, X., FULTON, R. S., LEY, T. J., WILSON, R. K., DING, L. & MARDIS, E. R. 2009. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods,* 6**,** 677-81.

CHENG, X. & HUNG, M. C. 2007. Breast cancer brain metastases. *Cancer Metastasis Rev,* 26**,** 635-43.

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. *Nature Biotechnology,* 31**,** 213-219.

CINGOLANI, P., PLATTS, A., WANG, L. L., COON, M., NGUYEN, T., WANG, L., LAND, S. J., LU, X. Y. & RUDEN, D. M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w(1118); iso-2; iso-3. *Fly,* 6**,** 80-92.

CLAYTON, A. J., DANSON, S., JOLLY, S., RYDER, W. D. J., BURT, P. A., STEWART, A. L., WILKINSON, P. M., WELCH, R. S., MAGEE, B., WILSON, G., HOWELL, A. & WARDLEY, A. M. 2004. Incidence of cerebral metastases in patients treated with trastuzumab for metastatic breast cancer. *British Journal of Cancer,* 91**,** 639-643.

COMIJN, J., BERX, G., VERMASSEN, P., VERSCHUEREN, K., VAN GRUNSVEN, L., BRUYNEEL, E., MAREEL, M., HUYLEBROECK, D. & VAN ROY, F. 2001. The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. *Molecular Cell,* 7**,** 1267-1278.

COUSSENS, L. M., FINGLETON, B. & MATRISIAN, L. M. 2002. Matrix metalloproteinase inhibitors and cancer: trials and tribulations. *Science,* 295**,** 2387-92.

CREASMAN, W. T., MORROW, C. P., BUNDY, B. N., HOMESLEY, H. D., GRAHAM, J. E. & HELLER, P. B. 1987. Surgical pathologic spread patterns of endometrial cancer. A Gynecologic Oncology Group Study. *Cancer,* 60**,** 2035-41.

CRISSMAN, J. D., HATFIELD, J., SCHALDENBRAND, M., SLOANE, B. F. & HONN, K. V. 1985. Arrest and extravasation of B16 amelanotic melanoma in murine lungs. A light and electron microscopic study. *Lab Invest,* 53**,** 470-8.

CUIFFO, B. G., CAMPAGNE, A., BELL, G. W., LEMBO, A., ORSO, F., LIEN, E. C., BHASIN, M. K., RAIMO, M., HANSON, S. E., MARUSYK, A., EL-ASHRY, D., HEMATTI, P., POLYAK, K., MECHTA-GRIGORIOU, F., MARIANI, O., VOLINIA, S., VINCENT-SALOMON, A., TAVERNA, D. & KARNOUB, A. E. 2014. MSC-regulated microRNAs converge on the transcription factor FOXP2 and promote breast cancer metastasis. *Cell Stem Cell,* 15**,** 762-74.

DAVIES, H., BIGNELL, G. R., COX, C., STEPHENS, P., EDKINS, S., CLEGG, S., TEAGUE, J., WOFFENDIN, H., GARNETT, M. J., BOTTOMLEY, W., DAVIS, N., DICKS, N., EWING, R., FLOYD, Y., GRAY, K., HALL, S., HAWES, R., HUGHES, J., KOSMIDOU, V., MENZIES, A., MOULD, C., PARKER, A., STEVENS, C., WATT, S., HOOPER, S., WILSON, R., JAYATILAKE, H., GUSTERSON, B. A., COOPER, C., SHIPLEY, J., HARGRAVE, D., PRITCHARD-JONES, K., MAITLAND, N., CHENEVIX-TRENCH, G., RIGGINS, G. J., BIGNER, D. D., PALMIERI, G., COSSU, A., FLANAGAN, A., NICHOLSON, A., HO, J. W. C., LEUNG, S. Y., YUEN, S. T., WEBER, B. L., SIEGLER, H. F., DARROW, T. L., PATERSON, H., MARAIS, R., MARSHALL, C. J., WOOSTER, R., STRATTON, M. R. & FUTREAL, P. A. 2002. Mutations of the BRAF gene in human cancer. *Nature,* 417**,** 949-954.

DAVIES, M. A., STEMKE-HALE, K., LIN, E., TELLEZ, C., DENG, W., GOPAL, Y. N., WOODMAN, S. E., CALDERONE, T. C., JU, Z., LAZAR, A. J., PRIETO, V. G., ALDAPE, K., MILLS, G. B. & GERSHENWALD, J. E. 2009. Integrated Molecular and Clinical Analysis of AKT Activation in Metastatic Melanoma. *Clin Cancer Res,* 15**,** 7538-7546.

DAVIS, P. C., HUDGINS, P. A., PETERMAN, S. B. & HOFFMAN, J. C. 1991. Diagnosis of Cerebral Metastases - Double-Dose Delayed Ct Vs Contrast-Enhanced Mr Imaging. *American Journal of Neuroradiology,* 12**,** 293-300.

DE ANGELIS, P. M., CLAUSEN, O. P., SCHJOLBERG, A. & STOKKE, T. 1999. Chromosomal gains and losses in primary colorectal carcinomas detected by CGH and their associations with tumour DNA ploidy, genotypes and phenotypes. *Br J Cancer,* 80**,** 526-35.

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**,** 251-256.

DE CRAENE, B. & BERX, G. 2013. Regulatory networks defining EMT during cancer initiation and progression. *Nature Reviews Cancer,* 13**,** 97-110.

DELATTRE, J. Y., KROL, G., THALER, H. T. & POSNER, J. B. 1988. Distribution of brain metastases. *Arch Neurol,* 45**,** 741-4.

DELVA, E., TUCKER, D. K. & KOWALCZYK, A. P. 2009. The Desmosome. *Cold Spring Harbor Perspectives in Biology,* 1.

DESGROSELLIER, J. S. & CHERESH, D. A. 2010. Integrins in cancer: biological implications and therapeutic opportunities. *Nat Rev Cancer,* 10**,** 9-22.

DIAZ-LOPEZ, A., MORENO-BUENO, G. & CANO, A. 2014. Role of microRNA in epithelial to mesenchymal transition and metastasis and clinical perspectives. *Cancer Manag Res,* 6**,** 205-16.

DING, L., ELLIS, M. J., LI, S. Q., LARSON, D. E., CHEN, K., WALLIS, J., HARRIS, C. C., MCLELLAN, M. D., FULTON, R. S., FULTON, L. L., ABBOTT, R. M., HOOG, J., DOOLING, D. J., KOBOLDT, D. C., SCHMIDT, H., KALICKI, J., ZHANG, Q. Y., CHEN, L., LIN, L., WENDL, M. C., MCMICHAEL, J. F., MAGRINI, V. J., COOK, L., MCGRATH, S. D., VICKERY, T. L., APPELBAUM, E., DESCHRYVER, K., DAVIES, S., GUINTOLI, T., LIN, L., CROWDER, R., TAO, Y., SNIDER, J. E., SMITH, S. M., DUKES, A. F., SANDERSON, G. E., POHL, C. S., DELEHAUNTY, K. D., FRONICK, C. C., PAPE, K. A., REED, J. S., ROBINSON, J. S., HODGES, J. S., SCHIERDING, W., DEES, N. D., SHEN, D., LOCKE, D. P., WIECHERT, M. E., ELDRED, J. M., PECK, J. B., OBERKFELL, B. J., LOLOFIE, J. T., DU, F. Y., HAWKINS, A. E., O'LAUGHLIN, M. D., BERNARD, K. E., CUNNINGHAM, M., ELLIOTT, G., MASON, M. D., THOMPSON, D. M., IVANOVICH, J. L., GOODFELLOW, P. J., PEROU, C. M., WEINSTOCK, G. M., AFT, R., WATSON, M., LEY, T. J., WILSON, R. K. & MARDIS, E. R. 2010. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature,* 464**,** 999-1005.

DING, L., GETZ, G., WHEELER, D. A., MARDIS, E. R., MCLELLAN, M. D., CIBULSKIS, K., SOUGNEZ, C., GREULICH, H., MUZNY, D. M., MORGAN, M. B., FULTON, L., FULTON, R. S., ZHANG, Q., WENDL, M. C., LAWRENCE, M. S., LARSON, D. E., CHEN, K., DOOLING, D. J., SABO, A., HAWES, A. C., SHEN, H., JHANGIANI, S. N., LEWIS, L. R., HALL, O., ZHU, Y., MATHEW, T., REN, Y., YAO, J., SCHERER, S. E., CLERC, K., METCALF, G. A., NG, B., MILOSAVLJEVIC, A., GONZALEZ-GARAY, M. L., OSBORNE, J. R., MEYER, R., SHI, X., TANG, Y., KOBOLDT, D. C., LIN, L., ABBOTT, R., MINER, T. L., POHL, C., FEWELL, G., HAIPEK, C., SCHMIDT, H., DUNFORD-SHORE, B. H., KRAJA, A., CROSBY, S. D., SAWYER, C. S., VICKERY, T., SANDER, S., ROBINSON, J., WINCKLER, W., BALDWIN, J., CHIRIEAC, L. R., DUTT, A., FENNELL, T., HANNA, M., JOHNSON, B. E., ONOFRIO, R. C., THOMAS, R. K., TONON, G., WEIR, B. A., ZHAO, X., ZIAUGRA, L., ZODY, M. C., GIORDANO, T., ORRINGER, M. B., ROTH, J. A., SPITZ, M. R., WISTUBA, II, OZENBERGER, B., GOOD, P. J., CHANG, A. C., BEER, D. G., WATSON, M. A., LADANYI, M., BRODERICK, S., YOSHIZAWA, A., TRAVIS, W. D., PAO, W., PROVINCE, M. A., WEINSTOCK, G. M., VARMUS, H. E., GABRIEL, S. B., LANDER, E. S., GIBBS, R. A., MEYERSON, M. & WILSON, R. K. 2008. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature,* 455**,** 1069-75.

DISKIN, S. J., ECK, T., GRESHOCK, J., MOSSE, Y. P., NAYLOR, T., STOECKERT, C. J., WEBER, B. L., MARIS, J. M. & GRANT, G. R. 2006. STAC: A method for testing the significance of DNA copy number aberrations across multiple array-CGH experiments. *Genome Research,* 16**,** 1149-1158.

DOGAN, S., SHEN, R., ANG, D. C., JOHNSON, M. L., D'ANGELO, S. P., PAIK, P. K., BRZOSTOWSKI, E. B., RIELY, G. J., KRIS, M. G., ZAKOWSKI, M. F. & LADANYI, M. 2012. Molecular epidemiology of EGFR and KRAS mutations in 3,026 lung adenocarcinomas: higher susceptibility of women to smoking-related KRAS-mutant cancers. *Clin Cancer Res,* 18**,** 6169-77.

DURKIN, S. G. & GLOVER, T. W. 2007. Chromosome fragile sites. *Annu Rev Genet,* 41**,** 169-92.

EICHLER, A. F., CHUNG, E., KODACK, D. P., LOEFFLER, J. S., FUKUMURA, D. & JAIN, R. K. 2011. The biology of brain metastases-translation to new therapies. *Nat Rev Clin Oncol,* 8**,** 344-56.

EICHLER, A. F., KAHLE, K. T., WANG, D. L., JOSHI, V. A., WILLERS, H., ENGELMAN, J. A., LYNCH, T. J. & SEQUIST, L. V. 2010. EGFR mutation status and survival after diagnosis of brain metastasis in nonsmall cell lung cancer. *Neuro-Oncology,* 12**,** 1193-1199.

ELL, B. & KANG, Y. 2013. Transcriptional control of cancer metastasis. *Trends Cell Biol,* 23**,** 603-11.

ELLIS, C. A. & CLARK, G. 2000. The importance of being K-Ras. *Cellular Signalling,* 12**,** 425-434.

ELLIS, M. J., DING, L., SHEN, D., LUO, J. Q., SUMAN, V. J., WALLIS, J. W., VAN TINE, B. A., HOOG, J., GOIFFON, R. J., GOLDSTEIN, T. C., NG, S., LIN, L., CROWDER, R., SNIDER, J., BALLMAN, K., WEBER, J., CHEN, K., KOBOLDT, D. C., KANDOTH, C., SCHIERDING, W. S., MCMICHAEL, J. F., MILLER, C. A., LU, C., HARRIS, C. C., MCLELLAN, M. D., WENDL, M. C., DESCHRYVER, K., ALLRED, D. C., ESSERMAN, L., UNZEITIG, G., MARGENTHALER, J., BABIERA, G. V., MARCOM, P. K., GUENTHER, J. M., LEITCH, M., HUNT, K., OLSON, J., TAO, Y., MAHER, C. A., FULTON, L. L., FULTON, R. S., HARRISON, M., OBERKFELL, B., DU, F. Y., DEMETER, R., VICKERY, T. L., ELHAMMALI, A., PIWNICA-WORMS, H., MCDONALD, S., WATSON, M., DOOLING, D. J., OTA, D., CHANG, L. W., BOSE, R., LEY, T. J., PIWNICA-WORMS, D., STUART, J. M., WILSON, R. K. & MARDIS, E. R. 2012. Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature,* 486**,** 353-360.

ETTINGER, D. S., AKERLEY, W., BORGHAEI, H., CHANG, A. C., CHENEY, R. T., CHIRIEAC, L. R., D'AMICO, T. A., DEMMY, T. L., GOVINDAN, R., GRANNIS, F. W., JR., GRANT, S. C., HORN, L., JAHAN, T. M., KOMAKI, R., KONG, F. M., KRIS, M. G., KRUG, L. M., LACKNER, R. P., LENNES, I. T., LOO, B. W., JR., MARTINS, R., OTTERSON, G. A., PATEL, J. D., PINDER-SCHENCK, M. C., PISTERS, K. M., RECKAMP, K., RIELY, G. J., ROHREN, E., SHAPIRO, T. A., SWANSON, S. J., TAUER, K., WOOD, D. E., YANG, S. C., GREGORY, K., HUGHES, M. & NATIONAL COMPREHENSIVE CANCER, N. 2013. Non-small cell lung cancer, version 2.2013. *J Natl Compr Canc Netw,* 11**,** 645-53; quiz 653.

FALCHOOK, G. S., LONG, G. V., KURZROCK, R., KIM, K. B., ARKENAU, T. H., BROWN, M. P., HAMID, O., INFANTE, J. R., MILLWARD, M., PAVLICK, A. C., O'DAY, S. J., BLACKMAN, S. C., CURTIS, C. M., LEBOWITZ, P., MA, B., OUELLET, D. & KEFFORD, R. F. 2012. Dabrafenib in patients with melanoma, untreated brain metastases, and other solid tumours: a phase 1 dose-escalation trial. *Lancet,* 379**,** 1893-1901.

FEARON, E. R. 2011. Molecular genetics of colorectal cancer. *Annu Rev Pathol,* 6**,** 479-507.

FIDLER, I. J. 2003. Timeline - The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nature Reviews Cancer,* 3**,** 453-458.

FIDLER, I. J. & LIEBER, S. 1972. Quantitative analysis of the mechanism of glucocorticoid enhancement of experimental metastasis. *Res Commun Chem Pathol Pharmacol,* 4**,** 607-13.

FISHER, B., JEONG, J. H., ANDERSON, S., BRYANT, J., FISHER, E. R. & WOLMARK, N. 2002. Twenty-five-year follow-up of a randomized trial comparing radical mastectomy, total mastectomy, and total mastectomy followed by irradiation. *New England Journal of Medicine,* 347**,** 567-575.

FITZGERALD, D. P., PALMIERI, D., HUA, E., HARGRAVE, E., HERRING, J. M., QIAN, Y., VEGA-VALLE, E., WEIL, R. J., STARK, A. M., VORTMEYER, A. O. & STEEG, P. S. 2008. Reactive glia are recruited by highly proliferative brain metastases of breast cancer and promote tumor cell colonization. *Clin Exp Metastasis,* 25**,** 799-810.

FORBES, S. A., BINDAL, N., BAMFORD, S., COLE, C., KOK, C. Y., BEARE, D., JIA, M., SHEPHERD, R., LEUNG, K., MENZIES, A., TEAGUE, J. W., CAMPBELL, P. J., STRATTON, M. R. & FUTREAL, P. A. 2011. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res,* 39**,** D945-50.

FRIBERG, S. & MATTSON, S. 1997. On the growth rates of human malignant tumors: Implications for medical decision making. *Journal of Surgical Oncology,* 65**,** 284-297.

FRIDMAN, W. H., PAGES, F., SAUTES-FRIDMAN, C. & GALON, J. 2012. The immune contexture in human tumours: impact on clinical outcome. *Nature Reviews Cancer,* 12**,** 298-306.

FRIEDL, P., LOCKER, J., SAHAI, E. & SEGALL, J. E. 2012. Classifying collective cancer cell invasion. *Nature Cell Biology,* 14**,** 777-783.

FUCHS, C. S. & OGINO, S. 2013. Aspirin therapy for colorectal cancer with PIK3CA mutation: simply complex! *J Clin Oncol,* 31**,** 4358-61.

GARCIA, J., DURAN, A., TABERNERO, M. D., GARCIA PLAZA, A., FLORES CORRAL, T., NAJERA, M. L., GOMEZ-ALONSO, A. & ORFAO, A. 2003. Numerical abnormalities of chromosomes 17 and 18 in sporadic colorectal cancer: Incidence and correlation with clinical and biological findings and the prognosis of the disease. *Cytometry B Clin Cytom,* 51**,** 14-20.

GASPAR, L., SCOTT, C., ROTMAN, M., ASBELL, S., PHILLIPS, T., WASSERMAN, T., MCKENNA, W. G. & BYHARDT, R. 1997. Recursive partitioning analysis (RPA) of prognostic factors in three Radiation Therapy Oncology Group (RTOG) brain metastases trials. *Int J Radiat Oncol Biol Phys,* 37**,** 745-51.

GAY, L. J. & FELDING-HABERMANN, B. 2011. Contribution of platelets to tumour metastasis. *Nature Reviews Cancer,* 11**,** 123-134.

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. *New England Journal of Medicine,* 366**,** 883-892.

GIAMPIERI, S., MANNING, C., HOOPER, S., JONES, L., HILL, C. S. & SAHAI, E. 2009. Localized and reversible TGFbeta signalling switches breast cancer cells from cohesive to single cell motility. *Nat Cell Biol,* 11**,** 1287-96.

GIL-GIL, M. J., MARTINEZ-GARCIA, M., SIERRA, A., CONESA, G., DEL BARCO, S., GONZALEZ-JIMENEZ, S. & VILLA, S. 2014. Breast cancer brain metastases: a review of the literature and a current multidisciplinary management guideline. *Clinical & Translational Oncology,* 16**,** 436-446.

GILBERT, M. R. 2001. Brain metastases: still an 'orphan' disease? *Curr Oncol Rep,* 3**,** 463-6.

GILBERT, M. T. P., HASELKORN, T., BUNCE, M., SANCHEZ, J. J., LUCAS, S. B., JEWELL, L. D., VAN MARCK, E. & WOROBEY, M. 2007. The Isolation of Nucleic Acids from Fixed, Paraffin-Embedded Tissues-Which Methods Are Useful When? *Plos One,* 2.

GOLDSTRAW, P., CROWLEY, J., CHANSKY, K., GIROUX, D. J., GROOME, P. A., RAMI-PORTA, R., POSTMUS, P. E., RUSCH, V., SOBIN, L., INTERNATIONAL ASSOCIATION FOR THE STUDY OF LUNG CANCER INTERNATIONAL STAGING, C. & PARTICIPATING, I. 2007. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours. *J Thorac Oncol,* 2**,** 706-14.

GONZALEZ-PEREZ, A., PEREZ-LLAMAS, C., DEU-PONS, J., TAMBORERO, D., SCHROEDER, M. P., JENE-SANZ, A., SANTOS, A. & LOPEZ-BIGAS, N. 2013. IntOGen-mutations identifies cancer drivers across tumor types. *Nature Methods,* 10**,** 1081-1082.

GOODENOUGH, D. A. & PAUL, D. L. 2009. Gap Junctions. *Cold Spring Harbor Perspectives in Biology,* 1.

GOVINDAN, R., DING, L., GRIFFITH, M., SUBRAMANIAN, J., DEES, N. D., KANCHI, K. L., MAHER, C. A., FULTON, R., FULTON, L., WALLIS, J., CHEN, K., WALKER, J., MCDONALD, S., BOSE, R., ORNITZ, D., XIONG, D., YOU, M., DOOLING, D. J., WATSON, M., MARDIS, E. R. & WILSON, R. K. 2012. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell,* 150**,** 1121-34.

GOW, C. H., CHANG, Y. L., HSU, Y. C., TSAI, M. F., WU, C. T., YU, C. J., YANG, C. H., LEE, Y. C., YANG, P. C. & SHIH, J. Y. 2009. Comparison of epidermal growth factor receptor mutations between primary and corresponding metastatic tumors in tyrosine kinase inhibitor-naive non-small-cell lung cancer. *Annals of Oncology,* 20**,** 696-702.

GREAVES, M. & MALEY, C. C. 2012. Clonal evolution in cancer. *Nature,* 481**,** 306-313.

GRINBERG-RASHI, H., OFEK, E., PERELMAN, M., SKARDA, J., YARON, P., HAJDUCH, M., JACOB-HIRSCH, J., AMARIGLIO, N., KRUPSKY, M., SIMANSKY, D. A., RAM, Z., PFEFFER, R., GALERNTER, I., STEINBERG, D. M., BEN-DOV, I., RECHAVI, G. & IZRAELI, S. 2009. The Expression of Three Genes in Primary Non-Small Cell Lung Cancer Is Associated with Metastatic Spread to the Brain. *Clinical Cancer Research,* 15**,** 1755-1761.

GROMMES, C., OXNARD, G. R., KRIS, M. G., MILLER, V. A., PAO, W., HOLODNY, A. I., CLARKE, J. L. & LASSMAN, A. B. 2011. "Pulsatile" high-dose weekly erlotinib for CNS metastases from EGFR mutant non-small cell lung cancer. *Neuro-Oncology,* 13**,** 1364-1369.

GUPTA, G. P. & MASSAGUE, J. 2006. Cancer metastasis: Building a framework. *Cell,* 127**,** 679-695.

GUTENBERG, A., GERDES, J. S., JUNG, K., SANDER, B., GUNAWAN, B., BOCK, H. C., LIERSCH, T., BRUCK, W., ROHDE, V. & FUZESI, L. 2010. High chromosomal instability in brain metastases of colorectal carcinoma. *Cancer Genetics and Cytogenetics,* 198**,** 47-51.

HAJRA, K. M., CHEN, D. Y. S. & FEARON, E. R. 2002. The SLUG zinc-finger protein represses E-cadherin in breast cancer. *Cancer Research,* 62**,** 1613-1618.

HANIBUCHI, M., KIM, S. J., FIDLER, I. J. & NISHIOKA, Y. 2014. The molecular biology of lung cancer brain metastasis: an overview of current comprehensions and future perspectives. *J Med Invest,* 61**,** 241-53.

HART, I. R. & FIDLER, I. J. 1980. Role of organ selectivity in the determination of metastatic patterns of B16 melanoma. *Cancer Res,* 40**,** 2281-7.

HEINTZ, A. P., ODICINO, F., MAISONNEUVE, P., QUINN, M. A., BENEDET, J. L., CREASMAN, W. T., NGAN, H. Y., PECORELLI, S. & BELLER, U. 2006. Carcinoma of the ovary. FIGO 26th Annual Report on the Results of Treatment in Gynecological Cancer. *Int J Gynaecol Obstet,* 95 Suppl 1**,** S161-92.

HERBST, R. S., HEYMACH, J. V. & LIPPMAN, S. M. 2008. Lung cancer. *N Engl J Med,* 359**,** 1367-80.

HESS, K. R., VARADHACHARY, G. R., TAYLOR, S. H., WEI, W., RABER, M. N., LENZI, R. & ABBRUZZESE, J. L. 2006. Metastatic patterns in adenocarcinoma. *Cancer,* 106**,** 1624-1633.

HOFREITER, M., JAENICKE, V., SERRE, D., VON HAESELER, A. & PAABO, S. 2001. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research,* 29**,** 4793-4799.

HORNER, D. S., PAVESI, G., CASTRIGNANO, T., DE MEO, P. D., LIUNI, S., SAMMETH, M., PICARDI, E. & PESOLE, G. 2010. Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. *Brief Bioinform,* 11**,** 181-97.

HU, M., YAO, J., CARROLL, D. K., WEREMOWICZ, S., CHEN, H., CARRASCO, D., RICHARDSON, A., VIOLETTE, S., NIKOLSKAYA, T., NIKOLSKY, Y., BAUERLEIN, E. L., HAHN, W. C., GELMAN, R. S., ALLRED, C., BISSELL, M. J., SCHNITT, S. & POLYAK, K. 2008. Regulation of in situ to invasive breast carcinoma transition. *Cancer Cell,* 13**,** 394-406.

HUANG, R. Y., GUILFORD, P. & THIERY, J. P. 2012. Early events in cell adhesion and polarity during epithelial-mesenchymal transition. *J Cell Sci,* 125**,** 4417-22.

HUSEMANN, Y., GEIGL, J. B., SCHUBERT, F., MUSIANI, P., MEYER, M., BURGHART, E., FORNI, G., EILS, R., FEHM, T., RIETHMULLER, G. & KLEIN, C. A. 2008. Systemic spread is an early step in breast cancer. *Cancer Cell,* 13**,** 58-68.

IMIELINSKI, M., BERGER, A. H., HAMMERMAN, P. S., HERNANDEZ, B., PUGH, T. J., HODIS, E., CHO, J., SUH, J., CAPELLETTI, M., SIVACHENKO, A., SOUGNEZ, C., AUCLAIR, D., LAWRENCE, M. S., STOJANOV, P., CIBULSKIS, K., CHOI, K., DE WAAL, L., SHARIFNIA, T., BROOKS, A., GREULICH, H., BANERJI, S., ZANDER, T., SEIDEL, D., LEENDERS, F., ANSEN, S., LUDWIG, C., ENGEL-RIEDEL, W., STOELBEN, E., WOLF, J., GOPARJU, C., THOMPSON, K., WINCKLER, W., KWIATKOWSKI, D., JOHNSON, B. E., JANNE, P. A., MILLER, V. A., PAO, W., TRAVIS, W. D., PASS, H. I., GABRIEL, S. B., LANDER, E. S., THOMAS, R. K., GARRAWAY, L. A., GETZ, G. & MEYERSON, M. 2012. Mapping the Hallmarks of Lung Adenocarcinoma with Massively Parallel Sequencing. *Cell,* 150**,** 1107-1120.

JEANES, A., GOTTARDI, C. J. & YAP, A. S. 2008. Cadherins and cancer: how does cadherin dysfunction promote tumor progression? *Oncogene,* 27**,** 6920-6929.

JONES, A. M., DOUGLAS, E. J., HALFORD, S. E., FIEGLER, H., GORMAN, P. A., ROYLANCE, R. R., CARTER, N. P. & TOMLINSON, I. P. 2005. Array-CGH analysis of microsatellite-stable, near-diploid bowel cancers and comparison with other types of colorectal carcinoma. *Oncogene,* 24**,** 118-29.

JONES, S., LI, M., PARSONS, D. W., ZHANG, X. S., WESSELING, J., KRISTEL, P., SCHMIDT, M. K., MARKOWITZ, S., YAN, H., BIGNER, D., HRUBAN, R. H., ESHLEMAN, J. R., IACOBUZIO-DONAHUE, C. A., GOGGINS, M., MAITRA, A., MALEK, S. N., POWELL, S., VOGELSTEIN, B., KINZLER, K. W., VELCULESCU, V. E. & PAPADOPOULOS, N. 2012. Somatic Mutations in the Chromatin Remodeling Gene ARID1A Occur in Several Tumor Types. *Human Mutation,* 33**,** 100-103.

JUANYIN, J., TRACY, K., ZHANG, L., MUNASINGHE, J., SHAPIRO, E., KORETSKY, A. & KELLY, K. 2009. Noninvasive imaging of the functional effects of anti-VEGF therapy on tumor cell extravasation and regional blood volume in an experimental brain metastasis model. *Clin Exp Metastasis,* 26**,** 403-14.

JUNTTILA, M. R. & DE SAUVAGE, F. J. 2013. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature,* 501**,** 346-354.

KAGEY, M. H., NEWMAN, J. J., BILODEAU, S., ZHAN, Y., ORLANDO, D. A., VAN BERKUM, N. L., EBMEIER, C. C., GOOSSENS, J., RAHL, P. B., LEVINE, S. S., TAATJES, D. J., DEKKER, J. & YOUNG, R. A. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature,* 467**,** 430-5.

KALLURI, R. & WEINBERG, R. A. 2009. The basics of epithelial-mesenchymal transition. *Journal of Clinical Investigation,* 119**,** 1420-1428.

KATAINEN, R., DAVE, K., PITKANEN, E., PALIN, K., KIVIOJA, T., VALIMAKI, N., GYLFE, A. E., RISTOLAINEN, H., HANNINEN, U. A., CAJUSO, T., KONDELIN, J., TANSKANEN, T., MECKLIN, J. P., JARVINEN, H., RENKONEN-SINISALO, L., LEPISTO, A., KAASINEN, E., KILPIVAARA, O., TUUPANEN, S., ENGE, M., TAIPALE, J. & AALTONEN, L. A. 2015. CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat Genet,* 47**,** 818-21.

KERICK, M., ISAU, M., TIMMERMANN, B., SULTMANN, H., HERWIG, R., KROBITSCH, S., SCHAEFER, G., VERDORFER, I., BARTSCH, G., KLOCKER, H., LEHRACH, H. & SCHWEIGER, M. R. 2011. Targeted high throughput sequencing in clinical cancer Settings: formaldehyde fixed-paraffin embedded (FFPE) tumor tissues, input amount and tumor heterogeneity. *Bmc Medical Genomics,* 4.

KESSENBROCK, K., PLAKS, V. & WERB, Z. 2010. Matrix Metalloproteinases: Regulators of the Tumor Microenvironment. *Cell,* 141**,** 52-67.

KIENAST, Y., VON BAUMGARTEN, L., FUHRMANN, M., KLINKERT, W. E. F., GOLDBRUNNER, R., HERMS, J. & WINKLER, F. 2010. Real-time imaging reveals the single steps of brain metastasis formation. *Nature Medicine,* 16**,** 116-U157.

KIM, L. S., HUANG, S. Y., LU, W. X., LEV, D. C. & PRICE, J. E. 2004. Vascular endothelial growth factor expression promotes the growth of breast cancer brain metastases in nude mice. *Clinical & Experimental Metastasis,* 21**,** 107-118.

KIM, S. J., KIM, J. S., PARK, E. S., LEE, J. S., LIN, Q., LANGLEY, R. R., MAYA, M., HE, J., KIM, S. W., WEIHUA, Z., BALASUBRAMANIAN, K., FAN, D., MILLS, G. B., HUNG, M. C. & FIDLER, I. J. 2011. Astrocytes upregulate survival genes in tumor cells and induce protection from chemotherapy. *Neoplasia,* 13**,** 286-98.

KITCHENER, H., SWART, A. M. C., QIAN, W., AMOS, C. & PARMER, M. K. B. 2009. Efficacy of systematic pelvic lymphadenectomy in endometrial cancer (MRC ASTEC trial): a randomised study. (vol 373, pg 125, 2009). *Lancet,* 373**,** 1764-1764.

KLEIN, C. A. 2009. Parallel progression of primary tumours and metastases. *Nat Rev Cancer,* 9**,** 302-12.

KLOOSTERMAN, W. P., HOOGSTRAAT, M., PALING, O., TAVAKOLI-YARAKI, M., RENKENS, I., VERMAAT, J. S., VAN ROOSMALEN, M. J., VAN LIESHOUT, S., NIJMAN, I. J., ROESSINGH, W., VAN 'T SLOT, R., VAN DE BELT, J., GURYEV, V., KOUDIJS, M., VOEST, E. & CUPPEN, E. 2011. Chromothripsis is a common mechanism driving genomic rearrangements in primary and metastatic colorectal cancer. *Genome Biology,* 12.

KOBOLDT, D. C., FULTON, R. S., MCLELLAN, M. D., SCHMIDT, H., KALICKI-VEIZER, J., MCMICHAEL, J. F., FULTON, L. L., DOOLING, D. J., DING, L., MARDIS, E. R., WILSON, R. K., ALLY, A., BALASUNDARAM, M., BUTTERFIELD, Y. S. N., CARLSEN, R., CARTER, C., CHU, A., CHUAH, E., CHUN, H. J. E., COOPE, R. J. N., DHALLA, N., GUIN, R., HIRST, C., HIRST, M., HOLT, R. A., LEE, D., LI, H. Y. I., MAYO, M., MOORE, R. A., MUNGALL, A. J., PLEASANCE, E., ROBERTSON, A. G., SCHEIN, J. E., SHAFIEI, A., SIPAHIMALANI, P., SLOBODAN, J. R., STOLL, D., TAM, A., THIESSEN, N., VARHOL, R. J., WYE, N., ZENG, T., ZHAO, Y. J., BIROL, I., JONES, S. J. M., MARRA, M. A., CHERNIACK, A. D., SAKSENA, G., ONOFRIO, R. C., PHO, N. H., CARTER, S. L., SCHUMACHER, S. E., TABAK, B., HERNANDEZ, B., GENTRY, J., NGUYEN, H., CRENSHAW, A., ARDLIE, K., BEROUKHIM, R., WINCKLER, W., GETZ, G., GABRIEL, S. B., MEYERSON, M., CHIN, L., PARK, P. J., KUCHERLAPATI, R., HOADLEY, K. A., AUMAN, J. T., FAN, C., TURMAN, Y. J., SHI, Y., LI, L., TOPAL, M. D., HE, X. P., CHAO, H. H., PRAT, A., SILVA, G. O., IGLESIA, M. D., ZHAO, W., USARY, J., BERG, J. S., ADAMS, M., BOOKER, J., WU, J. Y., GULABANI, A., BODENHEIMER, T., HOYLE, A. P., SIMONS, J. V., SOLOWAY, M. G., MOSE, L. E., JEFFERYS, S. R., BALU, S., PARKER, J. S., HAYES, D. N., PEROU, C. M., MALIK, S., MAHURKAR, S., SHEN, H., WEISENBERGER, D. J., TRICHE, T., et al. 2012. Comprehensive molecular portraits of human breast tumours. *Nature,* 490**,** 61-70.

KOCHER, M., SOFFIETTI, R., ABACIOGLU, U., VILLA, S., FAUCHON, F., BAUMERT, B. G., FARISELLI, L., TZUK-SHINA, T., KORTMANN, R. D., CARRIE, C., BEN HASSEL, M., KOURI, M., VALEINIS, E., VAN DEN BERGE, D., COLLETTE, S., COLLETTE, L. & MUELLER, R. P. 2011. Adjuvant whole-brain radiotherapy versus observation after radiosurgery or surgical resection of one to three cerebral metastases: results of the EORTC 22952-26001 study. *J Clin Oncol,* 29**,** 134-41.

KOVAC, M., NAVAS, C., HORSWELL, S., SALM, M., BARDELLA, C., ROWAN, A., STARES, M., CASTRO-GINER, F., FISHER, R., DE BRUIN, E. C., KOVACOVA, M., GORMAN, M., MAKINO, S., WILLIAMS, J., JAEGER, E., JONES, A., HOWARTH, K., LARKIN, J., PICKERING, L., GORE, M., NICOL, D. L., HAZELL, S., STAMP, G., O'BRIEN, T., CHALLACOMBE, B., MATTHEWS, N., PHILLIMORE, B., BEGUM, S., RABINOWITZ, A., VARELA, I., CHANDRA, A., HORSFIELD, C., POLSON, A., TRAN, M., BHATT, R., TERRACCIANO, L., EPPENBERGER-CASTORI, S., PROTHEROE, A., MAHER, E., EL BAHRAWY, M., FLEMING, S., RATCLIFFE, P., HEINIMANN, K., SWANTON, C. & TOMLINSON, I. 2015. Recurrent chromosomal gains and heterogeneous driver mutations characterise papillary renal cancer evolution. *Nat Commun,* 6**,** 6336.

KUONG, K. J. & LOEB, L. A. 2013. Apobec3b Mutagenesis in Cancer. *Nature Genetics,* 45**,** 964-965.

LAMOUILLE, S., XU, J. & DERYNCK, R. 2014. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol,* 15**,** 178-96.

LANGLEY, R. R., FAN, D., GUO, L., ZHANG, C., LIN, Q., BRANTLEY, E. C., MCCARTY, J. H. & FIDLER, I. J. 2009. Generation of an immortalized astrocyte cell line from H-2Kb-tsA58 mice to study the role of astrocytes in brain metastasis. *Int J Oncol,* 35**,** 665-72.

LARSEN, M. J., THOMASSEN, M., GERDES, A. M. & KRUSE, T. A. 2014. Hereditary breast cancer: clinical, pathological and molecular characteristics. *Breast Cancer (Auckl),* 8**,** 145-55.

LARSON, D. E., HARRIS, C. C., CHEN, K., KOBOLDT, D. C., ABBOTT, T. E., DOOLING, D. J., LEY, T. J., MARDIS, E. R., WILSON, R. K. & DING, L. 2012. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics,* 28**,** 311-317.

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. H., 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.

LEBLEU, V. S., MACDONALD, B. & KALLURI, R. 2007. Structure and function of basement membranes. *Exp Biol Med (Maywood),* 232**,** 1121-9.

LEE, W., JIANG, Z. S., LIU, J. F., HAVERTY, P. M., GUAN, Y. H., STINSON, J., YUE, P., ZHANG, Y., PANT, K. P., BHATT, D., HA, C., JOHNSON, S., KENNEMER, M. I., MOHAN, S., NAZARENKO, I., WATANABE, C., SPARKS, A. B., SHAMES, D. S., GENTLEMAN, R., DE SAUVAGE, F. J., STERN, H., PANDITA, A., BALLINGER, D. G., DRMANAC, R., MODRUSAN, Z., SESHAGIRI, S. & ZHANG, Z. M. 2010. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature,* 465**,** 473-477.

LEHEMBRE, F., YILMAZ, M., WICKI, A., SCHOMBER, T., STRITTMATTER, K., ZIEGLER, D., KREN, A., WENT, P., DERKSEN, P. W. B., BERNS, A., JONKERS, J. & CHRISTOFORI, G. 2008. NCAM-induced focal adhesion assembly: a functional switch upon loss of E-cadherin. *Embo Journal,* 27**,** 2603-2615.

LEYLAND-JONES, B. 2009. Human epidermal growth factor receptor 2-positive breast cancer and central nervous system metastases. *J Clin Oncol,* 27**,** 5278-86.

LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 1754-60.

LIN, C. W., CHANG, Y. L., CHANG, Y. C., LIN, J. C., CHEN, C. C., PAN, S. H., WU, C. T., CHEN, H. Y., YANG, S. C., HONG, T. M. & YANG, P. C. 2013. MicroRNA-135b promotes lung cancer metastasis by regulating multiple targets in the Hippo pathway and LZTS1. *Nat Commun,* 4**,** 1877.

LIN, N. U. 2014. Targeted Therapies in Brain Metastases. *Current Treatment Options in Neurology,* 16.

LIN, N. U. & WINER, E. P. 2007a. Brain metastases: The HER2 paradigm. *Clinical Cancer Research,* 13**,** 1648-1655.

LIN, N. U. & WINER, E. P. 2007b. Brain metastases: the HER2 paradigm. *Clin Cancer Res,* 13**,** 1648-55.

LIN, Q., BALASUBRAMANIAN, K., FAN, D., KIM, S. J., GUO, L., WANG, H., BAR-ELI, M., ALDAPE, K. D. & FIDLER, I. J. 2010. Reactive astrocytes protect melanoma cells from chemotherapy by sequestering intracellular calcium through gap junction communication channels. *Neoplasia,* 12**,** 748-54.

LIN, Y. C., YANG, Z., XU, A., DONG, P., HUANG, Y., LIU, H., LI, F. D., WANG, H. F., XU, Q., WANG, Y. Q., SUN, D., ZOU, Y., ZOU, X. W., WANG, Y., ZHANG, D., LIU, H. J., WU, X., ZHANG, M., FU, Y., CAI, Z. M., LIU, C. X. & WU, S. 2015. PIK3R1 negatively regulates the epithelial-mesenchymal transition and stem-like phenotype of renal cancer cells through the AKT/GSK3 beta/CTNNB1 signaling pathway. *Scientific Reports,* 5.

LIPPITZ, B., LINDQUIST, C., PADDICK, I., PETERSON, D., O'NEILL, K. & BEANEY, R. 2014. Stereotactic radiosurgery in the treatment of brain metastases: the current evidence. *Cancer Treat Rev,* 40**,** 48-59.

LIU, L., LI, Y., LI, S., HU, N., HE, Y., PONG, R., LIN, D., LU, L. & LAW, M. 2012a. Comparison of next-generation sequencing systems. *J Biomed Biotechnol,* 2012**,** 251364.

LIU, P. F., CARVALHO, C. M. B., HASTINGS, P. J. & LUPSKI, J. R. 2012b. Mechanisms for recurrent and complex human genomic rearrangements. *Current Opinion in Genetics & Development,* 22**,** 211-220.

LIU, P. Y., MORRISON, C., WANG, L., XIONG, D. H., VEDELL, P., CUI, P., HUA, X., DING, F., LU, Y., JAMES, M., EBBEN, J. D., XU, H. M., ADJEI, A. A., HEAD, K., ANDRAE, J., TSCHANNEN, M. R., JACOB, H., PAN, J., ZHANG, Q., VAN DEN BERGH, F., XIAO, H. J., LO, K. C., PATEL, J., RICHMOND, T., WATT, M. A., ALBERT, T., SELZER, R., ANDERSON, M., WANG, J., WANG, Y. A., STARNES, S., YANG, P. & YOU, M. 2012c. Identification of somatic mutations in non-small cell lung carcinomas using whole-exome sequencing. *Carcinogenesis,* 33**,** 1270-1276.

LIU, Y. N., YIN, J. J., ABOU-KHEIR, W., HYNES, P. G., CASEY, O. M., FANG, L., YI, M., STEPHENS, R. M., SENG, V., SHEPPARD-TILLMAN, H., MARTIN, P. & KELLY, K. 2013. MiR-1 and miR-200 inhibit EMT via Slug-dependent and tumorigenesis via Slug-independent mechanisms. *Oncogene,* 32**,** 296-306.

LOCKMAN, P. R., MITTAPALLI, R. K., TASKAR, K. S., RUDRARAJU, V., GRIL, B., BOHN, K. A., ADKINS, C. E., ROBERTS, A., THORSHEIM, H. R., GAASCH, J. A., HUANG, S. Y., PALMIERI, D., STEEG, P. S. & SMITH, Q. R. 2010. Heterogeneous Blood-Tumor Barrier Permeability Determines Drug Efficacy in Experimental Brain Metastases of Breast Cancer. *Clinical Cancer Research,* 16**,** 5664-5678.

LONG, G. V., TREFZER, U., DAVIES, M. A., KEFFORD, R. F., ASCIERTO, P. A., CHAPMAN, P. B., PUZANOV, I., HAUSCHILD, A., ROBERT, C., ALGAZI, A., MORTIER, L., TAWBI, H., WILHELM, T., ZIMMER, L., SWITZKY, J., SWANN, S., MARTIN, A. M., GUCKERT, M., GOODMAN, V., STREIT, M., KIRKWOOD, J. M. & SCHADENDORF, D. 2012. Dabrafenib in patients with Val600Glu or Val600Lys BRAF-mutant melanoma metastatic to the brain (BREAK-MB): a multicentre, open-label, phase 2 trial. *Lancet Oncology,* 13**,** 1087-1095.

LORGER, M. & FELDING-HABERMANN, B. 2010. Capturing changes in the brain microenvironment during initial steps of breast cancer brain metastasis. *Am J Pathol,* 176**,** 2958-71.

LUNTER, G. & GOODSON, M. 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res,* 21**,** 936-9.

MA, L., REINHARDT, F., PAN, E., SOUTSCHEK, J., BHAT, B., MARCUSSON, E. G., TERUYA-FELDSTEIN, J., BELL, G. W. & WEINBERG, R. A. 2010. Therapeutic silencing of miR-10b inhibits metastasis in a mouse mammary tumor model. *Nature Biotechnology,* 28**,** 341-U67.

MA, L., TERUYA-FELDSTEIN, J. & WEINBERG, R. A. 2007. Tumour invasion and metastasis initiated by microRNA 10b in breast cancer. *Nature,* 449**,** 682-U2.

MARDIS, E. R. 2013. Next-Generation Sequencing Platforms. *Annual Review of Analytical Chemistry, Vol 6,* 6**,** 287-303.

MARTIN, S. A., LORD, C. J. & ASHWORTH, A. 2010. Therapeutic Targeting of the DNA Mismatch Repair Pathway. *Clinical Cancer Research,* 16**,** 5107-5113.

MASSARELLI, E., VARELLA-GARCIA, M., TANG, X. M., XAVIER, A. C., OZBURN, N. C., LIU, D. D., BEKELE, B. N., HERBST, R. S. & WISTUBA, I. I. 2007. KRAS mutation is an important predictor of resistance to therapy with epidermal growth factor receptor tyrosine kinase inhibitors in non-small cell lung cancer. *Clinical Cancer Research,* 13**,** 2890-2896.

MCBRIDE, K. A., BALLINGER, M. L., KILLICK, E., KIRK, J., TATTERSALL, M. H. N., EELES, R. A., THOMAS, D. M. & MITCHELL, G. 2014. Li-Fraumeni syndrome: cancer risk assessment and clinical management. *Nature Reviews Clinical Oncology,* 11**,** 260-271.

MCCAFFREY, L. M. & MACARA, I. G. 2011. Epithelial organization, cell polarity and tumorigenesis. *Trends Cell Biol,* 21**,** 727-35.

MELCHOR, L. & BENITEZ, J. 2013. The complex genetic landscape of familial breast cancer. *Hum Genet,* 132**,** 845-63.

METZKER, M. L. 2010. Sequencing technologies - the next generation. *Nat Rev Genet,* 11**,** 31-46.

MEYERSON, M., GABRIEL, S. & GETZ, G. 2010. Advances in understanding cancer genomes through second-generation sequencing. *Nature Reviews Genetics,* 11**,** 685-696.

MINN, A. J., GUPTA, G. P., PADUA, D., BOS, P., NGUYEN, D. X., NUYTEN, D., KREIKE, B., ZHANG, Y., WANG, Y. X., ISHWARAN, H., FOEKENS, J. A., VAN DE VIJVER, M. & MASSAGUE, J. 2007. Lung

metastasis genes couple breast tumor size and metastatic spread. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 6740-6745.

MOES, M., LE BECHEC, A., CRESPO, I., LAURINI, C., HALAVATYI, A., VETTER, G., DEL SOL, A. & FRIEDERICH, E. 2012. A Novel Network Integrating a miRNA-203/SNAI1 Feedback Loop which Regulates Epithelial to Mesenchymal Transition. *Plos One,* 7.

MOK, T. S., WU, Y. L., THONGPRASERT, S., YANG, C. H., CHU, D. T., SAIJO, N., SUNPAWERAVONG, P., HAN, B. H., MARGONO, B., ICHINOSE, Y., NISHIWAKI, Y., OHE, Y., YANG, J. J., CHEWASKULYONG, B., JIANG, H. Y., DUFFIELD, E. L., WATKINS, C. L., ARMOUR, A. A. & FUKUOKA, M. 2009. Gefitinib or Carboplatin-Paclitaxel in Pulmonary Adenocarcinoma. *New England Journal of Medicine,* 361**,** 947-957.

MUNDY, G. R. 2002. Metastasis to bone: causes, consequences and therapeutic opportunities. *Nat Rev Cancer,* 2**,** 584-93.

MUNFUS-MCCRAY, D., HARADA, S., ADAMS, C., ASKIN, F., CLARK, D., GABRIELSON, E. & LI, Q. K. 2011. EGFR and KRAS mutations in metastatic Lung adenocarcinomas. *Human Pathology,* 42**,** 1447-1453.

MUZNY, D. M., BAINBRIDGE, M. N., CHANG, K., DINH, H. H., DRUMMOND, J. A., FOWLER, G., KOVAR, C. L., LEWIS, L. R., MORGAN, M. B., NEWSHAM, I. F., REID, J. G., SANTIBANEZ, J., SHINBROT, E., TREVINO, L. R., WU, Y. Q., WANG, M., GUNARATNE, P., DONEHOWER, L. A., CREIGHTON, C. J., WHEELER, D. A., GIBBS, R. A., LAWRENCE, M. S., VOET, D., JING, R., CIBULSKIS, K., SIVACHENKO, A., STOJANOV, P., MCKENNA, A., LANDER, E. S., GABRIEL, S., GETZ, G., DING, L., FULTON, R. S., KOBOLDT, D. C., WYLIE, T., WALKER, J., DOOLING, D. J., FULTON, L., DELEHAUNTY, K. D., FRONICK, C. C., DEMETER, R., MARDIS, E. R., WILSON, R. K., CHU, A., CHUN, H. J. E., MUNGALL, A. J., PLEASANCE, E., ROBERTSON, A. G., STOLL, D., BALASUNDARAM, M., BIROL, I., BUTTERFIELD, Y. S. N., CHUAH, E., COOPE, R. J. N., DHALLA, N., GUIN, R., HIRST, C., HIRST, M., HOLT, R. A., LEE, D., LI, H. I., MAYO, M., MOORE, R. A., SCHEIN, J. E., SLOBODAN, J. R., TAM, A., THIESSEN, N., VARHOL, R., ZENG, T., ZHAO, Y., JONES, S. J. M., MARRA, M. A., BASS, A. J., RAMOS, A. H., SAKSENA, G., CHERNIACK, A. D., SCHUMACHER, S. E., TABAK, B., CARTER, S. L., PHO, N. H., NGUYEN, H., ONOFRIO, R. C., CRENSHAW, A., ARDLIE, K., BEROUKHIM, R., WINCKLER, W., GETZ, G., MEYERSON, M., PROTOPOPOV, A., ZHANG, J., HADJIPANAYIS, A., LEE, E., XI, R., YANG, L., REN, X., ZHANG, H., SATHIAMOORTHY, N., SHUKLA, S., CHEN, P. C., HASELEY, P., et al. 2012. Comprehensive molecular characterization of human colon and rectal cancer. *Nature,* 487**,** 330-337.

NAROD, S. A. 2012. Tumour size predicts long-term survival among women with lymph node-positive breast cancer. *Current Oncology,* 19**,** 249-+.

NEUWELT, E. A. 2004. Mechanisms of disease: the blood-brain barrier. *Neurosurgery,* 54**,** 131-40; discussion 141-2.

NG, P. C. & HENIKOFF, S. 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res,* 31**,** 3812-4.

NGUYEN, D. X., BOS, P. D. & MASSAGUE, J. 2009a. Metastasis: from dissemination to organ-specific colonization. *Nature Reviews Cancer,* 9**,** 274-U65.

NGUYEN, D. X., CHIANG, A. C., ZHANG, X. H. F., KIM, J. Y., KRIS, M. G., LADANYI, M., GERALD, W. L. & MASSAGUE, J. 2009b. WNT/TCF Signaling through LEF1 and HOXB9 Mediates Lung Adenocarcinoma Metastasis. *Cell,* 138**,** 51-62.

NIEDER, C., SPANNE, O., MEHTA, M. P., GROSU, A. L. & GEINITZ, H. 2011. Presentation, patterns of care, and survival in patients with brain metastases: what has changed in the last 20 years? *Cancer,* 117**,** 2505-12.

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. M., APARICIO, S. A. J. R., BORG, A., SALOMON, A. V., THOMAS, G., BORRESEN-DALE, A. L., RICHARDSON, A. L., NEUBERGER, M. S., FUTREAL, P. A., CAMPBELL, P. J., STRATTON, M. R. & CONSORTIUM, I. C. G. 2012. Mutational Processes Molding the Genomes of 21 Breast Cancers. *Cell,* 149**,** 979-993.

NIU, B. F., YE, K., ZHANG, Q. Y., LU, C., XIE, M. C., MCLELLAN, M. D., WENDL, M. C. & DING, L. 2014. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics,* 30**,** 1015-1016.

NORDEN, A. D., WEN, P. Y. & KESARI, S. 2005. Brain metastases. *Current Opinion in Neurology,* 18**,** 654-661.

OLIVIER, M., LANGEROD, A., CARRIERI, P., BERGH, J., KLAAR, S., EYFJORD, J., THEILLET, C., RODRIGUEZ, C., LIDEREAU, R., BIECHE, I., VARLEY, J., BIGNON, Y., UHRHAMMER, N., WINQVIST, R., JUKKOLA-VUORINEN, A., NIEDERACHER, D., KATO, S., ISHIOKA, C., HAINAUT, P. & BORRESEN-DALE, A. L. 2006. The clinical value of somatic TP53 gene mutations in 1,794 patients with breast cancer. *Clin Cancer Res,* 12**,** 1157-67.

OUDEJANS, J. J., SLEBOS, R. J. C., ZOETMULDER, F. A. N., MOOI, W. J. & RODENHUIS, S. 1991. Differential Activation of Ras Genes by Point Mutation in Human Colon Cancer with Metastases to Either Lung or Liver. *International Journal of Cancer,* 49**,** 875-879.

OUYANG, L., LEE, J., PARK, C. K., MAO, M., SHI, Y., GONG, Z., ZHENG, H., LI, Y., ZHAO, Y., WANG, G., FU, H., KIM, J. & LIM, H. Y. 2014. Whole-genome sequencing of matched primary and metastatic hepatocellular carcinomas. *BMC Med Genomics,* 7**,** 2.

OWONIKOKO, T. K., ARBISER, J., ZELNAK, A., SHU, H. K. G., SHIM, H., ROBIN, A. M., KALKANIS, S. N., WHITSETT, T. G., SALHIA, B., TRAN, N. L., RYKEN, T., MOORE, M. K., EGAN, K. M. & OLSON, J. J. 2014. Current approaches to the treatment of metastatic brain tumours. *Nature Reviews Clinical Oncology,* 11**,** 203-222.

PADUA, D., ZHANG, X. H., WANG, Q., NADAL, C., GERALD, W. L., GOMIS, R. R. & MASSAGUE, J. 2008. TGFbeta primes breast tumors for lung metastasis seeding through angiopoietin-like 4. *Cell,* 133**,** 66-77.

PAGET, S. 1989. The distribution of secondary growths in cancer of the breast. 1889. *Cancer Metastasis Rev,* 8**,** 98-101.

PALMA, D. A., SALAMA, J. K., LO, S. S., SENAN, S., TREASURE, T., GOVINDAN, R. & WEICHSELBAUM, R. 2014. The oligometastatic state - separating truth from wishful thinking. *Nat Rev Clin Oncol,* 11**,** 549-57.

PALMIERI, D., BRONDER, J. L., HERRING, J. M., YONEDA, T., WEIL, R. J., STARK, A. M., KUREK, R., VEGA-VALLE, E., FEIGENBAUM, L., HALVERSON, D., VORTMEYER, A. O., STEINBERG, S. M., ALDAPE, K. & STEEG, P. S. 2007. Her-2 overexpression increases the metastatic outgrowth of breast cancer cells in the brain. *Cancer Research,* 67**,** 4190-4198.

PALMIERI, D., FITZGERALD, D., SHREEVE, S. M., HUA, E., BRONDER, J. L., WEIL, R. J., DAVIS, S., STARK, A. M., MERINO, M. J., KUREK, R., MEHDORN, H. M., DAVIS, G., STEINBERG, S. M., MELTZER, P. S., ALDAPE, K. & STEEG, P. S. 2009. Analyses of Resected Human Brain Metastases of Breast Cancer Reveal the Association between Up-Regulation of Hexokinase 2 and Poor Prognosis. *Molecular Cancer Research,* 7**,** 1438-1445.

PAOLI, P., GIANNONI, E. & CHIARUGI, P. 2013. Anoikis molecular pathways and its role in cancer progression. *Biochimica Et Biophysica Acta-Molecular Cell Research,* 1833**,** 3481-3498.

PARDOLL, D. M. 2012. The blockade of immune checkpoints in cancer immunotherapy. *Nature Reviews Cancer,* 12**,** 252-264.

PARELHO, V., HADJUR, S., SPIVAKOV, M., LELEU, M., SAUER, S., GREGSON, H. C., JARMUZ, A., CANZONETTA, C., WEBSTER, Z., NESTEROVA, T., COBB, B. S., YOKOMORI, K., DILLON, N., ARAGON, L., FISHER, A. G. & MERKENSCHLAGER, M. 2008. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell,* 132**,** 422-33.

PARKIN, D. M., BRAY, F., FERLAY, J. & PISANI, P. 2005. Global cancer statistics, 2002. *Ca-a Cancer Journal for Clinicians,* 55**,** 74-108.

PATCHELL, R. A. 2003. The management of brain metastases. *Cancer Treatment Reviews,* 29**,** 533-540.

PATCHELL, R. A., TIBBS, P. A., REGINE, W. F., DEMPSEY, R. J., MOHIUDDIN, M., KRYSCIO, R. J., MARKESBERY, W. R., FOON, K. A. & YOUNG, B. 1998. Postoperative radiotherapy in the treatment of single metastases to the brain - A randomized trial. *Jama-Journal of the American Medical Association,* 280**,** 1485-1489.

PAVLIDIS, N. & PENTHEROUDAKIS, G. 2012. Cancer of unknown primary site. *Lancet,* 379**,** 1428-35.

PEKMEZCI, M. & PERRY, A. 2013. Neuropathology of brain metastases. *Surg Neurol Int,* 4**,** S245-55.

PENCHEVA, N. & TAVAZOIE, S. F. 2013. Control of metastatic progression by microRNA regulatory networks. *Nature Cell Biology,* 15**,** 546-554.

PETITJEAN, A., MATHE, E., KATO, S., ISHIOKA, C., TAVTIGIAN, S. V., HAINAUT, P. & OLIVIER, M. 2007. Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum Mutat,* 28**,** 622-9.

PETRIE, R. J., DOYLE, A. D. & YAMADA, K. M. 2009. Random versus directionally persistent cell migration. *Nat Rev Mol Cell Biol,* 10**,** 538-49.

PFEIFER, G. P., DENISSENKO, M. F., OLIVIER, M., TRETYAKOVA, N., HECHT, S. S. & HAINAUT, P. 2002. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene,* 21**,** 7435-7451.

POLYAK, K. 2011. Heterogeneity in breast cancer. *J Clin Invest,* 121**,** 3786-8.

POWELL, G. J., SOUTHBY, J., DANKS, J. A., STILLWELL, R. G., HAYMAN, J. A., HENDERSON, M. A., BENNETT, R. C. & MARTIN, T. J. 1991. Localization of Parathyroid Hormone-Related Protein in Breast-Cancer Metastases - Increased Incidence in Bone Compared with Other Sites. *Cancer Research,* 51**,** 3059-3061.

PRESS, M. F., SAUTER, G., BERNSTEIN, L., VILLALOBOS, I. E., MIRLACHER, M., ZHOU, J. Y., WARDEH, R., LI, Y. T., GUZMAN, R., MA, Y. L., SULLIVAN-HALLEY, J., SANTIAGO, A., PARK, J. M., RIVA, A. & SLAMON, D. J. 2005. Diagnostic evaluation of HER-2 as a molecular target: An assessment of accuracy and reproducibility of laboratory testing in large, prospective, randomized clinical trials. *Clinical Cancer Research,* 11**,** 6598-6607.

PSAILA, B. & LYDEN, D. 2009. The metastatic niche: adapting the foreign soil. *Nat Rev Cancer,* 9**,** 285-93.

QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26**,** 841-842.

RAY, P. S., WANG, J. H., QU, Y., SIM, M. S., SHAMONKI, J., BAGARIA, S. P., YE, X., LIU, B. Y., ELASHOFF, D., HOON, D. S., WALTER, M. A., MARTENS, J. W., RICHARDSON, A. L., GIULIANO, A. E. & CUI, X. J. 2010. FOXC1 Is a Potential Prognostic Biomarker with Functional Significance in Basal-like Breast Cancer. *Cancer Research,* 70**,** 3870-3876.

REYMOND, N., IM, J. H., GARG, R., VEGA, F. M., BORDA D'AGUA, B., RIOU, P., COX, S., VALDERRAMA, F., MUSCHEL, R. J. & RIDLEY, A. J. 2012. Cdc42 promotes transendothelial migration of cancer cells through beta1 integrin. *J Cell Biol,* 199**,** 653-68.

RIDLEY, A. J. 2011. Life at the Leading Edge. *Cell,* 145**,** 1012-1022.

RIMMER, A., PHAN, H., MATHIESON, I., IQBAL, Z., TWIGG, S. R. F., WILKIE, A. O. M., MCVEAN, G., LUNTER, G. & CONSORTIUM, W. 2014. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nature Genetics,* 46**,** 912-918.

ROBERTS, S. A., LAWRENCE, M. S., KLIMCZAK, L. J., GRIMM, S. A., FARGO, D., STOJANOV, P., KIEZUN, A., KRYUKOV, G. V., CARTER, S. L., SAKSENA, G., HARRIS, S., SHAH, R. R., RESNICK, M. A., GETZ, G. & GORDENIN, D. A. 2013. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nature Genetics,* 45**,** 970-+.

ROBERTS, S. A., STERLING, J., THOMPSON, C., HARRIS, S., MAV, D., SHAH, R., KLIMCZAK, L. J., KRYUKOV, G. V., MALC, E., MIECZKOWSKI, P. A., RESNICK, M. A. & GORDENIN, D. A. 2012. Clustered Mutations in Yeast and in Human Cancers Can Arise from Damaged Long Single-Strand DNA Regions. *Molecular Cell,* 46**,** 424-435.

RODRIGUEZ-BOULAN, E. & MACARA, I. G. 2014. Organization and execution of the epithelial polarity programme. *Nature Reviews Molecular Cell Biology,* 15**,** 225-242.

ROTH, A., KHATTRA, J., YAP, D., WAN, A., LAKS, E., BIELE, J., HA, G., APARICIO, S., BOUCHARD-COTE, A. & SHAH, S. P. 2014. PyClone: statistical inference of clonal population structure in cancer. *Nature Methods,* 11**,** 396-+.

ROTHBERG, J. M., HINZ, W., REARICK, T. M., SCHULTZ, J., MILESKI, W., DAVEY, M., LEAMON, J. H., JOHNSON, K., MILGREW, M. J., EDWARDS, M., HOON, J., SIMONS, J. F., MARRAN, D., MYERS, J. W., DAVIDSON, J. F., BRANTING, A., NOBILE, J. R., PUC, B. P., LIGHT, D., CLARK, T. A., HUBER, M., BRANCIFORTE, J. T., STONER, I. B., CAWLEY, S. E., LYONS, M., FU, Y. T., HOMER, N., SEDOVA, M., MIAO, X., REED, B., SABINA, J., FEIERSTEIN, E., SCHORN, M., ALANJARY, M., DIMALANTA, E., DRESSMAN, D., KASINSKAS, R., SOKOLSKY, T., FIDANZA, J. A., NAMSARAEV, E., MCKERNAN, K. J., WILLIAMS, A., ROTH, G. T. & BUSTILLO, J. 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature,* 475**,** 348-352.

RUBIO, E. D., REISS, D. J., WELCSH, P. L., DISTECHE, C. M., FILIPPOVA, G. N., BALIGA, N. S., AEBERSOLD, R., RANISH, J. A. & KRUMM, A. 2008. CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A,* 105**,** 8309-14.

RUDENSTAM, C. M., ZAHRIEH, D., FORBES, J. F., CRIVELLARI, D., HOLMBERG, S. B., REY, P., DENT, D., CAMPBELL, I., BERNHARD, J., PRICE, K. N., CASTIGLIONE-GERTSCH, M., GOLDHIRSCH, A., GELBER, R. D., COATES, A. S., COLLINS, J., THURLIMANN, B., SENN, H. J., VERONESI, A., CORTES-FUNES, H., HILTBRUNNER, A., JENATSCH, C., EGLI, G., RABAGLIO, M., MAIBACH, R., ZAHRIEH, D., GELBER, S., BONETTI, M., KESHAVIAH, A., LI, S., O'NEILL, A., REGAN, M., HURNY, C., WESCHLER, Y., VIALE, G., GUSTERSON, B., SPATARO, V., BLACHER, L., HINKLE, R., LIPPERT, S., ISLEY, M., SCOTT, A., MONFARDINI, S., GALLIGIONI, E., MAGRI, M. D., BUONADONNA, A., MASSARUT, S., ROSSI, C., CANDIANI, E., CARBONE, A., VOLPE, R., RONCADIN, M., ARCICASA, M., CORAN, F., SIMONCINI, E., MARINI, G., MARPICATI, P., BRAGA, M., GRIGOLATO, P., LUCINI, L., MURRAY, E., WERNER, I. D., DENT, D. M., GUDGEON, A., STEYNOR, P., TOOP, J., WALLGREN, A., OTTOSSON-LONN, S., HULTBORN, R., COLLDAHL-JADESTROM, G., CAHLIN, E., MATTSSON, J., HOLMBERG, S. B., RUUSVIK, O., NIKLASSON, L. G., DAHLIN, S., KARLSSON, G., LINDBERG, B., SUNDBACK, A., BERGEGARCH, S., GROOT, O., DAHLBACK, L. O., SALANDER, H., ANDERSSON, C., HEIDEMAN, M., NISSBORG, A., WALLIN, A., CLAES, G., RAMHULT, T., SVENSSON, J. H., LIEDBERG, P., NILSSON, A., HAVEL, G., OESTBERG, G., PERSSON, S., SUURKULA, M., MATUSIK, J., LINDTNER, J., ERZEN, D., CUTER, T., CERVEK, J., et al. 2006. Randomized trial comparing axillary clearance versus no axillary clearance in older patients with breast cancer: First results of International Breast Cancer Study Group Trial 10-93. *Journal of Clinical Oncology,* 24**,** 337-344.

SANTOS, G. D., SHEPHERD, F. A. & TSAO, M. S. 2011. EGFR Mutations and Lung Cancer. *Annual Review of Pathology: Mechanisms of Disease, Vol 6,* 6**,** 49-69.

SCHAEFER, P. W., BUDZIK, R. F. & GONZALEZ, R. G. 1996. Imaging of cerebral metastases. *Neurosurgery Clinics of North America,* 7**,** 393-&.

SCHARDT, J. A., MEYER, M., HARTMANN, C. H., SCHUBERT, F., SCHMIDT-KITTLER, O., FUHRMANN, C., POLZER, B., PETRONIO, M., EILS, R. & KLEIN, C. A. 2005. Genomic analysis of single cytokeratin-positive cells from bone marrow reveals early mutational events in breast cancer. *Cancer Cell,* 8**,** 227-239.

SCHMID, K., OEHL, N., WRBA, F., PIRKER, R., PIRKER, C. & FILIPITS, M. 2009. EGFR/KRAS/BRAF mutations in primary lung adenocarcinomas and corresponding locoregional lymph node metastases. *Clin Cancer Res,* 15**,** 4554-60.

SCHMIDT-KITTLER, O., RAGG, T., DASKALAKIS, A., GRANZOW, M., AHR, A., BLANKENSTEIN, T. J. F., KAUFMANN, M., DIEBOLD, J., ARNHOLDT, H., MULLER, P., BISCHOFF, J., HARICH, D., SCHLIMOK, G., RIETHMULLER, G., EILS, R. & KLEIN, C. A. 2003. From latent disseminated cells to overt metastasis: Genetic analysis of systemic breast cancer progression. *Proceedings of the National Academy of Sciences of the United States of America,* 100**,** 7737-7742.

SCHOUTEN, L. J., RUTTEN, J., HUVENEERS, H. A. & TWIJNSTRA, A. 2002. Incidence of brain metastases in a cohort of patients with carcinoma of the breast, colon, kidney, and lung and melanoma. *Cancer,* 94**,** 2698-705.

SEVENICH, L., BOWMAN, R. L., MASON, S. D., QUAIL, D. E., RAPAPORT, F., ELIE, B. T., BROGI, E., BRASTIANOS, P. K., HAHN, W. C., HOLSINGER, L. J., MASSAGUE, J., LESLIE, C. S. & JOYCE, J. A. 2014. Analysis of tumour-and stroma-supplied proteolytic networks reveals a brain-metastasis-promoting role for cathepsin S. *Nature Cell Biology,* 16**,** 876-888.

SHAH, S. P., MORIN, R. D., KHATTRA, J., PRENTICE, L., PUGH, T., BURLEIGH, A., DELANEY, A., GELMON, K., GULIANY, R., SENZ, J., STEIDL, C., HOLT, R. A., JONES, S., SUN, M., LEUNG, G., MOORE, R., SEVERSON, T., TAYLOR, G. A., TESCHENDORFF, A. E., TSE, K., TURASHVILI, G., VARHOL, R., WARREN, R. L., WATSON, P., ZHAO, Y. J., CALDAS, C., HUNTSMAN, D., HIRST, M., MARRA, M. A. & APARICIO, S. 2009. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature,* 461**,** 809-U67.

SHAH, S. P., ROTH, A., GOYA, R., OLOUMI, A., HA, G., ZHAO, Y. J., TURASHVILI, G., DING, J. R., TSE, K., HAFFARI, G., BASHASHATI, A., PRENTICE, L. M., KHATTRA, J., BURLEIGH, A., YAP, D., BERNARD, V., MCPHERSON, A., SHUMANSKY, K., CRISAN, A., GIULIANY, R., HERAVI-MOUSSAVI, A., ROSNER, J., LAI, D., BIROL, I., VARHOL, R., TAM, A., DHALLA, N., ZENG, T., MA, K., CHAN, S. K., GRIFFITH, M., MORADIAN, A., CHENG, S. W. G., MORIN, G. B., WATSON, P., GELMON, K., CHIA, S., CHIN, S. F., CURTIS, C., RUEDA, O. M., PHAROAH, P. D., DAMARAJU, S., MACKEY, J., HOON, K., HARKINS, T., TADIGOTLA, V., SIGAROUDINIA, M., GASCARD, P., TLSTY, T., COSTELLO, J. F., MEYER, I. M., EAVES, C. J., WASSERMAN, W. W., JONES, S., HUNTSMAN, D., HIRST, M., CALDAS, C., MARRA, M. A. & APARICIO, S. 2012. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature,* 486**,** 395-399.

SHEN, M. M. 2013. Chromoplexy: A New Category of Complex Rearrangements in the Cancer Genome. *Cancer Cell,* 23**,** 567-569.

SHENDURE, J. & LIEBERMAN AIDEN, E. 2012. The expanding scope of DNA sequencing. *Nat Biotechnol,* 30**,** 1084-94.

SHISEKI, M., KOHNO, T., NISHIKAWA, R., SAMESHIMA, Y., MIZOGUCHI, H. & YOKOTA, J. 1994. Frequent allelic losses on chromosomes 2q, 18q, and 22q in advanced non-small cell lung carcinoma. *Cancer Res,* 54**,** 5643-8.

SIEGEL, R., MA, J. M., ZOU, Z. H. & JEMAL, A. 2014. Cancer Statistics, 2014. *Ca-a Cancer Journal for Clinicians,* 64**,** 9-29.

SIEGEL, R., NAISHADHAM, D. & JEMAL, A. 2013. Cancer statistics, 2013. *Ca-a Cancer Journal for Clinicians,* 63**,** 11-30.

SIMMEN, M. W. 2008. Genome-scale relationships between cytosine methylation and dinucleotide abundances in animals. *Genomics,* 92**,** 33-40.

SINDELAR, W. F., TRALKA, T. S. & KETCHAM, A. S. 1975. Electron microscopic observations on formation of pulmonary metastases. *J Surg Res,* 18**,** 137-61.

SJOBLOM, T., JONES, S., WOOD, L. D., PARSONS, D. W., LIN, J., BARBER, T. D., MANDELKER, D., LEARY, R. J., PTAK, J., SILLIMAN, N., SZABO, S., BUCKHAULTS, P., FARRELL, C., MEEH, P., MARKOWITZ, S. D., WILLIS, J., DAWSON, D., WILLSON, J. K. V., GAZDAR, A. F., HARTIGAN, J., WU, L., LIU, C. S., PARMIGIANI, G., PARK, B. H., BACHMAN, K. E., PAPADOPOULOS, N., VOGELSTEIN, B., KINZLER, K. W. & VELCULESCU, V. E. 2006. The consensus coding sequences of human breast and colorectal cancers. *Science,* 314**,** 268-274.

SLAMON, D. J., LEYLAND-JONES, B., SHAK, S., FUCHS, H., PATON, V., BAJAMONDE, A., FLEMING, T., EIERMANN, W., WOLTER, J., PEGRAM, M., BASELGA, J. & NORTON, L. 2001. Use of

chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *New England Journal of Medicine,* 344**,** 783-792.

SMITH, G., BOUNDS, R., WOLF, H., PRATT, N., CAREY, F., STEELE, R. & WOLF, C. 2010. Activating K-Ras Mutations Outwith "Hotspot" Codons in Sporadic Colorectal Tumours-Implications for Personalised Cancer Medicine. *Journal of Pathology,* 222**,** S12-S12.

SMITH, H. C., BENNETT, R. P., KIZILYER, A., MCDOUGALL, W. M. & PROHASKA, K. M. 2012. Functions and regulation of the APOBEC family of proteins. *Seminars in Cell & Developmental Biology,* 23**,** 258-268.

SNEED, P. K., SUH, J. H., GOETSCH, S. J., SANGHAVI, S. N., CHAPPELL, R., BUATTI, J. M., REGINE, W. F., WELTMAN, E., KING, V. J., BRENEMAN, J. C., SPERDUTO, P. W. & MEHTA, M. P. 2002. A multi-institutional review of radiosurgery alone vs. radiosurgery with whole brain radiotherapy as the initial management of brain metastases. *International Journal of Radiation Oncology Biology Physics,* 53**,** 519-526.

SOFFIETTI, R., TREVISAN, E. & RUDA, R. 2012. Targeted therapy in brain metastasis. *Current Opinion in Oncology,* 24**,** 679-686.

SONOSHITA, M., AOKI, M., FUWA, H., AOKI, K., HOSOGI, H., SAKAI, Y., HASHIDA, H., TAKABAYASHI, A., SASAKI, M., ROBINE, S., ITOH, K., YOSHIOKA, K., KAKIZAKI, F., KITAMURA, T., OSHIMA, M. & TAKETO, M. M. 2011. Suppression of colon cancer metastasis by Aes through inhibition of Notch signaling. *Cancer Cell,* 19**,** 125-37.

SOSA, M. S., BRAGADO, P. & AGUIRRE-GHISO, J. A. 2014. Mechanisms of disseminated cancer cell dormancy: an awakening field. *Nat Rev Cancer,* 14**,** 611-22.

STEMMLER, H. J., KAHLERT, S., SIEKIERA, W., UNTCH, M., HEINRICH, B. & HEINEMANN, V. 2006. Characteristics of patients with brain metastases receiving trastuzumab for HER2 overexpressing metastatic breast cancer. *Breast,* 15**,** 219-225.

STEPHENS, P. J., GREENMAN, C. D., FU, B. Y., YANG, F. T., 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. 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.

STINGL, J. & CALDAS, C. 2007. Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis. *Nat Rev Cancer,* 7**,** 791-9.

STRATTON, M. R., CAMPBELL, P. J. & FUTREAL, P. A. 2009. The cancer genome. *Nature,* 458**,** 719-724.

SUN, M., BEHRENS, C., FENG, L., OZBURN, N., TANG, X., YIN, G., KOMAKI, R., VARELLA-GARCIA, M., HONG, W. K., ALDAPE, K. D. & WISTUBA, II 2009. HER family receptor abnormalities in lung cancer brain metastases and corresponding primary tumors. *Clin Cancer Res,* 15**,** 4829-37.

SUTHERLAND, S., ASHLEY, S., MILES, D., CHAN, S., WARDLEY, A., DAVIDSON, N., BHATTI, R., SHEHATA, M., NOURAS, H., CAMBURN, T. & JOHNSTON, S. R. D. 2010. Treatment of HER2-positive metastatic breast cancer with lapatinib and capecitabine in the lapatinib expanded access programme, including efficacy in brain metastases - the UK experience. *British Journal of Cancer,* 102**,** 995-1002.

TAKAHASHI, K., KOHNO, T., MATSUMOTO, S., NAKANISHI, Y., ARAI, Y., YAMAMOTO, S., FUJIWARA, T., TANAKA, N. & YOKOTA, J. 2007. Clonal and parallel evolution of primary lung cancers and their metastases revealed by molecular dissection of cancer cells. *Clin Cancer Res,* 13**,** 111-20.

TALMADGE, J. E. & FIDLER, I. J. 2010. AACR centennial series: the biology of cancer metastasis: historical perspective. *Cancer Res,* 70**,** 5649-69.

TAN, W. S., HO, K. S. & EU, K. W. 2009. Brain Metastases in Colorectal Cancers. *World Journal of Surgery,* 33**,** 817-821.

TANIGUCHI, C. M., TRAN, T. T., KONDO, T., LUO, J., UEKI, K., CANTLEY, L. C. & KAHN, C. R. 2006. Phosphoinositide 3-kinase regulatory subunit p85alpha suppresses insulin action via positive regulation of PTEN. *Proc Natl Acad Sci U S A,* 103**,** 12093-7.

TARIN, D. 2005. The fallacy of epithelial mesenchymal transition in neoplasia. *Cancer Research,* 65**,** 5996-6000.

TARIN, D., PRICE, J. E., KETTLEWELL, M. G., SOUTER, R. G., VASS, A. C. & CROSSLEY, B. 1984. Mechanisms of human tumor metastasis studied in patients with peritoneovenous shunts. *Cancer Res,* 44**,** 3584-92.

TAVAZOIE, S. F., ALARCON, C., OSKARSSON, T., PADUA, D., WANG, Q. Q., BOS, P. D., GERALD, W. L. & MASSAGUE, J. 2008. Endogenous human microRNAs that suppress breast cancer metastasis. *Nature,* 451**,** 147-U3.

THIERY, J. P., ACLOQUE, H., HUANG, R. Y. J. & NIETO, M. A. 2009. Epithelial-Mesenchymal Transitions in Development and Disease. *Cell,* 139**,** 871-890.

THOMPSON, E. W., NEWGREEN, D. F. & TARIN, D. 2005. Carcinoma invasion and metastasis: a role for epithelial-mesenchymal transition? *Cancer Res,* 65**,** 5991-5; discussion 5995.

THORVALDSDOTTIR, H., ROBINSON, J. T. & MESIROV, J. P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform,* 14**,** 178-92.

THUNNISSEN, E., BUBENDORF, L., DIETEL, M., ELMBERGER, G., KERR, K., LOPEZ-RIOS, F., MOCH, H., OLSZEWSKI, W., PAUWELS, P., PENAULT-LLORCA, F. & ROSSI, G. 2012. EML4-ALK testing in non-small cell carcinomas of the lung: a review with recommendations. *Virchows Arch,* 461**,** 245-57.

TRIOZZI, P. L., ENG, C. & SINGH, A. D. 2008. Targeted therapy for uveal melanoma. *Cancer Treatment Reviews,* 34**,** 247-258.

TSUKADA, Y., FOUAD, A., PICKREN, J. W. & LANE, W. W. 1983. Central nervous system metastasis from breast carcinoma. Autopsy study. *Cancer,* 52**,** 2349-54.

TURAJLIC, S., FURNEY, S. J., LAMBROS, M. B., MITSOPOULOS, C., KOZAREWA, I., GEYER, F. C., MACKAY, A., HAKAS, J., ZVELEBIL, M., LORD, C. J., ASHWORTH, A., THOMAS, M., STAMP, G., LARKIN, J., REIS, J. S. & MARAIS, R. 2012a. Whole genome sequencing of matched primary and metastatic acral melanomas. *Genome Research,* 22**,** 196-207.

TURAJLIC, S., FURNEY, S. J., LAMBROS, M. B., MITSOPOULOS, C., KOZAREWA, I., GEYER, F. C., MACKAY, A., HAKAS, J., ZVELEBIL, M., LORD, C. J., ASHWORTH, A., THOMAS, M., STAMP, G., LARKIN, J., REIS-FILHO, J. S. & MARAIS, R. 2012b. Whole genome sequencing of matched primary and metastatic acral melanomas. *Genome Res,* 22**,** 196-207.

ULAHANNAN, D., KOVAC, M. B., MULHOLLAND, P. J., CAZIER, J. B. & TOMLINSON, I. 2013. Technical and implementation issues in using next-generation sequencing of cancers in clinical practice. *British Journal of Cancer,* 109**,** 827-835.

VALASTYAN, S. & WEINBERG, R. A. 2011. Tumor Metastasis: Molecular Insights and Evolving Paradigms. *Cell,* 147**,** 275-292.

VALIENTE, M., OBENAUF, A. C., JIN, X., CHEN, Q., ZHANG, X. H. F., LEE, D. J., CHAFT, J. E., KRIS, M. G., HUSE, J. T., BROGI, E. & MASSAGUE, J. 2014. Serpins Promote Cancer Cell Survival and Vascular Co-Option in Brain Metastasis. *Cell,* 156**,** 1002-1016.

VAN DIJK, E. L., AUGER, H., JASZCZYSZYN, Y. & THERMES, C. 2014. Ten years of next-generation sequencing technology. *Trends in Genetics,* 30**,** 418-426.

VAN ZANDWIJK, N., MATHY, A., BOERRIGTER, L., RUIJTER, H., TIELEN, I., DE JONG, D., BAAS, P., BURGERS, S. & NEDERLOF, P. 2007. EGFR and KRAS mutations as criteria for treatment with tyrosine kinase inhibitors: retro- and prospective observations in non-small-cell lung cancer. *Annals of Oncology,* 18**,** 99-103.

VERONESI, U., MARUBINI, E., MARIANI, L., VALAGUSSA, P. & ZUCALI, R. 1999. The dissection of internal mammary nodes does not improve the survival of breast cancer patients. 30-year results of a randomised trial. *European Journal of Cancer,* 35**,** 1320-1325.

WAN, L., PANTEL, K. & KANG, Y. 2013. Tumor metastasis: moving new biological insights into the clinic. *Nat Med,* 19**,** 1450-64.
WANG, K., LI, M. Y. & HAKONARSON, H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research,* 38.
WATERS, T. R. & SWANN, P. F. 2000. Thymine-DNA glycosylase and G to A transition mutations at CpG sites. *Mutat Res,* 462**,** 137-47.
WECKERMANN, D., POLZER, B., RAGG, T., BLANA, A., SCHLIMOK, G., ARNHOLDT, H., BERTZ, S., HARZMANN, R. & KLEIN, C. A. 2009. Perioperative Activation of Disseminated Tumor Cells in Bone Marrow of Patients With Prostate Cancer. *Journal of Clinical Oncology,* 27**,** 1549-1556.
WEICHSELBAUM, R. R. & HELLMAN, S. 2011. Oligometastases revisited. *Nat Rev Clin Oncol,* 8**,** 378-82.
WEIL, R. J., PALMIERI, D. C., BRONDER, J. L., STARK, A. M. & STEEG, P. S. 2005. Breast cancer metastasis to the central nervous system. *American Journal of Pathology,* 167**,** 913-920.
WEINBERG, R. A. 2008. Mechanisms of malignant progression. *Carcinogenesis,* 29**,** 1092-5.
WEIS, S. M. & CHERESH, D. A. 2011. Tumor angiogenesis: molecular pathways and therapeutic targets. *Nat Med,* 17**,** 1359-70.
WELSH, J. W., KOMAKI, R., AMINI, A., MUNSELL, M. F., UNGER, W., ALLEN, P. K., CHANG, J. Y., WEFEL, J. S., MCGOVERN, S. L., GARLAND, L. L., CHEN, S. S., HOLT, J., LIAO, Z. X., BROWN, P., SULMAN, E., HEYMACH, J. V., KIM, E. S. & STEA, B. 2013. Phase II Trial of Erlotinib Plus Concurrent Whole-Brain Radiation Therapy for Patients With Brain Metastases From Non-Small-Cell Lung Cancer. *Journal of Clinical Oncology,* 31**,** 895-902.
WHEELOCK, M. J., SHINTANI, Y., MAEDA, M., FUKUMOTO, Y. & JOHNSON, K. R. 2008. Cadherin switching. *Journal of Cell Science,* 121**,** 727-735.
WILKERSON, M. D. & HAYES, D. N. 2010. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics,* 26**,** 1572-3.
WONG, S. Q., LI, J., TAN, A. Y. C., VEDURURU, R., PANG, J. M. B., DO, H. D., ELLUL, J., DOIG, K., BELL, A., MACARTHUR, G. A., FOX, S. B., THOMAS, D. M., FELLOWES, A., PARISOT, J. P., DOBROVIC, A. & COHORT, C. 2014. Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. *Bmc Medical Genomics,* 7.
WU, C., LI, Y. L., WANG, Z. M., LI, Z., ZHANG, T. X. & WEI, Z. 2007. Gefitinib as palliative therapy for lung adenocarcinoma metastatic to the brain. *Lung Cancer,* 57**,** 359-364.
WU, J. N. & ROBERTS, C. W. 2013. ARID1A mutations in cancer: another epigenetic tumor suppressor? *Cancer Discov,* 3**,** 35-43.
XIE, T., CHO, Y. B., WANG, K., HUANG, D., HONG, H. K., CHOI, Y. L., KO, Y. H., NAM, D. H., JIN, J., YANG, H., FERNANDEZ, J., DENG, S., REJTO, P. A., LEE, W. Y. & MAO, M. 2014. Patterns of somatic alterations between matched primary and metastatic colorectal tumors characterized by whole-genome sequencing. *Genomics*.
XIE, T. X., HUANG, F. J., ALDAPE, K. D., KANG, S. H., LIU, A. G., GERSHENWALD, J. E., XIE, K. P., SAWAYA, R. & HUANG, S. Y. 2006. Activation of Stat3 in human melanoma promotes brain metastasis. *Cancer Research,* 66**,** 3188-3196.
YANG, Y. M., BAUER, C., STRASSER, G., WOLLMAN, R., JULIEN, J. P. & FUCHS, E. 1999. Integrators of the cytoskeleton that stabilize microtubules. *Cell,* 98**,** 229-238.
YANG, Y. M., DOWLING, J., YU, Q. C., KOUKLIS, P., CLEVELAND, D. W. & FUCHS, E. 1996. An essential cytoskeletal linker protein connecting actin microfilaments to intermediate filaments. *Cell,* 86**,** 655-665.
YANO, S., SHINOHARA, H., HERBST, R. S., KUNIYASU, H., BUCANA, C. D., ELLIS, L. M., DAVIS, D. W., MCCONKEY, D. J. & FIDLER, I. J. 2000. Expression of vascular endothelial growth factor is necessary but not sufficient for production and growth of brain metastasis. *Cancer Research,* 60**,** 4959-4967.

YAO, I., IIDA, J., NISHIMURA, W. & HATA, Y. 2002. Synaptic and nuclear localization of brain-enriched guanylate kinase-associated protein. *Journal of Neuroscience,* 22**,** 5354-5364.

YATES, L. R. & CAMPBELL, P. J. 2012. Evolution of the cancer genome. *Nature Reviews Genetics,* 13**,** 795-806.

YAU, C. 2013. OncoSNP-SEQ: a statistical approach for the identification of somatic copy number alterations from next-generation sequencing of cancer genomes. *Bioinformatics,* 29**,** 2482-2484.

YAU, C., MOURADOV, D., JORISSEN, R. N., COLELLA, S., MIRZA, G., STEERS, G., HARRIS, A., RAGOUSSIS, J., SIEBER, O. & HOLMES, C. C. 2010. A statistical approach for detecting genomic aberrations in heterogeneous tumor samples from single nucleotide polymorphism genotyping data. *Genome Biology,* 11.

YILMAZ, M. & CHRISTOFORI, G. 2009. EMT, the cytoskeleton, and cancer cell invasion. *Cancer and Metastasis Reviews,* 28**,** 15-33.

ZHANG, C. Z., LEIBOWITZ, M. L. & PELLMAN, D. 2013. Chromothripsis and beyond: rapid genome evolution from complex chromosomal rearrangements. *Genes Dev,* 27**,** 2513-30.

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**,** 179-84.

ZHANG, J. J., FUJIMOTO, J., ZHANG, J. H., WEDGE, D. C., SONG, X. Z., ZHANG, J. X., SETH, S., CHOW, C. W., CAO, Y., GUMBS, C., GOLD, K. A., KALHOR, N., LITTLE, L., MAHADESHWAR, H., MORAN, C., PROTOPOPOV, A., SUN, H. D., TANG, J. B., WU, X. F., YE, Y. Q., WILLIAM, W. N., LEE, J. J., HEYMACH, J. V., HONG, W. K., SWISHER, S., WISTUBA, I. I. & FUTREAL, A. 2014. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science,* 346**,** 256-259.