IDENTIFICATION OF DNA METHYLATION CHANGES IN SPORADIC BREAST AND OTHER CANCERS

Victoria Kate Hill

Thesis submitted to the Faculty of Medicine and Dentistry of the University of Birmingham for the Degree of Doctor of Philosophy

Medical and Molecular Genetics

School of Clinical and Experimental Medicine

The University of Birmingham

July 2011

ACKNOWLEDGEMENTS

Firstly, I would like to thank Prof. Farida Latif for her supervision during this PhD and her efforts to ensure all of this work has been published. I would also like to say a huge thank you to Breast Cancer Campaign for funding this research along with other funding bodies of the Latif Lab.

Day to day I have had a lot of help from all Medical and Molecular Genetics lab members but I would particularly like to thank Dean Gentle and Drs Chris Ricketts, Luke Hesson and Nick Underhill-Day. I also have to thank Jan Hatfield for her daily chats and endless supply of tips!

On a personal note, I have to thank my family, but particularly my parents for their constant support.

Finally, I have to thank Chris for his constant help and support. Only now am I beginning to realise how patient he has been without ever a hint of complaint – I still have a lot to learn! Thank you - and sorry.

ABSTRACT

Aberrant DNA methylation is a well known characteristic of cancer genomes. It can be used as a way of identifying important genes in tumourigenesis and can have diagnostic/prognostic value.

Using a genome-wide methylated DNA affinity enrichment approach this work identified five genes (*DBC1*, *CIDE-A*, *EMILIN2*, *FBLN2* and *SALL1*) that are hypermethylated in sporadic breast cancer. Methylation of one of these genes, *EMILIN2*, was found to associate with worse disease free survival (DFS). A second genome wide approach assessing over 27,000 CpG loci was carried out on sporadic breast cancer patient samples and identified greater overall methylation in ER positive tumours and those that relapsed. Individual locus analysis identified six genes where methylation associated with worse DFS. Of these, promising candidates for further analysis were identified, including *RECK*, *ACADL* and *C1orf114*.

Candidate gene approaches also identified methylation of two newly characterised cancer-related genes, *RASSF10* and *KIBRA*. Analysis of a range of solid cancers identified hypermethylation in multiple tumour types for *RASSF10*, the most frequent being gliomas. Frequent hypermethylation of *KIBRA* was identified in childhood acute lymphocytic leukaemia (ALL).

This study has used genome wide methods and candidate gene approaches to identify several novel methylated genes in a range of tumour types.

TABLE OF CONTENTS

Chapter 1 – Introduction..... 1

1.1 Cancer..... 1

1.1.1 Cancer incidence..... 1

1.1.2 Cellular basis of cancer..... 5

1.1.3 Cancer genetics..... 7

1.1.3.1 Tumour suppressor genes..... 7

1.1.3.2 Oncogenes..... 9

1.1.3.3 Regions of chromosomal loss and gain..... 10

1.1.3.4 Genomic rearrangements..... 11

1.1.3.5 Genome wide analysis..... 13

1.2 Cancer epigenetics..... 15

1.2.1 Histone modifications in cancer.....15

1.2.2 DNA methylation..... 17

1.2.2.1 CpG islands..... 19

1.2.2.2 DNA methylation machinery..... 20

1.2.2.3 DNA methyltransferases and cancer..... 23

1.2.2.4 DNA methylation and gene expression..... 23

1.2.2.5 Recruitment of methyl DNA binding proteins..... 24

1.2.3 Aberrant DNA methylation changes in cancer..... 28

1.2.3.1 DNA hypomethylation..... 28

1.2.3.2 DNA hypermethylation..... 28

1.2.3.3 DNA hypermethylation events as biomarkers..... 30

1.2.3.4 CpG island methylator phenotype (CIMP)..... 32

1.2.3.5 Loss of imprinting (LOI) in cancer..... 33

1.2.3.5 Cancer epigenomics..... 34

1.2.3.6 Epigenetic therapies..... 34

1.3 Analysis of DNA methylation..... 36
1.3.1 DNA modification by sodium bisulphite treatment..... 37
1.3.2 CoBRA (combined bisulphite restriction analysis)..... 37
1.3.3 Methylation specific PCR (MSP)..... 38
1.3.4 Pyrosequencing..... 40
1.3.5 MethyLight™..... 41
1.3.6 EpiTYPER™..... 41
1.3.7 Larger Scale/Genome wide methylation analysis..... 43
1.3.7.1 De-methylation (5azaDC) specific mRNA expression array analysis..... 43
1.3.7.2 Methylated DNA enrichment assays..... 45
1.3.7.3 Direct hybridisation of sodium bisulphite modified DNA to microarrays..... 48
1.3.8 DNA methylation analysis methods and 5-hydroxymethylcytosine 50
1.4 Breast cancer..... 51
1.4.1 Breast cancer subtypes and classification..... 55
1.4.2 Breast cancer genetics..... 59
1.4.2.1 Familial/Inherited predisposition to breast cancer..... 59
1.4.2.2 Somatic genetic alterations associated with breast cancer..... 64
1.4.3 Breast cancer epigenetics and epigenomics..... 68
1.5 The RASSF family..... 74
1.5.1 Mechanisms of tumour suppressor effects..... 76
1.5.2 The Classical RASSFs and promoter hypermethylation..... 77
1.5.3 The N-terminal RASSFs..... 80
1.5.3.1 RASSF7..... 80
1.5.3.2 RASSF8..... 80
1.5.3.3 RASSF9..... 81
1.5.3.4 RASSF10..... 81

1.6 The SWH pathway..... 83
1.6.1 Core SWH components..... 83
1.6.2 Downstream SWH pathway members..... 84
1.6.3 Upstream SWH pathway members..... 84
1.6.4 The SWH pathway and cancer..... 85
1.7 Research aims..... 88

Chapter 2 – Materials and methods..... 90
2.1 Materials..... 90
2.2 DNA samples..... 91
2.2.1 Cell lines..... 91
2.2.1.1 Breast cell lines..... 91
2.2.1.2 Lung cell lines..... 92
2.2.1.3 Glioma cell lines..... 92
2.2.1.4 Kidney cell lines..... 92
2.2.1.5 Prostate cell lines..... 92
2.2.1.6 Leukaemia cell lines..... 93
2.2.1.7 Colorectal cell lines..... 93
2.2.2 Patient samples..... 93
2.2.2.1 Breast tissue DNA..... 93
. 2.2.2.2 Glioma samples..... 94
2.2.2.3 Childhood acute lymphoblastic leukaemia samples..... 94
2.2.2.4 NSCLC tumours..... 94
2.2.2.5 RCC tumours..... 95
2.2.2.6 Ovarian tumours..... 95
2.3 Nucleic acid extraction..... 95
2.3.1 DNA extraction..... 95
2.3.2 RNA extraction..... 96

2.3.3 Quantification of nucleic acid preparations..... 97

2.4 DNA modification for methylation analysis..... 97

2.4.1 Bisulphite modification of DNA..... 97

2.4.1.1 Bisulphite modification of DNA using the traditional method..... 98

2.4.1.2 Bisulphite modification of DNA using the Qiagen EpiTect Kit.....100

2.5 Combined Bisulphite Restriction Analysis (CoBRA) and Methylation Specific PCR (MSP)..... 102

2.5.1 CoBRA primer design..... 102

2.5.2 CoBRA PCR..... 103

2.5.3 Agarose gel electrophoresis..... 105

2.5.4 Restriction digest..... 105

2.5.5 Methylation Specific PCR (MSP)..... 106

2.6 Bisulphite sequencing analysis..... 106

2.6.1 PCR product purification..... 106

2.6.2 Ligation of PCR products into plasmid vector..... 107

2.6.3 Transformation of bacterial cells with PCR product ligated into the pGEM-T easy vector..... 107

2.6.4 Single colony PCR..... 109

2.6.5 PCR product clean-up for sequencing..... 110

2.6.6 Sequencing reaction..... 110

2.6.7 Ethanol precipitation following sequencing reaction..... 111

2.7 Gene expression analysis..... 111

2.7.1 DNase treatment of RNA..... 111

2.7.2 cDNA synthesis..... 112

2.7.3 Reverse transcript polymerase chain reaction (RT-PCR) primer Design..... 112

2.7.4 RT-PCR..... 113

2.7.5 Real-time quantitative expression PCR..... 113
2.8 Tissue culture..... 114
2.8.1 Preparation of growth media..... 114
2.8.2 Resurrection of cell lines..... 115
2.8.3 Maintenance and passaging of cell lines..... 115
2.8.4 Counting of cells using a haemocytometer..... 115
2.8.5 Treatment of cell lines with 5-aza-2'deoxycitidine..... 116
2.8.6 Producing cell pellets..... 116
2.8.7 Producing frozen aliquots..... 117
2.9 Array analysis..... 117
2.9.1 MIRA assay..... 117
2.9.2 Illumina Infinium HumanMethylation27 BeadChip array..... 118
2.10 Statistical methods..... 118
2.10.1 Fisher's exact test..... 118
2.10.2 Student's t-test..... 119
2.10.3 ANOVA.....119
2.10.4 Kaplan-Meier analysis..... 119
2.10.5 Multiple correction..... 119
2.11 Bioinformatics..... 120
2.11.1 Gathering gene structure information..... 120
2.11.2 Functional annotation of gene lists..... 121
2.11.3 Individual gene information..... 121
2.11.4 Gene mutation and methylation information..... 121

Chapter 3 – Use of the Methylated CpG Island Recovery Assay (MIRA) coupled with human CpG island arrays to identify methylated genes in breast cancer..... 122

3.1 Abstract..... 122

3.2 Introduction..... 124

3.3 Aims..... 126

3.4 Results..... 127

3.4.1 Array analysis..... 127

3.4.2 Analysis of selected genes in breast cancer cell lines..... 129

3.4.3 Analysis of selected genes in primary breast tumour samples..... 135

3.4.4 Expression analysis of selected genes..... 135

3.4.5 Analysis of selected genes in primary breast tumour/normal paired samples..... 139

3.4.6 Association of methylation and clinical features..... 141

3.4.7 Expression of *EMILIN2* in primary tumour samples..... 141

3.4.8 Analysis of *CIDE-A*, *DBC1*, *EMILIN2*, *FBLN2* and *SALL1* in other common solid tumour types..... 147

3.4.8.1 *EMILIN2*..... 147

3.4.8.2 *SALL1*..... 149

3.4.8.3 *FBLN2*..... 149

3.4.8.4 *CIDE-A*..... 150

3.4.8.5 *DBC1*..... 150

3.4.9 miRNA analysis..... 150

3.5 Discussion..... 154

3.5.1 *CIDE-A*..... 156

3.5.2 *DBC1*..... 158

3.5.3 *EMILIN2*..... 160

3.5.4 *FBLN2*..... 163

3.5.5 *SALL1*..... 165

3.5.6 Conclusions..... 166

Chapter 4 – Illumina infinium methylation analyses of sporadic breast tumours....170

4.1 Abstract..... 170

4.2 Introduction..... 171

4.3 Aims..... 172

4.4 Results..... 173

4.4.1 Use of Illumina Infinium HumanMethylation27 BeadChip to measure genome wide DNA methylation changes.....173

4.4.2 Identification of candidate hypermethylated loci.....176

4.4.3 Array validation.....179

4.4.3.1 Assessment of individual β-values.....179

4.4.3.2 Functional assessment of gene list.....181

4.4.4 Analysis of candidate genes.....185

4.4.4.1 Highly methylated genes..... 185

4.4.4.2 Genes with multiple probes..... 191

4.4.5 Analysis of clinical features..... .195

4.4.5.1 Hierarchical clustering.....195

4.4.5.2 Analysis of Variance (ANOVA) analysis.....195

4.4.5.3 Identification of individual genes with association to clinical features.....199

4.4.6 Analysis of MIRA genes..... 210

4.5 Discussion.....220

4.5.1 Oestrogen (ER) and progesterone (PR) positive tumours.....221

4.5.1.1 *TNFRSF10D*.....222

4.5.1.2 *COL1A2*.....223

4.5.1.3 *C1orf114*.....224

4.5.2 Relapse and DFS 225

4.5.2.1 *SFRP2*.....225

4.5.2.2 *RECK*.....226

4.5.2.3 *ACADL*.....228

4.5.3 Use of Infinium array for candidate gene identification.....229

4.5.3.1 Highly Methylated Genes.....229

4.5.3.2 Multiple probe genes.....231

4.5.4 Comparison with the MIRA study (chapter 3).....232

4.5.5 Final Conclusions.....233

Chapter 5 – DNA methylation analysis of *RASSF10* in common solid tumours..... 236

5.1 Abstract..... 236

5.2 Introduction..... 238

5.3 Aims..... 239

5.4 Results..... 240

5.4.1 Methylation analysis of solid tumour cell lines..... 240

5.4.2. Methylation analysis of primary solid tumour samples..... 240

5.4.3 Methylation analysis of primary glioma samples..... 244

5.4.4 Association of *RASSF10* methylation with survival..... 248

5.5 Discussion..... 253

5.5.1 *RASSF10* methylation..... 253

5.5.3 *RASSF10* methylation in gliomas..... 254

5.5.3 *RASSF10* as a tumour suppressor..... 258

5.5.4 *RASSF10* as a member of the N-terminal RASSFs..... 260

Chapter 6 – DNA methylation analysis of the newly identified SALVADOR/WARTS/HIPPO pathway member, *KIBRA*..... 262

6.1 Abstract..... 262

6.2 Introduction..... 263

6.3 Aims..... 264

6.4 Results..... 265

6.4.1 Analysis of *KIBRA* in cancer cell lines..... 265

6.4.2 Analysis of *KIBRA* in primary glioma tumour samples..... 265

6.4.3 Analysis of *KIBRA* in primary leukaemia samples..... 270

6.4.4 Methylation and expression analysis in diagnosis and remission paired primary B-ALL samples..... 270

6.5 Discussion..... 274

6.5.1 *KIBRA* methylation in childhood ALL..... 274

6.5.2 *KIBRA* methylation as a deregulated member of the SWH pathway in haematological cancers..... 277

6.5.3 *KIBRA* as an upstream member of the SWH pathway..... 278

6.5.4 Conclusions and further work..... 278

Chapter 7 – Discussion.....280

7.1 Genome wide analysis of sporadic breast cancer..... 281

7.1.1 Experimental Design for Breast Cancer Array Analysis Studies.....284

7.2 Candidate genes analysis of *RASSF10* and *KIBRA* in multiple cancer types..... 288

7.3 Final conclusions 290

Appendix A.....I

Appendix B.....II

Appendix B.I.....II

Appendix B.II.....V

Appendix B.III.....IV

Appendix C.....XI

Appendix C.I.....XI

Appendix C.II.....XIII

Appendix C.III.....XIIII

Appendix D.....XVIII

Appendix D.I.....XVIII

Appendix D.II.....XIX

Appendix D.III.....XX

Appendix E.....XXI

Appendix E.I.....XXI

Appendix E.II.....XXII

References

LIST OF FIGURES

CHAPTER 1

Figure 1.1_Worlwide cancer incidence and death rates.....2

Figure 1.2_The top ten most commonly diagnosed cancers in the UK.....3

Figure 1.3_Timeline of major developments in cancer over the past 100 years.....8

Figure 1.4_Distribution of CpG dinucleotides throughout the genome.....20

Figure 1.5_Gene transcriptional silencing due to DNA methylation.....27

Figure 1.6_CoBRA assay for analysis of single gene methylation status.....39

Figure 1.7_Schematic of the pyrosequencing technique.....42

Figure 1.8_Schematic of the EpiTYPER technique.....44

Figure 1.9_Schematic of the MIRA coupled with microarray technique.....47

Figure 1.10_Schematic of the Illumina Infinium Methylation Assay.....49

Figure 1.11_Worldwide breast cancer incidence and death rates.....52

Figure 1.12_Influence of cancer stage and time of diagnosis on survival and mortality rates in the UK.....53

Figure 1.13_Breast cancer European age-standardised incidence and mortality rates in the UK 1975-2008.....54

Figure 1.14_Normal and cancerous breast anatomy.....57

Figure 1.15_The top 20 most significant genes somatically mutated in breast cancer.....67

Figure 1.16_Methylation of the top 20 most analysed in breast cancer.....71

Figure 1.17_Major isoforms of RASSF family members.....75

Figure 1.18_SWH pathway in *D.melanogaster* and mammals.....86

CHAPTER 2

Figure 2.1_pGEM T-easy plasmid map.....108

CHAPTER 3

Figure 3.1_Flow diagram of MIRA list analysis.....128

Figure 3.2_CoBRA results for selected genes in cancer cell lines.....133
Figure 3.3_CoBRA results for selected genes in primary tumour samples.....137
Figure 3.4_Expression analysis of selected genes.....138
Figure 3.5_Tumour specificity of selected genes in tumour-normal paired samples.....140
Figure 3.6_Sequencing results for *EMILIN2*.....142
Figure 3.7_Sequencing results for *CIDE-A*.....143
Figure 3.8_Sequencing results for *SALL1*.....144
Figure 3.9_Kaplan-Meier analysis for *EMILIN2*.....145
Figure 3.10_*EMILIN2* expression in primary tumour samples.....146
Figure 3.11_CoBRA results for other solid tumour cell samples.....151

CHAPTER 4

Figure 4.1_Correlation of a replicated sample.....175
Figure 4.2_β value correlation between normals.....177
Figure 4.3_β value correlation between tumour normal paired samples.....178
Figure 4.4_Identification of hypermethylated loci.....180
Figure 4.5_β-value validation for *SFRP5*.....182
Figure 4.6_β-value validation for *SIPA1*.....183
Figure 4.7_β-value validation for *CLDN6*.....184
Figure 4.8_DAVID analysis of hypermethylated loci.....186
Figure 4.9_Top ten highly methylated probes.....188
Figure 4.10_Cell line CoBRA and expression analysis for highly methylated genes.....189
Figure 4.11_CoBRA tumour/normal pair results for *COX7A1* and *SST*.....190
Figure 4.12_Genes with multiple probes.....192
Figure 4.13_Cell CoBRA and expression analysis for highly methylated genes.....193
Figure 4.14_CoBRA tumour/normal pair results for *CDKL2* and *ZNF154*.....194
Figure 4.15_Hierarchical clustering splits tumour samples into three main groups.....196

Figure 4.16_ANOVA analysis of methylation and clinical features.....198

Figure 4.17_ANOVA analysis of disease free survival and clinical features.....200

Figure 4.18_Probes that show significant association between methylation and relapse.....202

Figure 4.19_Kaplan-Meier analysis.....204

Figure 4.20_Cell line CoBRA and expression analysis of *RECK*, *ACADL* and *SFRP2*.....205

Figure 4.21_Clone sequencing and CoBRA results for *RECK* tumour normal pairs.....206

Figure 4.22_Clone sequencing and CoBRA results for *ACADL* tumour normal pairs.....207

Figure 4.23_Quantitative real-time expression results for *RECK* and *ACADL*.....208

Figure 4.24_Probes showing a significant association between methylation and ER and PR expressing tumours 209

Figure 4.25_Cell line CoBRA and expression analysis for ER and PR associated genes.....211

Figure 4.26_CoBRA tumour/normal pair results for *TNFRSF10D*, *COL1A2* and *C1orf114*.....212

Figure 4.27_Clone sequencing and CoBRA results for *C1orf114* tumour normal pairs.....213

Figure 4.28_Quantitative real-time expression for *TNFRSF10D*.....214

CHAPTER 5

Figure 5.1_*RASSF10* CoBRA region.....241

Figure 5.2_*RASSF10* CoBRA results for solid tumour cell lines.....242

Figure 5.3_Methylation analysis for glioma cell lines.....243

Figure 5.4_CoBRA analysis of primary colorectal, lung, kidney and breast samples.....245

Figure 5.5_CoBRA and sequencing of primary glioma tumours and non-cancerous brain tissue I.....246

Figure 5.6_CoBRA and sequencing of primary glioma tumours and non-cancerous brain tissue II.....247

Figure 5.7_*RASSF10* methylation at grade II and IV gliomas.....249

Figure 5.8_*RASSF1A* and *MGMT* methylation resulst.....250

Figure 5.9_Kaplan-Meier analysis for *RASSF10* in sGBM tumours.....251

Figure 5.10_*RASSF10* colony forming ability.....259

CHAPTER 6

Figure 6.1_*KIBRA* CoBRA region.....266

Figure 6.2_CoBRA results for solid tumour cell lines.....267

Figure 6.3_CoBRA and expression for methylated cell lines.....268

Figure 6.4_CoBRA results for primary glioma samples.....269

Figure 6.5_CoBRA results for primary ALL samples.....271

Figure 6.6_Sequencing of ALL samples.....272

Figure 6.7_CoBRA and expression results for diagnosis and remission B-ALL samples.....273

LIST OF TABLES

CHAPTER 1

Table 1.1_Breast cancer susceptibility genes.....65

CHAPTER 2

Table 2.1_Bisulphite modification conditions for use with EpiTect kit.....101

Table 2.2_Standard and touchdown CoBRA PCR conditions.....104

CHAPTER 3

Table 3.1_Reason for selecting the 32 genes for analysis in cell lines.....130

Table 3.2_CoBRA results for 32 selected genes in breast cancer cell lines.....131

Table 3.3_CoBRA results for selected genes analysed in primary tumour samples.....136

Table 3.4_Results of other solid tumour types.....148

Table 3.5_Mutational data of *CIDE-A*, *DBC1*, *EMILIN2*, *FBLN2*, and *SALL1* from COSMIC.....159

CHAPTER 4

Table 4.1_ Clinical features of the 39 sporadic breast tumours used in the array.....174

Table 4.2_Individual gene methylation associated with clinical features.....201

Table 4.3_Infinium results for genes identified through MIRA.....215

Table 4.4_Genes identified through both MIRA and Infinium analysis.....219

CHAPTER 5

Table 5.1_Methyation frequency of RASSF10 in multiple cancers.....255

CHAPTER 1

INTRODUCTION

1.1 CANCER

Cancer is ‘a disease caused by an uncontrolled division of abnormal cells in a part of the body’ (Oxford English Dictionary).

1.1.1 Cancer incidence

Cancer is the leading cause of death worldwide and second leading cause of death in developing countries (Jemal *et al*¸ 2011). In 2008, over 12 million cases of cancer were diagnosed worldwide in males and females combined. In total, 2008 saw 4.2 and 3.3 million deaths occurring as a result of cancer in males and females respectively (Jemal *et al*, 2011). The most common forms of cancer diagnosed worldwide are lung cancers in males and breast cancers in females, these two cancers are also responsible for the greatest number of cancer attributable deaths worldwide (figure 1.1). Due to differences in healthcare, cancer mortality rates are higher in economically developing countries, occurring at a frequency of 1.7 times that in developed countries. For example, more than 85% of cervical cancer cases and deaths occur in developing countries as a result of a lack of screening techniques to detect pre-cancerous lesions which are now commonplace in many developed countries (Jemal *et al*, 2011). In the UK, approximately 309,500 people were diagnosed with cancer in 2008; based on this 2008 data, breast cancer is the most common cancer in females, accounting for 31% of all adult female cancers, whilst prostate cancer is the most common cancer in males, accounting for 24% of all adult male cancers. For both males and females, the next most common cancers are lung and colorectal (CRUK cancer statistics) (figure 1.2). Despite the

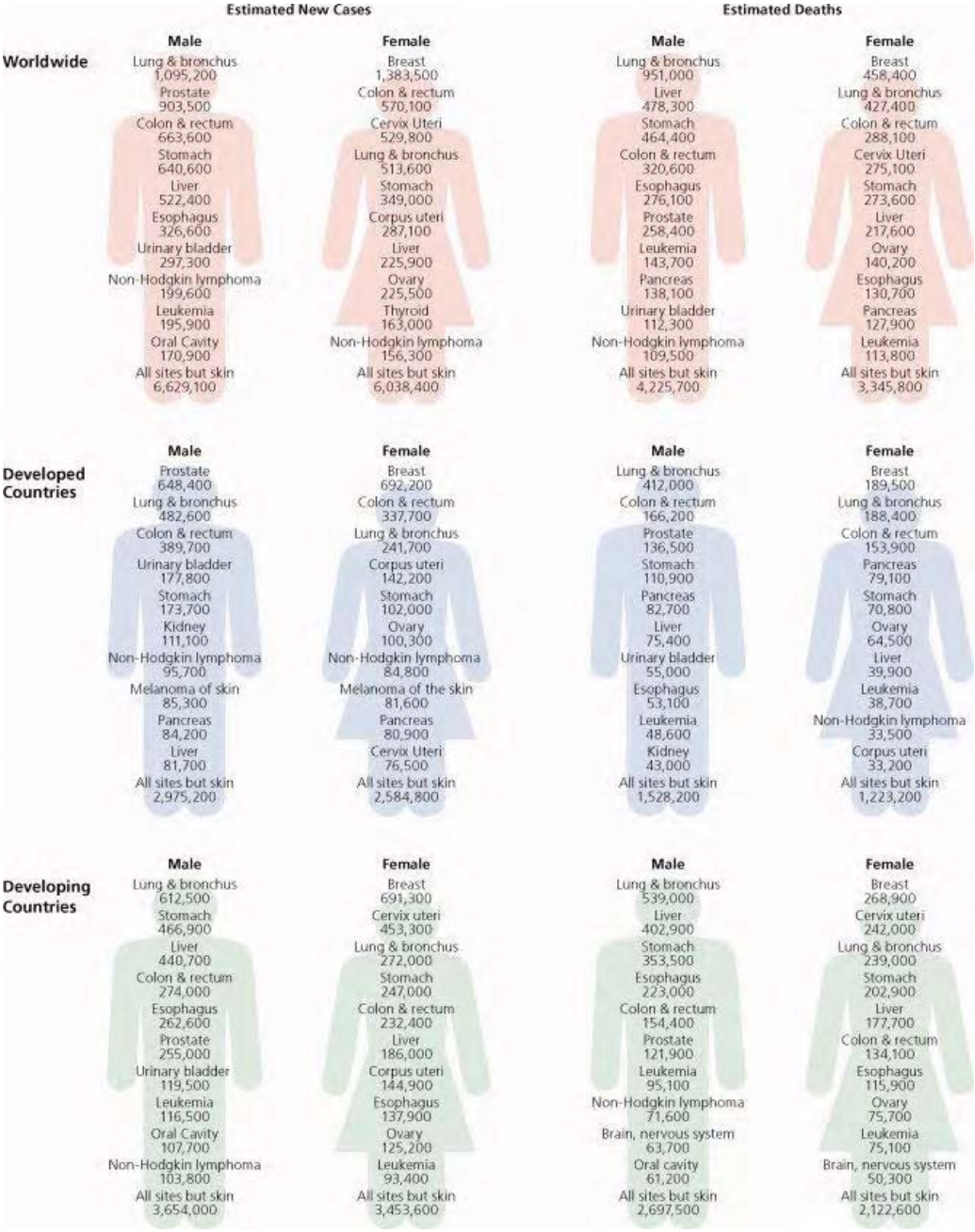
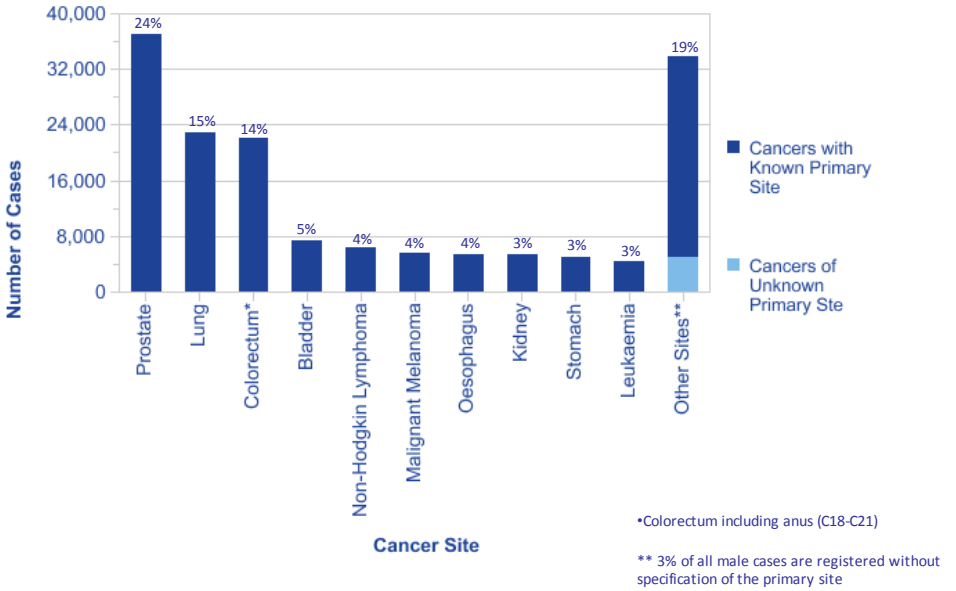


Figure 1.1_Worldwide cancer incidence and death rates

This figure has been taken from Jemal *et al*, 2011 and shows the estimated numbers of diagnoses and deaths attributable to different cancer types in 2008. Frequencies are broken down into males and females, and are shown for developed and developing countries both combined and separately.

(a)



(b)

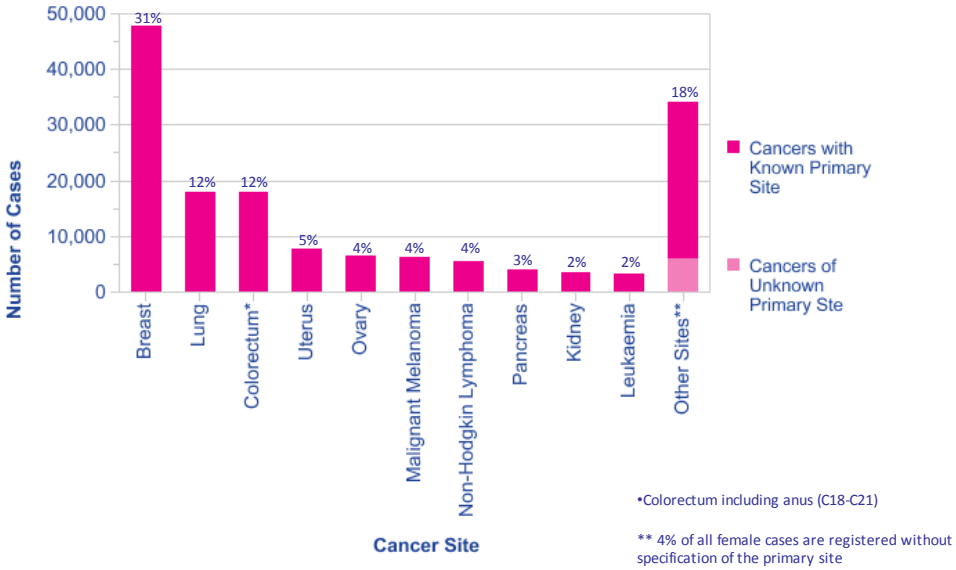


Figure 1.2_The top ten most commonly diagnosed cancers in the UK
The top ten most commonly diagnosed cancers in the UK in 2008 are shown for males (a) and females (b). These graphs have been taken from the Cancer Research UK (CRUK) website (www.cancerresearchuk.org).

overwhelming proportion of breast cancers occurring in females, the disease had the third highest mortality rate in 2005, accounting for 8% of all UK cancer deaths. However, the highest mortality rate in females was still from lung cancer, as in males (CRUK cancer statistics). The most common childhood cancers in the UK are leukaemias and brain and CNS cancers, accounting for approximately 30% and 20% respectively of all childhood cancers (CRUK cancer statistics).

In the developed world the incidence of cancer has been increasing, and is likely to continue to increase as a result of an ageing population and changes in lifestyle. For example, the incidence of breast cancer increased in many Western countries between the 1980s and 1990s as a result of changes in reproductive factors such as the increasing use of post-menopausal hormone therapies. Colorectal cancer incidences have also been increasing as a result of changes in diet, increased obesity and smoking, while lung cancer has seen an increase in Asian and African countries as a result of increased/unchanged smoking habits. Liver cancers are increasing as a result of increasing obesity rates and hepatitis C infection (Jemal *et al*, 2011). However, it is important to note that a proportion of these increased cases are attributable to the development and increased usage of effective screening techniques. Since the rapid increase in breast cancer rates in the 1980s and 1990s in Western countries, the incidence rates decreased in the 2000s due to the heavily sequestered use of post-menopausal hormone therapy after its causative effects were realised (Colditz, 1998). Despite the overall increasing incidence rates of many cancers, the death rates of cancers including lung, colorectal, breast and prostate cancers have been decreasing in the Western world as a result of earlier detection through screening methods and better treatment of the cancer through increased understanding of the disease. This is exemplified by testicular cancer for which the survival rate has increased massively, even for advanced disease (Feldman *et* al, 2008).

Additionally, due to the unrelenting efforts of research into all aspects of cancer, disease specific tailored therapy is now available for many cancer types, such as the much publicised use of trastuzumab (Herceptin) (an inhibiting monoclonal antibody against HER2) in late stage HER2/neu (*ERBB2*) over-expressing breast tumours (Hudis *et al*, 2007) and the development of the tyrosine kinase inhibitor, gefitinib, against specific EGFR mutations in non-small cell lung cancer (Lynch *et al*¸ 2004; Paez *et al*, 2004). These developments have markedly increased the outcome for many patients, although unfortunately the effects of this have been observed primarily in developed countries. However, the impact of a cancer diagnosis on an individual and their family members remains huge, as does the burden of cancer on healthcare services across the World. Further identification of good therapeutic targets and prognostic, diagnostic and treatment outcome markers are essential for benefitting patient outcome.

1.1.2 Cellular basis of cancer

Whist there are over 100 types of cancers consisting of multiple subtypes able to form in a single organ or tissue type, the majority of cancers can be classified, on a very general level, based on their tissue site of origin. The most common form of cancer, accounting for over 80% of all cancers, arises in epithelial cells; these cancers are termed carcinomas. The majority of carcinomas may be split into squamous cell carcinomas or adenocarcinomas depending on whether the epithelial cells of origin are part of the protective epithelial layer (squamous cell carcinomas) or have secretory properties (adenocarcinomas) (Weinberg, 2007). Examples of squamous cell carcinomas include squamous cell carcinoma of the skin and squamous cell carcinoma of the cervix, whilst adenocarcinomas include adenocarcinomas of the breast and colon. Although some carcinomas can be defined as pure squamous cell carcinomas or adenocarcinomas, many carcinomas are located where both cell types co-exist.

After carcinomas, the majority of malignancies are either sarcomas, haematological or arise within the central or peripheral nervous system (Weinberg, 2007). Sarcomas arise from mesenchymal cells such as connective tissue, adipocytes or osteoblasts. Examples include bone cancers, such as osteosarcomas and chondrosarcomas, and muscle tissue based malignancies, such as rhabdomyosarcomas. Haematological malignancies can be split into leukaemias and lymphomas; leukaemias arising from various lineages of white blood cells and lymphomas arising from lymphoid lineages that then go on to form solid tumour masses. Central and peripheral nervous system tumours include gliomas, neuroblastomas and medulloblastomas, among others. There are a small number of other cancers, such as melanomas, that do not fit into any of the above categories (Weinberg, 2007).

Despite the site of origin of malignancies, cells must acquire various biological traits to form a malignant tumour. Reviewed in Hanahan *et al* (2000), these biological attributes include:

i. Self-sufficiency in growth i.e. not requiring external growth signals to proliferate.
ii. Insensitivity to antigrowth signals i.e. becoming able to block the inhibition of proliferation initiated by external growth signals.
iii. The ability to evade apoptosis – the majority of cancer cells have acquired some mechanism of resistance to apoptosis.
iv. Potentially limitless replication – overcoming the intrinsic switch to senescence after numerous cell divisions
v. Sustained angiogenesis – angiogenic properties are normally tightly regulated but essential for transition from an aberrant proliferative lesion to a larger malignant tumour entity
vi. Ability to invade surrounding tissue and metastasize – tumours only become malignant upon invasion of surrounding tissue.

To be able to acquire these biological properties that enable malignant tumour growth a multitude of genetic changes must occur within the cell.

1.1.3 Cancer Genetics

Many alterations within the genome, both germline and somatic, can influence the process of tumourigenesis. Mutations, genome rearrangements, amplifications, deletions, genome instability and epigenetic changes, such as DNA methylation and histone modifications, can all contribute to the tumourigenic process. The view of cancer genetics has changed dramatically over the past 100 years and provides an almost unending wealth of knowledge that could be discussed. Figure 1.3 provides a timeline of major discoveries and initiatives during this time. The work in this thesis concentrates on a specific type of somatic alteration that can occur during tumourigenesis and tumour progression, DNA methylation. This introduction will provide a brief overview of the other forms of somatic alterations before going into details concerning epigenetics and specifically DNA methylation.

1.1.3.1 Tumour Suppressor Genes

In a normal cellular context, tumour suppressor genes are responsible for keeping cell behaviour under control. When inactivated, associated normal cell functions can spiral out of control causing, or adding to, the tumourigenic process. In most cases, both copies of a tumour suppressor must be inactivated to lead to cancer formation. This theory, Knudson's 'two hit hypothesis' was first suggested in 1971, based on hereditary aspects of retinoblastoma (Knudson, 1971). Occurring almost exclusively as either heritable bilateral cancers or sporadic unilateral cases, it was proposed that unilateral retinoblastomas occurred

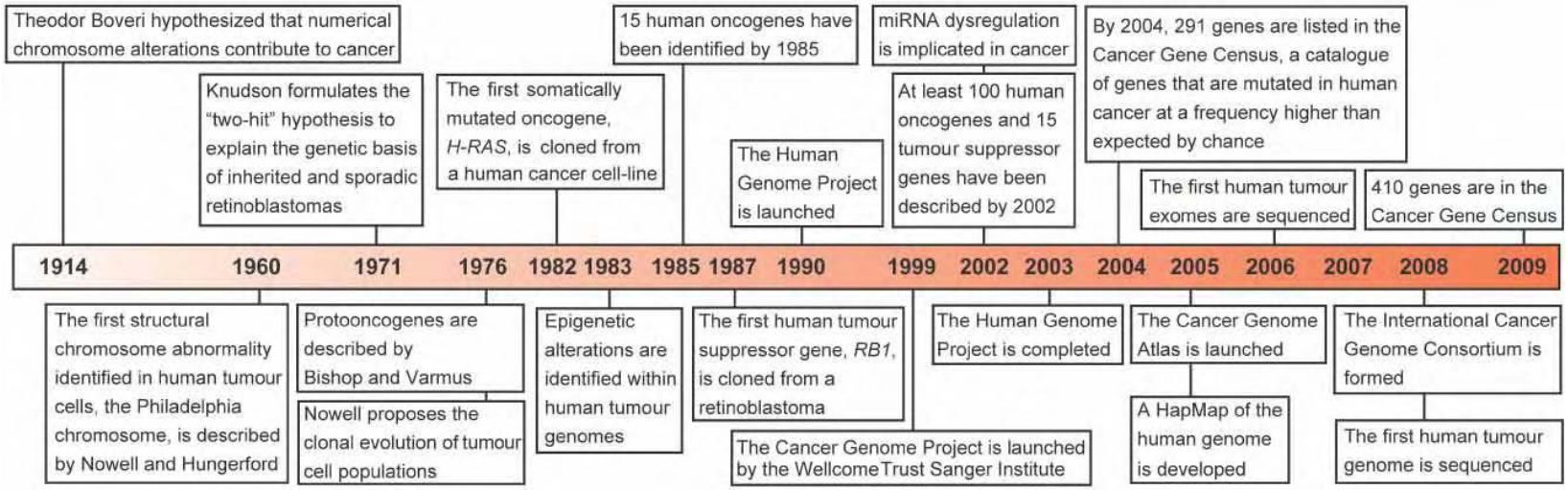


Figure 1.3_Timeline of major developments in cancer genetics over the past 100 years

The above diagram shows a timeline of major developments in cancer genetics including the first descriptions of cancer specific translocations, oncogenes, tumour suppressors, epigenetic changes and sequencing initiatives. This figure was taken from Bell, 2010.

as a result of acquired mutations in a then unknown causative gene during development. Bilateral retinoblastomas on the other hand, occurred when a germline copy of the then unknown causative gene was already mutated and therefore only one mutation was required to occur during development, making cancers bilateral and earlier events. The causative gene, *RB1* (*Retinoblastoma 1*), was subsequently isolated as the first tumour suppressor in 1987 (Lee *et al*, 1987; Fung *et al*, 1987). It remains a classical example of a tumour suppressor gene, involved heavily in cell cycle regulation. Inactivation of the *RB1* protein product, pRb, via phosphorylation by Cyclin D/Cdk-4 or -6 results in the activation of the E2F transcription factor which is required for upregulation of genes required for S phase initiation, such as cyclin E. Loss of pRb result in unregulated E2F activity and therefore unregulated cell cycle progression (Hanahan *et al*, 2000). This, however, is only one of the mechanisms by which *RB1* is able to function as a tumour suppressor gene – *RB1* has additionally been shown to be involved in chromosomal instability, angiogenesis and autophagy in response to hypoxia (Burkhart *et al*, 2008). In addition to retinoblastomas, *RB1* inactivation has been observed in many other cancers, including breast, prostate, lung and oesophageal cancers, among others (Burkhart *et al*, 2008) Other classical tumour suppressor genes include *TP53* (Bourdon, 2007; Vousden and Lane, 2007) and *CDKN2A* (Rocco and Sidransky, 2001) in multiple tumour types and *VHL* in kidney cancer (Latif *et al*, 1993).

1.1.3.2 Oncogenes

Oncogenes were first described in 1976, with the discovery that alteration of the normal activity of certain genes resulted in oncogenic transformation of cells. Altered, oncogenic versions of these genes are termed oncogenes while the unaltered version is called a proto-oncogene. *SRC* (*v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian)*), a homologue of the Rous sarcoma virus transforming gene, *v-src*, which causes sarcomas in

chickens, was the first proto-oncogene to be described (Stehelin *et al*, 1976). *SRC* functions as a tyrosine-kinase that becomes overactive in many cancers. Oncogenic mutation is common but unusual in the sense that the protein sequence must be altered to effect function and/or expression control but not result in either rendering the protein ineffective or resulting in complete loss of protein expression, as is the case with tumour suppressor genes. Mutations in oncogenes therefore have a tendency to be limited to specific mutation of single amino acids or specific genomic rearrangements, as discussed below. A good example of an oncogene is the *PIK3CA* (*phosphoinositide-3-kinase, catalytic, alpha polypeptide*) gene. It codes for the p110α protein, a class I PI 3-kinase catalytic substrate. *PIK3CA* mutations have been observed in numerous cancer types and, to date, *PIK3CA* represents the most frequently somatically mutated gene in breast cancer (Samuels *et al*, 2004; Campbell *et al*, 2004). This gene also demonstrates the specificity of oncogenic mutation, as for the 1435 instances of *PIK3CA* mutation in breast cancer listed by COSMIC (catalogue of somatic mutations in cancer – http://www.sanger.ac.uk/genetics/CGP/cosmic/), mutation of 1366 are represented by alterations to just three residues (p.E542 – 159 mutations, pE545 – 300 mutations, p.H1047 – 907 mutations).

1.1.3.3 Regions of chromosomal loss and gain

Regions of chromosomal loss and gain are also a common feature of cancer genomes, often termed copy number alterations. Both large regional aberrations and more focal aberrations are common and are capable of affecting well known tumour suppressor genes and oncogenes. For example, amplification of PIK3CA has been observed in numerous cancers including those with infrequent mutation of PIK3CA such as prostate cancer (Agell *et al*, 2011) and those with a higher mutation rate such as pituitary tumours, demonstrating mutation in approximately 9% of tumours and genomic amplification in 20-40% in one recent

study (Lin *et al*, 2009). PIK3CA amplification and amplification of associated PIK3 pathway members is a common alteration associated with bad prognosis in ovarian cancer, demonstrating the importance of over-activating an oncogenic pathway by numerous mechanisms (Huang *et al*, 2011). Up to 80% of sporadic chordomas have shown genomic loss of the well known tumour suppressor gene CDNK2A (Le *et al*, 2011) in addition to CDKN2A loss in other cancers such as paediatric astrocytomas (Schiffman *et al*, 2010), ovarian cancer (Gorringe *et al*, 2009) and gastric cancer (Belinsky *et al*, 2009). A recent genome wide study across just over 3000 cancers encompassing 26 histological subtypes identified a large number of somatic copy number alterations (SCNAs). On average, each sample demonstrated 24 gains and 18 losses (Beroukhim *et al*¸ 2010). In general, arm-level SCNAs were more common which will affect numerous genes rather than focal SCNAs which are likely to only affect only one or two (Beroukhim *et al*, 2010). The most common focal SCNAs affected *MYC* via amplification and *CDKN2A/B* via deletion.

1.1.3.4 Genomic Rearrangements

Genomic rearrangements are common facets of the cancer genome, particularly in haematological cancers. The first structural chromosomal abnormality was identified in 1960 by Nowell and Hungerford in chronic myeloid leukaemia (CML). The translocation was identified between chromosomes 9 and 22 [t(9;22)(q34;q11)] that results in an elongated form of chromosome 9 and a truncated version of chromosome 22 (the Philadelphia chromosome) (Nowell *et al*, 1960). A fusion protein between *BCR* (*breakpoint control region*) and *ABL* (*Abelson*) is produced on the Philadelphia chromosome that gives rise to a constitutively active tyrosine kinase capable of activating numerous pathways resulting in increased cell division, checkpoint inhibition and subsequent genomic instability (Laurent *et al*, 2001). This novel loss of functional control due to the production of a fusion protein is typical of this type

of rearrangement having potent oncogenic properties, although such rearrangements can also result in the disruption of tumour suppressor genes as an alternative to mutation, deletion or methylation. The Philadelphia chromosome is an extremely common event in CML, occurring in approximately 95% of cases (Nowell *et al*, 1960). Some adulthood and childhood acute lymphocytic leukaemia (ALL) patients also harbour this translocation at frequencies of 25-30% and 2-10% respectively (Chissoe *et al*, 1995). As an overactive tyrosine-kinase, *BCR*/*ABL* patients respond to tyrosine-kinase inhibitors such as Imatinib (Glivec) (Deininger *et al*, 2003). A common translocation in childhood ALL is the [t(12;21)(p13;q22)] translocation resulting in the *ETV6*/*RUNX1* fusion protein, which, like *BCR*/*ABL*, is a constitutively active tyrosine-kinase. This translocation is present in approximately 25% of childhood ALL cases (Shurtleff *et al*, 1995).

Genomic translocations producing potent oncogenic fusion proteins are much less common in solid cancer genomes. To date, the most well-known solid tumour translocations have been observed in prostate cancer, producing the *TMPRSS2-ERG*/*ETV1* fusion protein (Tomlins *et al*, 2005) in approximately 45% of cases (Mosquera *et* al, 2009) and in lung cancer, producing the *EML4*/*ALK* fusion protein, although this was only observed in approximately 6% of non-small cell lung cancers (NSCLC) (Soda *et al*, 2007). However, a recent study across 24 breast cancers (15 of which are were primary samples, the remaining being cell lines) using massive parallel paired end sequencing, identified a multitude of somatic rearrangements, suggesting genomic rearrangements in solid tumours may be more common than first thought (Stephens *et al*, 2009). In total, 2166 rearrangements were observed among all the samples, 50% of which occurred within protein coding genes. Of these, 29 were expected to produce in-frame fusion proteins. Another, more recent study, identified a median of 90 somatic rearrangements per prostate cancer genome when performing whole genome sequencing on 7 prostate cancers

and their corresponding normal counterparts (Berger *et al*, 2011). Again suggesting that although potent oncogenic fusion proteins have been primarily observed in haematological cancer genomes, somatic rearrangements in solid tumours are possibly a much more prevalent alteration than first thought.

1.1.3.5 Genome wide analysis

Recent technology advances have made it possible to perform large scale genome wide approaches including whole exome (all the coding genome sequence) and whole genome sequencing at a relatively affordable cost. This has given rise to cancer genome sequencing consortiums such as the International Cancer Genome Consortium (ICGC), the Cancer Genome Project (CGP) based at the Wellcome Trust Sanger Institute and The Cancer Genome Atlas (TCGA) based at the US National Institute of Health (NIH). Studies by these large consortiums and independent laboratories have produced a multitude of studies in various cancers.

Sequencing of coding regions of breast and colorectal cancers suggest that somewhere between 80 and 90 non-synonymous changes are present in an individual tumour (Sjöblom *et al*, 2006; Wood *et al*, 2007). Following this, the sequencing of 20,661 protein coding genes in 24 prostate cancers identified a total of 1,562 somatic mutations, of which 25.5% were synonymous changes (3.8% nonsense mutation, 62.4% missense mutation, 5% small insertions/deletions and 3.3% splice site/untranslated region (UTR) mutation). Classification of the mutated genes identified in this study highlighted 12 core pathways that harbour mutations in 67-100% of all the prostate cancers studied (Jones *et al*, 2008). These pathways are; TGFβ signalling, JNK signalling, integrin signalling, Wnt/Notch signalling, Hedgehog signalling, control of the G1/S phase transition, apoptosis, DNA damage control, small

GTPase signalling, invasion, homophillic cell adhesion and K-RAS signalling (Jones *et al*, 2008). This work, and data from other studies, allowed for two important propositions to be made. Firstly, due to the large number of observed mutations, not all mutations within a cancer cell are directly related to tumourigenesis. Some mutations are thought of as driver mutations, necessary for tumourigenesis or tumour progression, and the rest as passenger mutations, randomly selected alongside driver mutations and thus retained. Secondly, that all these important cancer specific genetic alterations occur within the highlighted 12 core pathways (Jones *et al*, 2008; Ledford, 2010; Bell, 2010).

Recent exome sequencing studies have also identified novel tumour type specific somatic mutations in *GRIN2A* in 33% of metastatic melanoma cases and recurrent mutations of *TRAPP* in 4% of metastatic melanomas (Wei *et al*, 2011), frequent truncating mutations of the *PBRM1* gene in 41% of clear cell renal cell carcinomas (Varela *et al*, 2011) and mutations of *BAP1* in 81% of metastasizing uveal melanomas of the eye (Harbour *et al*, 2010).

These new techniques provide powerful new ways to look at the alterations at the base pair level for the whole genome of cancerous cells, providing important information. This information could also be combined with methods for looking at the genome wide epigenetic alterations to give a more complete picture of the cancer associated alterations as exemplified by a recent study in ovarian cancer where sequencing data, copy number data, DNA methylation data, and mRNA and microRNA expression data have all been combined (The Cancer Genome Atlas Network, 2011).

1.2 CANCER EPIGENETICS

There are two major epigenetic mechanisms that have a role in cancer; histone modifications and DNA methylation. The work in this thesis concentrates on the role of aberrant DNA methylation in specific cancers, but it is important to note that other forms of epigenetic control exist and can work both separately and in conjunction with DNA methylation.

1.2.1 Histone modifications in cancer

There are currently nine types of histone modifications known; acetylation, methylation, phosphorylation, ubiquitination, sumoylation, ADP-ribosylation, deimination, proline isomerism and propionylation (Sawan *et al*, 2010). In a normal cell, histone modifications function primarily in chromatin remodelling, various modifications or groups of modifications being able to adjust the chromatin from a less accessible heterochromatin form to a more accessible euchromatin form and vice-versa (Quina *et al*, 2006). Histone modifications are therefore able to have profound effects on gene transcription, DNA repair and DNA replication. Histone acetylation and methylation are the most commonly associated modifications associated with cancer.

Histone acetylation occurs at lysine residues. By removing the histone tail's positive charge, lysine acetylation of histones results in a weakening of nucleosome-nucleosome and nucleosome-DNA interactions, resulting in changes in chromatin conformation that render DNA more accessible. Acetylation of histones is achieved by histone acetylatransferases (HATs) and removed by histone deacetylases (HDACs). Both HATs and HDACs have been demonstrated to be involved in cancer. Two HATs, *p300* and *CBP*, are considered tumour suppressor genes and both proteins are involved in fusion proteins in leukaemias (Suganuma *et al*, 2002; Ida *et al*, 1997; Giles *et al*¸ 1997) and loss of heterozygosity (LOH) at the p300

lcous is associated with hyperacetylation in numerous cancers (Tillinghast *et al*, 2003; Koshiishi *et al*, 2004). Aberrant expression of HDACs has been shown in multiple cancers; upregulation of HDAC2, 3, 8, 6 and 7 has been observed in childhood ALL (Moreno *et al*, 2010), HDAC1, 2 and 3 upregulation has been observed in ovarian cancer (Hayashi *et al*, 2010), upregulation of HDAC6 has been observed in breast cancer and oral squamous cell carcinoma (Sakuma *et al*, 2006) and HDAC2 upregulation has been observed in numerous cancers (Langer *et al*, 2010; Mutze *et al*¸ 2010; Adams *et al*¸ 2010; Jin *et al*, 2008). Due to the overexpression of numerous HDACs in cancer they have become useful therapeutic targets, as discussed later.

Histone methylation can occur on arginine or lysine residues and is catalysed by different enzymes depending on where the methylation occurs. Arginine residues can be mono- or di-methylated whilst lysine residues can also be tri-methylated (Sawan *et al*, 2010). The effects of histone methylation on chromatin conformation are dependent on the context, aiding either a euchromatin or heterochromatin conformation. *LSD1* (*lysine (K)-specific demethylase 1A*), a histone demethylase, has been shown to be aberrantly expressed in cancer with overexpression observed in bladder cancer (Hayami *et al*, 2011), oestrogen receptor negative breast cancer (Lim *et al*, 2010) and neuroblastoma (Schulte *et al*, 2009). Recent genome wide sequencing studies have also identified mutations within lysine methyltransferase genes *MLL2* (*myeloid/lymphoid or mixed-lineage leukemia 2*) and *MLL3* (*myeloid/lymphoid or mixed-lineage leukemia 3*) in medulloblastomas (Parsons *et al*, 2011) and the demethylase gene *UTX* (*lysine (K)-specific demethylase 6A*) in various cancers (Dalgliesh *et al*, 2010; van-Haaften *et al*, 2009), reinforcing the potential role of histone methylation regulation in tumourigenesis. Additional mutations within the SWI/SNF (SWItch/Sucrose NonFermentable) chromatin remodelling complex genes, including *PBRM1* (*polybromo 1*) in

renal cell carcinoma (Varela *et al*, 2011) and *ARID1A* (*AT rich interactive domain 1A (SWI-like)*) in ovarian cancer (Jones *et* al, 2010; Wiegand *et al*, 2010), provide additional evidence for the important role of chromatin remodelling in cancer.

1.2.2 DNA methylation

DNA methylation is the addition of a methyl group to a DNA nucleotide. It occurs in many organisms, primarily at the carbon five position of cytosines, producing 5-methylcytosine (5mC). Methylation of other nucleotides can occur, for example, although DNA methylation in general is uncommon in bacteria, a small proportion of adenine nucleotides are sometimes methylated, along with cytosines, as a host defense mechanism (Strachan and Read, 2004). In multicellular organisms, the level of methylation varies between species. In drosophila, methylation is uncommon but when present often occurs at cytosines that are part of a CpT dinucleotide, whereas in mammals, methylation is common and occurs predominantly at cytosines that form part of CpG dinucleotides, although a small amount of methylation also occurs at CpNpG sequences, where N can be A, T or C. Due to the overwhelming proportion of methylation occurring at CpG dinculeotides rather than any other sites in the human genome, these other potential sites, although possible, are not normally considered when assessing methylation levels in human DNA.

A further possible modification of 5mC is the addition of a hydroxyl group producing 5-hydroxymethylcytosine (5hmC) and, until the recent discovery of 5hmC in human and mouse brains (Kriaucionis *et al*, 2009; Tahiliani *et al*, 2009), it was not thought to be present in higher organisms. Catalysis of 5hmC production can be achieved by the enzymes TET1 (tet oncogene 1), TET2 (tet oncogene family member 2) and TET3 (tet oncogene family member 3) (Tahiliani *et al*, 2009). Although it has been suggested that 5hmC may be produced as an

intermediary molecule during demethylation of 5mC (Kriaucionis *et al*, 2009; Tahiliani *et al*, 2009), the true significance and proportion of 5hmC in the human genome remains to be determined. A recent plethora of studies have tried to develop methods that reliably differentiate between 5hmC and 5mC for quantification of the levels of 5hmC in various tissues (Ficz *et al*, 2011; Globisch *et al*, 2011; Song *et al*, 2011; Szwagierczak *et al*, 2010). In the most recent study, by Ficz *et al* (2011), the greatest detection of 5hmC was at regions of euchromatin with 5mC being underrepresented at gene promoters and CpG islands. Studies of 5hmC in embryonic stem cells (ES cells) suggest 5hmC is likely to have a role in lineage commitment and pluripotency (Ficz *et al*, 2011). Recent studies have identified mutations in *TET2* in leukaemias (Ko *et al*, 2010; Figueroa *et al*, 2010; Abdel-Wahab *et al*¸ 2009). *TET2* mutated acute myeloid leukaemia (AML) samples demonstrate an increase in the levels of hypermethylation in the much the same way as *IDH1/2* mutated samples (Figueroa *et al*, 2010). Since the discovery of 5hmC is very recent and the true importance of its presence is not yet fully known, all further discussion will only refer to 5mC, as will the work in this thesis. However, it is important to bear in mind that the discovery of 5hmC in the human genome and recent demonstration of *TET2* mutations in human cancer is likely to add a new aspect to the field of cancer epigenetics over the coming years.

The average human genome content of G/C nucleotides is approximately 41%, however, the proportion of CpG dincleotides only occurs at about 20% of the statistically expected frequency based on the overall levels of G/C nucleotides (Lander *et al*, 2001 - International Human Genome Sequencing Consortium). This is due to the fact that cytosines at CpGs often become methylated, producing 5mCpG, and the 5mC is highly susceptible to sponataneous deamination, producing thymine. Following DNA replication, the mismatched T/G is repaired to T/A, producing a permanent alteration/mutation within the DNA sequence. Spontaneous

deamination of unmethylated cytosine produces uracil, which is recognised by the cell and corrected, unlike the thymine. As a result of this, the mutation rate of cytosine to thymine is much higher than the average nucleotide, estimated to be 8.5 times higher (Cooper *et al*, 2000), resulting in the loss of many CpG dinucleotides over time. Although the general level of CpG dinucleotides within the genome is low, high levels are observed at long repetitive sequences and regions known as CpG islands. CpGs at long repetitive sequences are often methylated whilst CpG islands are often unmethylated.

1.2.2.1 CpG islands

The accepted definition of a CpG island is a region having a GC content greater than 50% and an observed CpG to expected CpG ratio of greater than 60% (Gardiner-Garden *et al*, 1987). The overall G/C content within CpG islands is approximately 60% and the number of CpG dinucleotides can be ten times greater than any other similarly sized region of the genome. There are estimated to be approximately 27 000 CpG islands within the human genome (Lander *et al* - International Human Genome Sequencing Consortium, 2001) with approximately 50-60% of genes having at least one CpG island associated with them, a large number of which are important housekeeping genes (Wang *et al*, 2004; Larsen *et al*, 1992). Similarly, there are about 15 500 CpG islands within the mouse genome (Mouse Genome Sequencing Consortium, 2002). Often, but not exclusively, located at the 5' end of genes and in some cases extending over hundreds of nucleotides, the methylation status of CpG islands can have a large impact on gene expression, mainly via chromatin modifications. In general, it is thought that CpGs in CpG islands are much less likely to be methylated than other CpGs throughout the genome due to the selective pressure for them to remain unmethylated and active (Bird, 2002). Although it has now been shown that CpG islands can be methylated in normal tissues (Weber *et al*, 2005) this may relate to different types of cell differentiation and

activity. Also, a recent study identified regions of DNA that show differential methylation levels between different tissue types that are up to 2kb away from CpG islands, known as CpG shores (Irizarry *et al*, 2009). These regions have a median size of 255bp and have been suggested to be responsible for the majority of tissue differential methylation, rather than CpG islands (Irizarry *et al*, 2009). Differential methylation at these regions has also been implicated in cancer (Irizarry *et al*, 2009). Figure 1.4 illustrates examples of the locations of notable CpG regions within the genome and their general states of methylation in normal and cancer genomes.

1.2.2.2 DNA methylation machinery

DNA methylation is laid down by two classes of DNA methyltransferase enzymes, *DNMT1* (*DNA methyltransferase 1*) and *DNMT3a* and *3b* (*DNA methylatransferase 3a and 3b* respectively). Originally named *DNMT1* and *DNMT3* because of an additional molecule, *DNMT2*, discovered at the same time with high levels of similarity. However, little evidence for the ability of *DNMT2* to methylate cytosine was discovered and it has since been determined that *DNMT2* catalyses the methylation of position 38 in aspartic acid of tRNA (Goll *et al*, 2006). It is now also referred to as *TRDMT1* (*tRNA aspartic methyltransferase 1*).

Whilst both *DNMT1* and *DNMT3a/b* use the same methyl donor, S-adenosyl methionine, the two classes of enzyme target different states of DNA to methylate.

DNMT1

DNMT1 is often known as maintenance methyltransferase since it primarily methylates hemi-methylated DNA on the nascent strand of DNA following DNA replication (Bestor, 1992), thus providing a mechanism by which methylation status of DNA is maintained in daughter cells. *DNMT1* has been shown to be essential for maintaining DNA methylation patterns in

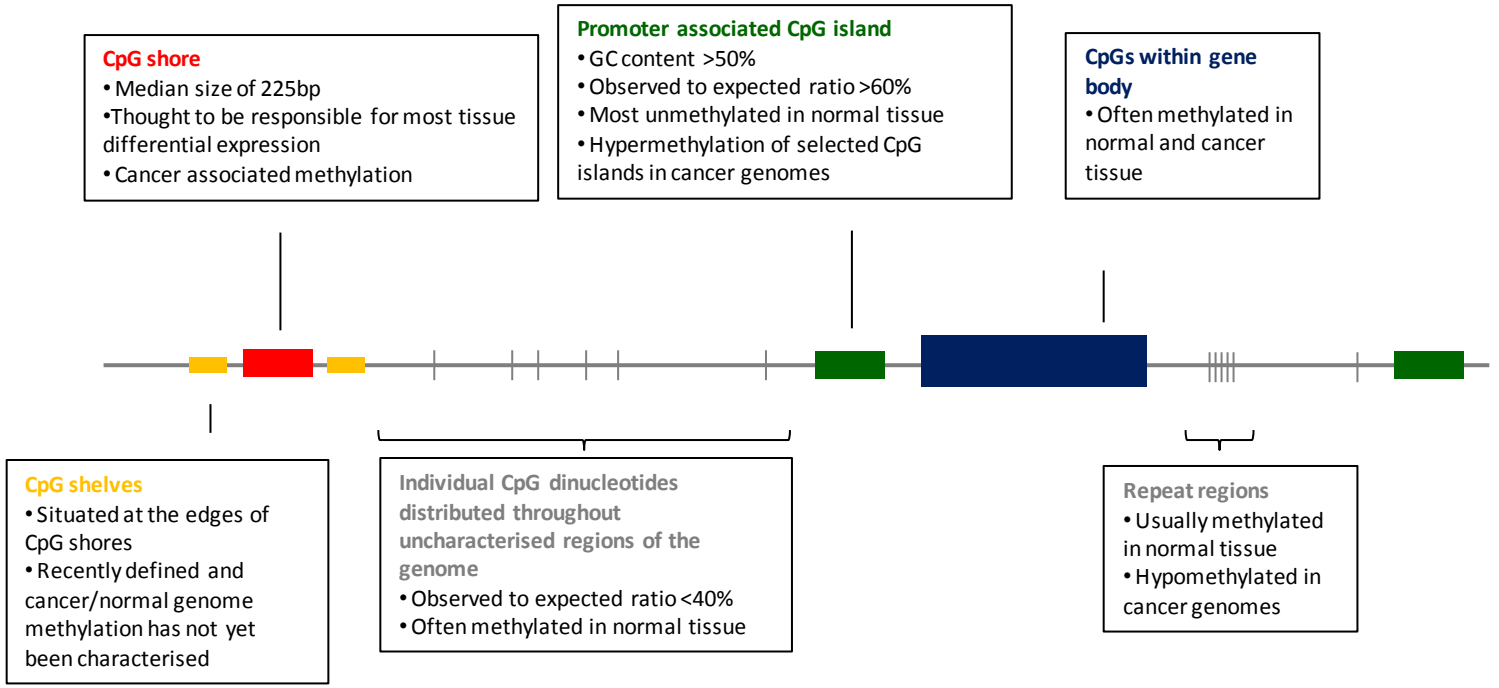


Figure 1.4_Distribution of CpG dinucleotides throughout the genome
This figure illustrates the location of CpG dinucleotides within the genome and whether they are methylated in normal or cancerous tissue.

proliferating cells (Li *et al*, 1992). Because of this, *DNMT1* is also essential for development, *Dnmt1*-/- mice demonstrate complete genome demethylation and developmental arrest at E8.5 (Li *et al*, 1992). DNMT1 lacks a methyl binding domain and therefore it is unknown exactly how DNMT1 is recruited to hemi-methylated DNA during replication. Binding of DNMT1 has been observed with a methyl-binding domain containing protein, MeCP2 (Kimura *et al*, 2003) and so this has been proposed as a possible mechanism.

DNMT3a and 3b

DNMT3a and *3b* are both considered *de novo* methyltransferases, required to establish new DNA methylation patterns (Okano *et al*, 1999). It is likely that *DNMT3a* and *3b* have different functions as *Dnmt3a*-/- mice demonstrate gut malformation, defects in spermatogenesis and are born alive, although die at around 4 weeks (Okano *et al*, 1999), whilst *Dnmt3b*-/- mice are embryonic lethal at E14.5-E18.5 demonstrating complete demethylation of minor satellite DNA and mid neural-tube defects (Okano *et al*, 1999). DNMT3b has been shown to interact with CENPC (centromere protein C), recruiting DNMT3b to pericentromeric satellite repeats; loss of either protein results in increased misalignment and segregation defects during mitosis (Gopalakrishnan *et al*, 2009). Mutations within *DNMT3b* cause ICF syndrome (Immunodeficiency-centromeric instability-facial anomaly syndrome), a rare autosomal recessive disease characterised by facial dysmorphism, immunoglobulin deficiency, branching of chromosomes 1, 9 and 16 after PHA (phytoheamaglutanin) stimulation of lymphocytes. A small number of patients also show hypomethylation of a fraction of the genome, which is likely to be a direct result of impaired *DNMT3b*.

1.2.2.3 DNA methyltransferases and cancer

A study by Chen *et al* (2007) demonstrated the necessity of cancer cells for functioning *DNMT1*. A conditional *DNMT1* knockout colorectal cancer cell line, HCT116, was made that demonstrated a reduction of CpG-CpG dyads within the genome and arrest at the G2 phase of the cell cycle, suggesting *DNMT1* is essential for maintenance of methylation in cancer cells and required for proliferation and survival (Chen *et al*, 2007). Whilst inhibition of *DNMT3a* and *3b* appear to have a lesser effect on cancer cells (Rhee *et* al, 2002; Robert *et al*, 2003), downregulation of the miRNA (miR)-29 family (29a, 29b, and 29c), which have complementary sites within the 3'-UTRs of *DNMT3a* and *DNMT3b*, in lung cancers have shown upregulation of DNMT3a and 3b, associating with worse prognosis (Fabbri *et al*, 2007). Upon re-expression of the miRNAs, normal patterns of DNA methylation were restored, as was the expression of some methylation silenced tumour suppressor genes (Fabbri *et al*, 2007). A recent study identified mutations in *DNMT3a* present in 22.1% of acute myeloid leukaemia patients (Ley *et al*, 2010). Whilst mutations were associated with worse overall survival, no significant differences in overall 5mC levels were observed in patients with mutations, however, it was noted that some regions showed significantly less methylation in a fraction of patients (Ley *et al*, 2010). Since this study, mutations in *DNMT3a* have also been identified in acute monocytic leukaemia (Yan *et al*¸ 2011) and myelodysplastic syndromes (Walter *et al*, 2011).

1.2.2.4 DNA methylation and gene expression

DNA methylation has little effect on gene expression if the DNA is not assembled into chromatin, as it would be *in vivo*, indicating the importance of the interactions affected by the methylation rather than the methylation itself. For the vast majority of genes, promoter CpG

island methylation results in the down regulation of gene expression and it is this effect that is observed as a frequent event in cancers (Feinberg *et al* 2004). However, DNA methylation has also been shown to produce enhanced expression, as seen with the IGF2 locus, by blocking the binding of repressor proteins to a silencer element within a gene promoter (Bell *et al*, 2000). Although this work concerns aberrant cancer related DNA methylation, DNA methylation is involved in the normal control of the cell. DNA methylation occurs on a large scale to cause the selected X chromosome inactivation early in embryogenesis (Heard, 2004), and on a single gene level when involved with imprinting, the specific methylation of either the maternal or paternal allele, resulting in single allelic expression, (Reik *et* al, 2003), or *de novo* methylation alteration associated with cell differentiation (Arney and Fisher, 2004). Repression of gene expression can be bought about by DNA methylation by either direct blocking of transcription factor binding or, as is more common, via the actions of methyl DNA binding proteins (Clouaire *et al*, 2008). Although this mechanism is rare *in vivo*, direct blocking of transcription factor binding by the presence of DNA methylation has been described in some instances. For example, the GFAP (glial fibrillary acidic protein) is a developmentally regulated gene activated during astrocyte differentiation. The GFAP gene promoter contains a STAT3 transcription factor binding site and upon removal of DNA methylation within this region, the GFAP gene is expressed due to STAT3 binding (Shimozaki *et* al, 2005).

1.2.2.5 Recruitment of methyl DNA binding proteins

The most common mechanism for DNA methylation to induce repression of gene expression is via the recruitment of DNA binding proteins and subsequent modification of chromatin organisation. There are three families of methyl DNA binding proteins: the Kaiso-like family

of proteins, the SRA (SET- and RING-associated) domain proteins and the methyl binding domain family of proteins.

The Kaiso-like family currently consists of three members, Kaiso, ZBTB4 and ZBTB38. Kaiso-like family members show affinity for both unmethylated and methylated DNA suggesting multiple roles are likely for this family. Kaiso itself is able to interact with the N-coR (nuclear receptor co-repressor) co-repressor complex which contains HDAC3 and therefore is capable of bringing about chromatin modification. For instance, the recruitment of Kaiso and N-coR due to DNA methylation of the MTA2 promoter is capable of causing MTA2 gene silencing (Yoon *et al*, 2003).

Little is known about the SRA (SET- and RING-associated) domain family of proteins other than one of the mammalian homologues, Np95, may have an important role in epigenetic inheritance and that the SRA domain is able to bind to CpNpG in addition to CpG (Johnson *et al*, 2007).

The methyl biding domain proteins are by far the biggest players in bringing about gene repression as a result of DNA methylation. The family consists of 5 members; MBD1, MBD2, MBD3, MBD4 and MeCP2, although not all can bind methylated DNA. The methyl binding domain is 75 amino acids long with a binding affinity dependent upon the presence of four specific residues, the loss of one of these residues in MBD3 renders it unable to bind methylated DNA (Hendrich *et al*, 1998).

MBD1

In the majority of cases, MBD1 brings about chromatin modification in a deacetylation-independent manner and most commonly acts via histone methylation. It has been shown to be able to interact with two histone H3 lysine 9 (H3K9) methyltransferases, SETDB1 (SET

domain, bifurcated 1) and SUV39H (suppressor of variegation 3-9 homolog), and thus upon binding with SETDB1, histone H3K9 becomes methylated. During the S phase of the cell cycle, the MBD1/SETDB1 complex is temporarily displaced and forms a comlpex with the p150 subunit of the chromatin assembly factor, CAF-1 (chromatin assembly factor 1), providing a mechanism by which the H3K9 chromatin modifications are transmitted to daughter cells (Sarraf *et al*, 2004).

MBD2 AND MBD3

Although MDB3 does not bind methylated DNA directly, both MBD2 and MBD3 separately interact with the DNA-binding NuRD (nucleosome remodelling and histone deacetylation) complex (Le Guezennec *et al*, 2006) which contains, among other things, a chromatin remodelling ATPase, HDAC1 and HDAC2.

MeCP2

MeCP2 contains a methyl binding domain and a transcriptional repression domain and is capable of binding both hemi-methylated and fully-methylated DNA. MeCP2 appears to be particularly important in neuron maturation; mutations in MeCP2 can cause Rett syndrome, a disorder characterised by mental retardation, small hands and feet, repetitive hand movements and encephalopathy (Amir *et al*, 1999). Mutations in MeCP2 have also been associated with autism (Swanberg *et al*, 2009). Gene repression by MeCP2 is achieved by recruitment of HDACs to remove the acetyl group from H3K9, methylation of H3K9 and subsequent recruitment of proteins that cause chromatin to become condensed, thus resulting in repression of gene expression (Bird *et al*, 1999) Interestingly, MeCP2 is able to both enhance and repress gene expression (Chahrour *et al*, 2008). Figure 1.5 shows a simplified schematic of how DNA methylation affects expression.

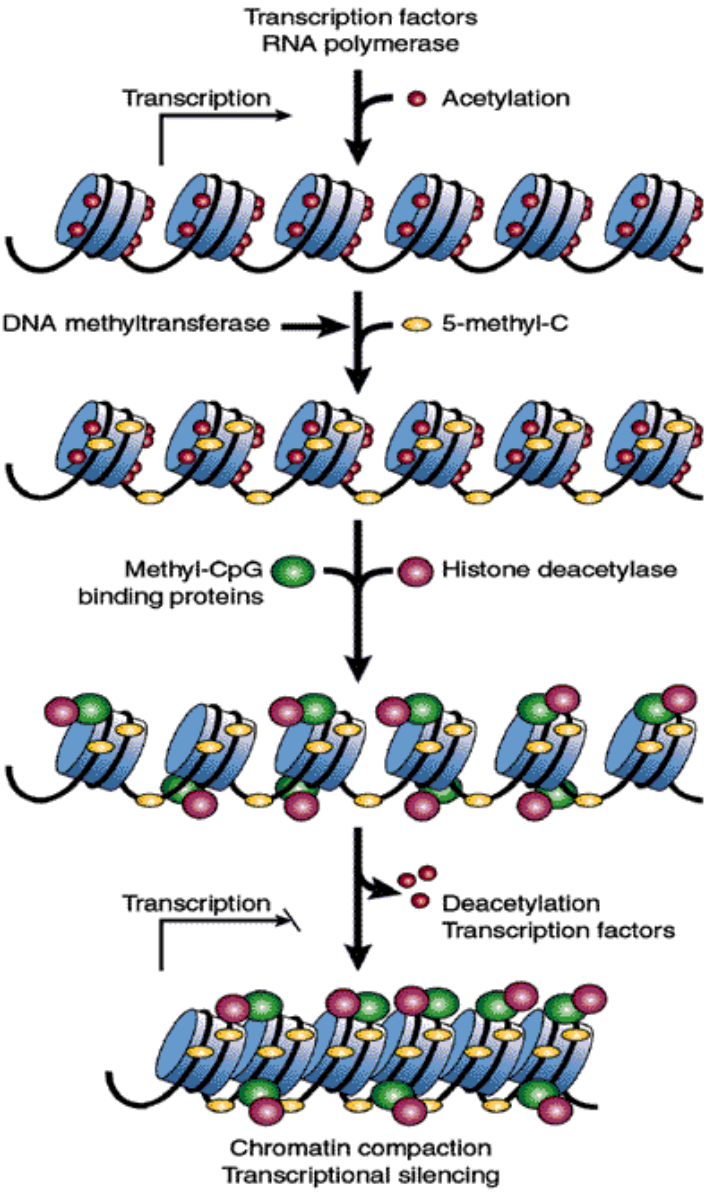


Figure 1.5_Gene transcriptional silencing due to DNA methylation
This figure illustrates how DNA methylation can induce transcriptional silencing. Areas of transcriptionally active chromatin regions tend to be acetylated and lack methylation to remain uncondensed and allow transcription factor binding. Transcriptionally silenced methylated regions have gained 5-methyl cytosines, via the action of DNA methylatransferases, which result in the recruitment of methyl-CpG binding proteins, histone deacetylases and other chromatin remodeling enzymes. Thus the region loses its acetylation and any bound transcription factors and the chromatin becomes condensed resulting in gene silencing and loss of transcription factor binding. This silencing would be replicated in any further DNA copies produced. This image was taken from: http://www.med.ufl.edu/biochem/keithr/fig1pt2.html

1.2.3 Aberrant DNA methylation changes in cancer

1.2.3.1 DNA hypomethylation

Despite cancer acquired hypermethylation of many genes, the overall 5mC levels within a cancer cell are often reduced due to global loss of methylation/hypomethylation. A common hallmark of cancer, hypomethylation most commonly occurs at repeat sequences, such as Alu and LINE1 elements, juxtacentromeric satellite 2 and tandem centromeric satellite α regions, possibly contributing to tumorigenicity by increasing the likelihood of genomic rearrangements and activating transposable elements (Ehrich *et al*, 2009). Although rarer, regional gene associated hypomethylation also occurs; for example, hypomethylation of synuclein γ has been demonstrated in breast and ovarian cancers (Gupta *et al*, 2003; Czekierdowski *et al*, 2006). Initially, it was supposed that hypomethylation was involved in more global events, such as genomic rearrangements, while hypermethylation was a gene specific event occurring as an alternative method of tumour suppressor gene silencing. However, more recently, an increasing number of gene specific hypomethylation studies have been performed. For example, a study in salivary gland adenoid cystic carcinoma demonstrated 8 genes that were tuomour specifically up-regulated due to loss of methylation (Shao *et al*, 2011). Nevertheless, there are far fewer examples of this and the overwhelming majority of studies involving cancer focus primarily on aberrant DNA hypermethylation.

1.2.3.2 DNA hypermethylation

Gene specific DNA hypermethylation in cancer cells as a form of gene inactivation was first demonstrated in 1986 with the discovery of hypermethylation and down regulation of the calcitonin gene in small cell lung cancer (Baylin *et al*, 1986). Subsequent analysis of *bone fide* tumour suppressor genes such as *RB1* (*retinoblastoma 1*) in retinoblastoma (Greger *et al*,

1989; Sakai *et al*, 1991), *VHL* (*von Hippel-Lindau*) in renal cell carcinoma (Herman *et al*, 1994) and *CDKN2A* (*cyclin-dependent kinase inhibitor 2A*) in multiple tumour types (Herman *et al*, 1995) demonstrated promoter hypermethylation as a common mechanism for inactivating certain tumour suppressor genes. Since then, many genes have been shown to be inactivated in this manner. In some cases, such as *CDH1* (*E-cadherin*), hypermethylation provides the 'second hit' to a tumour suppressor gene (Grady *et al*, 2000). In other cases, exemplified by *RASSF1A* (*Ras association (RalGDS/AF-6) domain family member 1*) and *HIC1* (*hypermethylated in cancer 1*), methylation is the only known mechanism of inactivation other than deletion. *RASSF1A* methylation is a common event in many cancers including lung, breast, glioma, colorectal and RCC (Dammann *et al*, 2000; Dammann *et al*, 2001; Hesson *et al*, 2004; van Engerland *et al*¸ 2002; Morrissey *et al*, 2001), yet no cancer acquired mutations have ever been identified (Lee *et al*, 2001; Dammann *et al*, 2000). *HIC1* methylation was first observed in leukaemias and breast cancers (Issa *et al*, 1997; Fujii *et al*, 1998; Melki *et al*, 1999) but has since been shown in multiple malignancies. The number of genes now shown to be hypermethylated in any cancer is well into the hundreds. Due to the large number of methylated genes in cancer it is unreasonable to think that they could all be *bone fide* tumour suppressor genes although many may contribute towards tumourigenicity or tumour progression, whilst others may be a by-product of general genome deregulation. They can be looked upon in a similar manner to the larger number of mutations per genome that are now being identified by next generation sequencing, where some are driver mutations (necessary for tumourigenesis) and the others are passenger mutations (randomly selected along with the driver mutations).

1.2.3.3 DNA hypermethylation events as biomarkers

The detection of single gene methylation status is a quick and easy procedure that could be undertaken in a diagnostic setting using PCR based techniques such as MSP and pyrosequencing, which will be discussed later. The presence of circulating cancer cells in bodily fluids for some types of cancer (for example, urine in prostate cancer and sputum in lung cancer), also allows for analysis of potential methylation markers from non-invasive techniques, an important consideration from the patients' point of view. Determination of gene specific DNA methylation events can be used to detect a small level of cancer cells (not visible by current imaging techniques), identify pre-cancerous lesions that are more likely to develop into cancers, aid in disease prognosis and indicate drug treatment outcomes. The following examples demonstrate these potential biomarker functions in numerous cancer types.

The *GSTP1* (*glutathione-S-transferase pi 1*) gene, located at 11q13 is involved in detoxification and has a large CpG island susceptible to tumour specific hypermethylation in multiple cancer types including prostate, breast and renal cell carcinoma (Esteller *et al*, 1998). In prostate cancer, very frequent methylation (80-90%) has been shown in cancerous tissue but not benign hyperplastic or normal prostate tissues (Jéronimo *et al*, 2001). Some success has been observed in studies trying to detect *GSTP1* methylation in urine/serum samples, which, if developed further, could offer the opportunity for diagnosis of malignant from benign prostate lesions using an easy, non-invasive technique. This could be used in combination with the prostate-specific antigen (PSA) test to improve prostate cancer prediction.

As previously mentioned, hypermethylation of *CDKN2A* is a common event in many cancers. It is a potent tumour suppressor gene and as such, cellular inactivation is likely to drive the tumourigenic process. For this reason, high levels of methylation are also observed in pre-cancerous lesions and often associate with those that become cancerous. For example, the detection of *CDKN2A* methylation in Barrett's oesophagus significantly associates with progression to oesophageal cancer (Schulman *et al*, 2005; Wang *et al*, 2009), a scenario that has also been observed in lung, colon and liver lesions (Licchesi *et al*, 2008; Kukitsu *et al*, 2008; Jicai *et al*, 2006).

The *SEPT9* (*septin 9*) gene, located at 17q19, is involved in cell cycle control and methylation has been observed in colorectal and head and neck cancers (Lofton-Day *et* al, 2008; Stanbery *et* al, 2010). When this tumour specific methylation is present in colorectal cancer, it can be detected in patient plasma samples and subsequently developed assays have demonstrated a high sensitivity of CRC detection from plasma samples (de Vos *et al*, 2009). This is potentially exploitable as a sensitive, non-invasive diagnostic or general screening technique for a cancer type that can often lie undetected for an unfortunately long amount of time before presenting as advanced disease.

The *DAPK1* (*death associated protein kinase 1*) gene, located at 9q21.33, is involved in apoptosis regulation and frequent promoter hypermethylation has been observed in many cancers including colorectal, NSCLC and cervical (Pehlivan et al, 2010; Feng *et al*, 2008; Missaoui *et al*, 2011). Methylation in NSCLC has been shown to significantly associate with poor patient outcome with numerous studies showing a significant association between patients demonstrating *DAPK1* methylation in an early tumour grade that go on to have poor disease free- and overall-survival (Lu *et al*, 2004; Buckingham *et al* 2010). Similar results for

other cancers also look promising but remain to be confirmed in multiple studies and large cohorts.

The *MGMT* (*O-6-methylguanine-DNA methyltransferase*) gene, located at 10q26, is a DNA repair gene responsible for removing alkyl groups from the O6 position of guanine. If not removed, alkyl groups can form cross linking of the DNA strands resulting in double strand breaks, ultimately resulting in cell death. Some cancer drugs, such as carmustine and temozolomide, exploit this mechanism by directly causing massive amounts of alkylation to greatly increase the DNA cross-linking. Therefore, when *MGMT* is down-regulated, the increased and unrepaired alkylation shifts the cells' fates rapidly towards cell death, thus improving the efficiency of these drugs. Cancer specific hypermethylation has been observed in many cancer types but in gliomas the methylation status of *MGMT* is both frequent (40%) and highly clinically relevant (Esteller *et al*, 1999; Esteller *et al*, 2000a). *MGMT* methylation status has been shown to indicate response to alkylating drugs such as carmustine and temozolomide (Esteller *et al*, 2000a) and is now used in a diagnostic setting to identify those patients most likely to benefit from alkylation based drug therapy. This is a wonderful example of how methylation status can be used to help tailor successful treatments. Other examples include *BRCA1* methylation and *ABCB1* and *GSTP1* methylation in breast cancer indicating increased sensitivity to cisplatin and doxorubicin respectively; *CHFR* methylation in endometrial cancer and increased sensitivity to taxanes; and *ASC/TMS1* methylation and increased resistance to 5-fluorouriacil in gastric cancer (Deng *et al*, 2010).

1.2.3.4 CpG island methylator phenotype (CIMP)

Studies in colorectal cancer identified a subset of cancers that showed a specific increase in both the number of methylated genes and the levels of methylation at these loci, often with

subsets of specific genes being methylated in these samples (Toyota *et al*, 1999a). This led to the proposal of the CpG island methylator phenotype (CIMP) (Toyota *et al*, 1999a). In colorectal cancer, CIMP is strongly associated with microsatellite instability and methylation of the mismatch repair gene, *MLH1* (*mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)*), although four molecular subtypes have been described: CIMP positive & MSI negative; CIMP & MSI positive; CIMP negative & MSI positive, and CIMP & MSI negative (Toyota *et al*¸ 2000). Although CIMP is mostly associated with colorectal cancer, other tumour types, such as gastric and ovarian (Toyota *et al*¸1999b; Strathdee *et al*¸ 2001), have exhibited a CIMP like phenotype. A CIMP phenotype in gliomas, termed G-CIMP has recently been described that is most prevalent in *IDH1* mutated samples (Noushmehr *et al*, 2010).

1.2.3.5 Loss of Imprinting (LOI) in cancer

Loss of imprinting (LOI) in cancer has been described both on a germline and somatic level, but is an unusual event. Some imprinting disorders, such as Beckwith-Weiderman syndrome (BWS), are associated with an increase in the likelihood of developing tumours; 5-10% of BWS patients develop tumours including Wilms Tumour, neuroblastomas and rhabodomyosarcomas (Lim *et al*, 2010). BWS (and Silver Russell syndrome) are caused by imprinting disorders at the 11p15.5 region which contains multiple genes, including *IGF2*, which is implicated in BWS. In addition to the increased likelihood of BWS patients developing cancers, somatic LOI of *IGF2* can occur as an early event in Wilms Tumour, colorectal cancer and ovarian cancer (Lim *et al*, 2010). Other imprinted regions, such as 14q32, are involved in cancer acquired changes, where hypermethylation of either the differential methylated region involved in the imprinting control or promoter region of *GLT2*

have been described in neuroblastomas, Wilms Tumour and phaeochromocytomas (Lim *et al*, 2010).

1.2.3.5 Cancer epigenomics

The development of genome wide techniques to assess DNA methylation, as described later, has changed the field from predominantly analysis of a single locus, candidate gene to using global aberrant DNA methylation data as a method of identifying multiple novel important loci as well as profiling whole cancer genomes with respect to their global methylation changes according to biological factors, such as clinical features. Recent studies include identifying candidate genes associating with longer progression free intervals in ovarian cancer (Bauerschiag *et al*, 2011) and prediction of relapse in prostate cancer (Kobayashi *et al*, 2011). Methylation profiles have also been described that associate with translocation events in acute myeloid leukaemia (Alvarez *et al*, 2010) and, as previously mentioned, CIMP in gliomas that associated with *IDH1* (*isocitrate dehydrogenase 1*) mutation and better survival (Noushmehr *et al*, 2010; Christensen *et al*, 2011). These studies are applicable to all cancer types and have been and will be extremely useful for identifying methylation patterns and methylated genes that associate with specific clinical features and tumour sub-types, thus producing biomarkers that are complete profiles of gene methylation as well as single or groups of genes. Further examples will be discussed later with specific association to breast cancer.

1.2.3.6 Epigenetic therapies

Due to the intrinsically reversible nature of DNA methylation and chromatin modifications, compounds or drugs designed to readjust the cancer acquired changes in these gene expression modifiers are attractive therapeutic targets for restoring more normal expression.

Drugs designed to block the DNA methyltransferases (DNMTs) have proved to be useful in cancer management, against haematological malignancies in particular. There are currently two DNMT inhibitors that have been approved for use in the clinical setting; 'Vidaza®' (5-aza-deoxycitidine) (Celgene Corporation) and 'Decitibine/Dacogen®' (5-aza-2'-deoxycitidine) (SuperGen, Inc.), both of which have been approved for the treatment of myelodysplastic syndrome. There are many more DNMT inhibiting drugs that are being developed, for example Zebularine, Procaine and Hydralazine. Unfortunately, due to their common mechanism of action, the DNMT inhibitors often affect the transcription of off-target, non-cancer involved genes, although still having an effect on genes important in leukaemogenesis (Rodríguez-Paredes *et al*, 2011). Clinical trials are also underway to assess the benefits of epigenetic drugs in solid tumours. A trial has been started in non-small cell lung cancer using a combination of Vidaza® and a HDAC inhibitor that is showing some signs of success (Kaiser, 2010).

A large abundance of drugs designed to block chromatin modifications are also being developed. HDAC inhibitors are currently the most heavily researched, falling into one of four categories; short chain fatty acids, hydroxamic acids, cyclic peptides and benzamide derivatives. Two HDAC inhibitors have so far been accepted for patient treatment, 'Vorinostat' and 'Romidepsin'; as with the DNMT inhibitors, the most promise for HDAC inhibitors has been observed in haematological malignancies. The HDAC inhibitors most commonly bring about anti-cancer effects via cell cycle arrest at the G1or G2-M phase, induction of differentiation and apoptosis (Rodríguez-Paredes *et al*, 2011). A range of other targets for drug induced blocking of chromatin modifications are also being investigated, for example, inhibitors of histone acetylases, methylases and demethylases along with SIRT inhibitors (Rodríguez-Paredes *et al*, 2011).

1.3 ANALYSIS OF DNA METHYLATION

DNA methylation status can be analysed on the scale of an individual gene or on a larger scale, such as a whole chromosome or the entire genome, although in reality, even when using large scale approaches, it is often necessary to further analyse selected targets in an individual manner. There are now many ways in which this can be achieved.

Since DNA methylation is not retained by artificial PCR based DNA amplification, most analysis is based around techniques that first modify the DNA, using sodium bisulphite to permanently retain the methylation markers, even after PCR has subsequently been performed. This modified, amplified DNA can be assessed for methylation.

Previous to this discovery, general methylation status could be ascertained using large amounts of unamplified genomic DNA using paired methylation sensitive or methylation insensitive restriction endonucleases such as HpaII and MspI. Both HpaII and MspI recognise the sequence 5'-CCGG-3'; however HpaII is 5mC sensitive and MspI is 5mC insensitive; therefore cleavage with HpaII will only occur when methylation is not present, and MspI will cleave DNA regardless. Analysis of results post digestion by 2 dimensional (2D) Southern Blotting will give an indication of methylation status and produce specific fragments due to methylation. This technique is labour intensive and requires the use of radioactive material for labelling and has now been almost entirely replaced by more modern methods, although it was used as recently as 2006 (Giotopoulos *et al*, 2006). A PCR based technique was also developed, HpaII-PCR, based on HpaII digestion where PCR primers were designed to surround a HpaII site, therefore only producing a product if DNA remained undigested, indicating CpG methylation (Singer-Sam *et al*, 1990). This technique is now also very rarely used.

1.3.1 DNA modification by sodium bisulphite treatment

It was discovered in the early 1970s that treatment of denatured DNA with sodium bisulphite caused rapid deamination of cytosine to uracil, whilst 5-methylcytosine (5mC) remained unaffected (Shapiro *et al*, 1970; Hayatsu *et al*, 1970). This discovery revolutionised the way 5mC levels in DNA were analysed. PCR of the bisulphite converted DNA resulted in the incorporation of T nucleotides where bisulphite conversion had produced deamination of cytosine to uracil whilst 5-methylcytosine nucleotides were unaffected and therefore incorporated as cytosines in the PCR reaction. This was first used to demonstrate the methylation status of cytosine residues within the DNA in a study by Frommer *et al* (1992) where PCR products were sequenced to demonstrate that retained Cs were originally methylated and newly created Ts occurred where no methylation was present.

Since then, many techniques have been developed to assess the methylation levels in particular DNA sequences, the majority of which rely on bisulphite conversion and subsequent PCR; a few of these will now be discussed below. It is important to note that sodium bisulphite will convert 5-hydroxymethycytosine as well as 5-methylcytosine and so does not allow for any discrimination between these two different types of methylation (Huang *et al*, 2010).

1.3.2 CoBRA (combined bisulphite restriction analysis)

CoBRA (combined bisulphite restriction analysis) was developed as a rapid way of detecting methylation in bisulphite modified DNA (Xong *et al*, 1997). Following bisulphite conversion, PCR amplification is performed on the region of interest, and subsequently digested with an enzyme that will only cleave DNA if methylation was present in the original unmodified DNA. The two most commonly used enzymes are BstUI and TaqαI. BstUI recognises the

sequence 5'-CGCG-3' and therefore if originally methylated, the DNA will remain 5'-CGCG-3' following bisulphite modification and subsequent PCR, whereas if originally unmethylated, the sequence will become 5'-TGTG-3, therefore only retaining a BstUI restriction site if methylated. TaqαI recognises the sequence 5'-TCGA-3' and therefore if originally methylated, the bisulphite modified and subsequently amplified DNA will remain 5'-TCGA-3' whilst unmethylated DNA will become 5'-TTGA-3', rendering the DNA uncleaved following incubation with TaqαI. Additionally, TaqαI has the advantage of also being more specific to successfully modified DNA since if the original methylated DNA had the sequence 5'-CCGA-3', the 5'-C will become T, producing a TaqαI restriction site (if the following C is methylated). Despite this, BstUI is more commonly used as BstUI sites are more common within CpG islands and digestion only occurs if both CpGs are methylated, implying a greater depth of methylation. A schematic is shown in figure 1.6 of the CoBRA assay.

CoBRA assay is very commonly used to assess methylation status as it is a reliable, cost effective and rapid way of determining methylation status. Additionally, the PCR products from the CoBRA reactions can be cloned into bacteria vectors, grown and subsequently sequenced to give an indication of the range and depth of methylation throughout the investigated region (numerous clones are picked and sequenced to produce a ratio of unmethylated to methylated for a given region and/or individual CpG)

1.3.3 Methylation specific PCR (MSP)

Methylation specific PCR (MSP) is another commonly used PCR based technique on bisulphite modified DNA. It is based upon designing primers that are capable of specifically amplifying either methylated DNA or unmethylated DNA. Thus for each sample, two PCRs

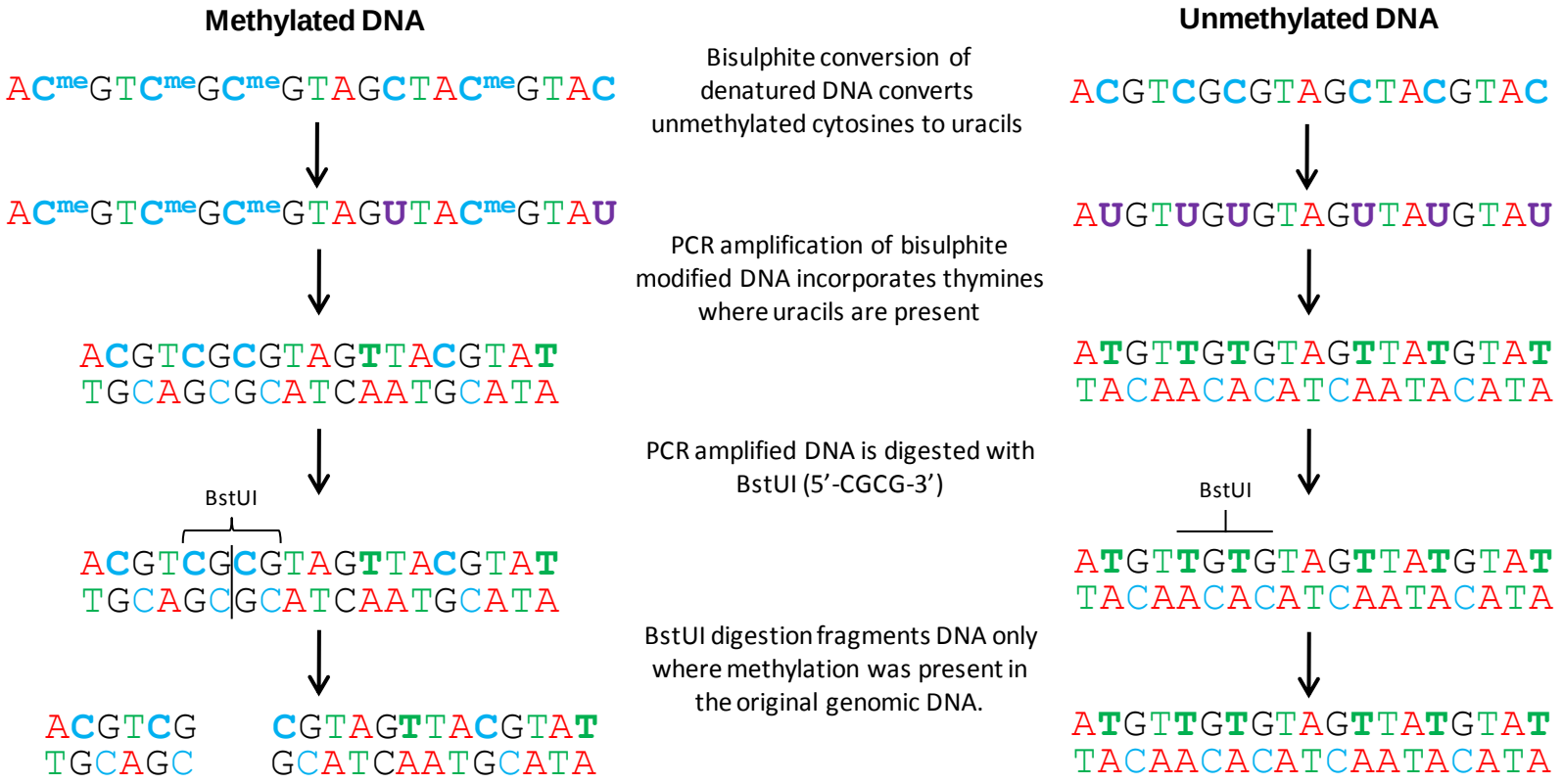


Figure 1.6_CoBRA assay for analysis of single gene methylation status
This figure shows a representative sequence of DNA in a methylated and unmethylated state as it goes through the bisulphite modification, subsequent PCR and digestion with a CG containing restriction endonuclease, in this example BstUI.

are performed, one designed to only amplify methylated DNA, the MSP (methylation specific PCR), and the other designed to only amplify unmethylated DNA, the USP (unmethylated specific PCR) (Herman *et al*, 1996). This technique can offer a more sensitive approach to detecting methylation than CoBRA and is often used to assess the methylation of known commonly methylated genes such as *RASSF1A* and *MGMT*. Due to the nature of primer design, it is often advantageous to know the regions of commonly methylated DNA first to ensure a sensitive assay. The only methylation assessed areas in a MSP lie directly beneath the primers, so for well characterised genes where the regions of important methylation alteration are known this is a very useful technique. Whilst MSP assays can be designed for the analysis of novel genes, prior information gathered from sequencing of CoBRA products can prove very useful.

1.3.4 Pyrosequencing

Pyrosequencing of methylated DNA was developed as a sensitive and quantitative method for analysing methylation status of CpGs. The basic pyrosequencing technique is based on incorporation of new dinucleotides to a template that are released in a pre-determined order; when the correct nucleotide is released, it binds to the template strand, releasing pyrophosphate. In turn, the pyrophosphate is converted during an enzyme catalysed reaction to produce light emissions, quantification of which gives a reading of the proportion of the incorporation of one nucleotide to another. The technique is therefore highly amenable to determining the proportion of C or T dinucleotides at a given CpG following bisulphite modification and PCR, allowing for quantitative assessment of single CpG methylation within a given sample (England *et al*, 2005; Uhlman *et al*, 2002). The technique offers a quantitative way of assessing methylation, which CoBRA and MSP assays both lack, however, the technique is less cost effective and is limited by relatively short sequence read-lengths and so

should be directed at a known region of importance. Figure 1.7 illustrates the pyrosequencing technique.

1.3.5 MethyLight™

The MethyLight™ method uses quantitative PCR techniques of the TaqMan™ system, which utilises a sequence specific probe, designed to anneal to PCR products, emitting a fluorescent signal when binding occurs which can be accurately measured throughout the PCR reaction. By performing qPCR on bisulphite modified DNA using a methylation pattern specific TaqMan probe, quantitative measurements can be made that determine the levels of methylation that lie within that particular probe sequence within a sample (Eads *et al*, 2000). The technique is quantitative, sensitive, rapid, amenable to high throughput analysis and relatively cost effective compared to some other methods however it is limited to analysing the methylation status at a specific, short sequence. It has become a common method of DNA methylation analysis since its discovery.

1.3.6 EpiTYPER™

EpiTYPER™ is a method commercialised by Sequenom to quantitavely measure 5mC levels at CpG sites by the use of MALDI-TOF (Matrix-assisted laser desorption/ionization-time-of-flight) mass spectrometry. It is based on a study by Ehrich *et al* (2005) in which it was demonstrated that bisulphite treatment followed by PCR and then *in vitro* transcription leads to a single stranded RNA molecule that can be base-specifically cleaved by an endoribonuclease, producing a mixture of fragments, the mass of which will depend on the CG content. Examination of the mixture of fragments by MALDI-TOF mass spectrometry provides enough information to calculate exactly which cytosines are methylated. Whilst this technique offers a quantitative way of assessing methylation without sequencing, it is a rather

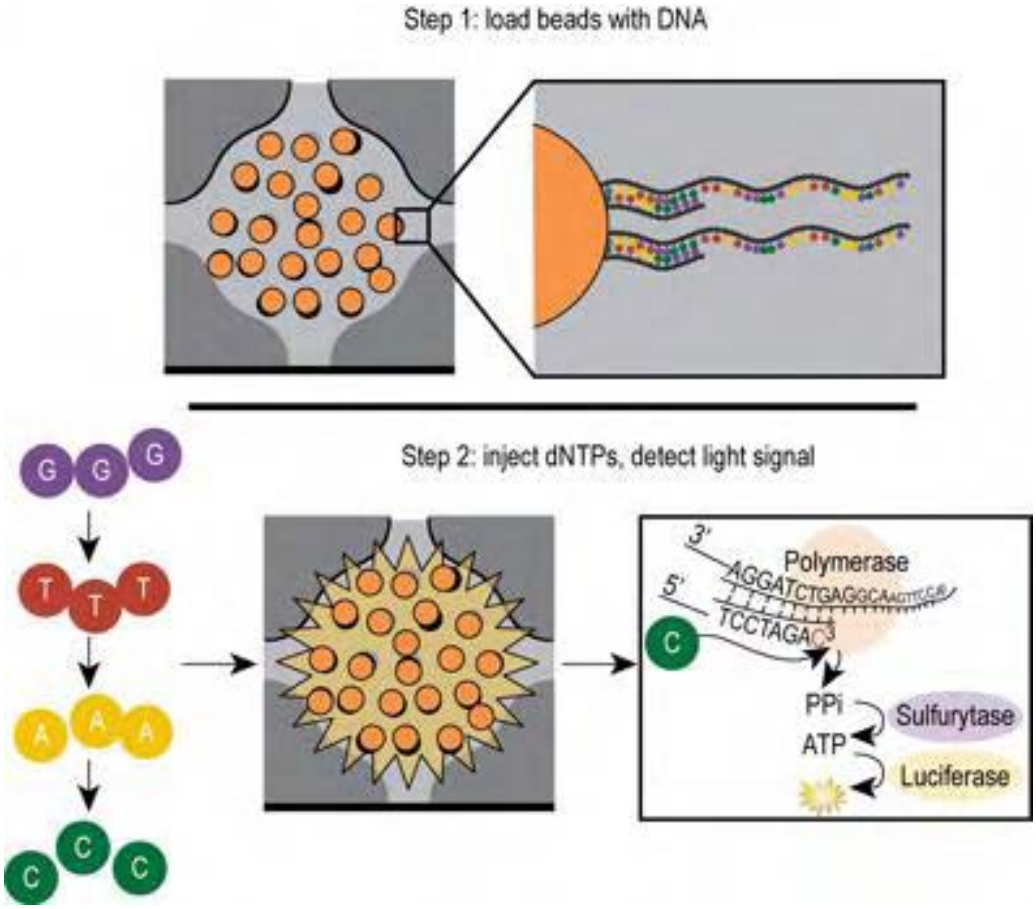


Figure 1.7_Schematic of the pyrosequencing technique

This figure shows the pyrosequencing technique for DNA methylation analysis. Bisulphite treated DNA is PCR amplified by using primers located around the CpGs of interest and a single strand loaded onto the beads. The bead probe is extended one nucleotide at a time by the sequential introduction of all four possible nucleotides, A, C, G or T. Incorporation of the dNTP creates by-products that result in an enzymatically produced light signal. This signal is measured after the introduction of each nucleotide and thus the sequence can be read. If the amplified region is partially methylated a percentage of amplified fragments will give a signal with the introduction of C and some with the introduction of T. The light signal can be accurately quantified and produce a quantitative level of DNA methylation. This figure was taken from http://www.dna-sequencing-service.com/dna-sequencing/pyrosequencing/

a cost heavy procedure. Recent uses have included validation of genome wide arrays (Christensen *et al*, 2010; Flanagan *et al*, 2010) and methylation studies of known candidate genes in clinically variable samples for biomarker discovery (Vanaja *et al*, 2009). Figure 1.8 illustrates the EpiTYPER™ technique.

1.3.7 Larger Scale/Genome wide methylation analysis

1.3.7.1 De-methylation (5azaDC) specific mRNA expression array analysis

This technique utilises expression array data generated from the total mRNA extracted from cell line samples either treated or untreated with the de-methylating drug 5-aza-2'-deoxycitidine (5azaDC). Analysis of the comparative gene expression with and without treatment can identify sets of genes that are up-regulated following de-methylation for each sample, suggesting gene methylation in the untreated sample is restricting expression. These arrays have the advantage of identifying methylated genes that are more likely to be biologically relevant, as the result of de-methylation is re-expression or up-regulation. However, this can only be performed on cell lines due to the necessity of 5azaDC treatment. It is also important to note that the up-regulated genes do not necessarily have to be themselves methylated. Methylation of upstream control genes, such as transcription factors, can also be removed and thus result in the up-regulation of their targets. This leads to a high degree of false positive gene selection and investigation of these genes often demonstrates no tumour specific DNA methylation within the cell lines (Morris *et al*, 2010; Dr C.J. Ricketts, personal communication). Nevertheless, numerous successful studies include the identification of potential methylation markers in lung cancer (Shames *et al*, 2006) and potential tumour suppressor genes in melanoma (Mithani *et al*, 2011), and renal cell carcinoma (Morris *et al*, 2010).

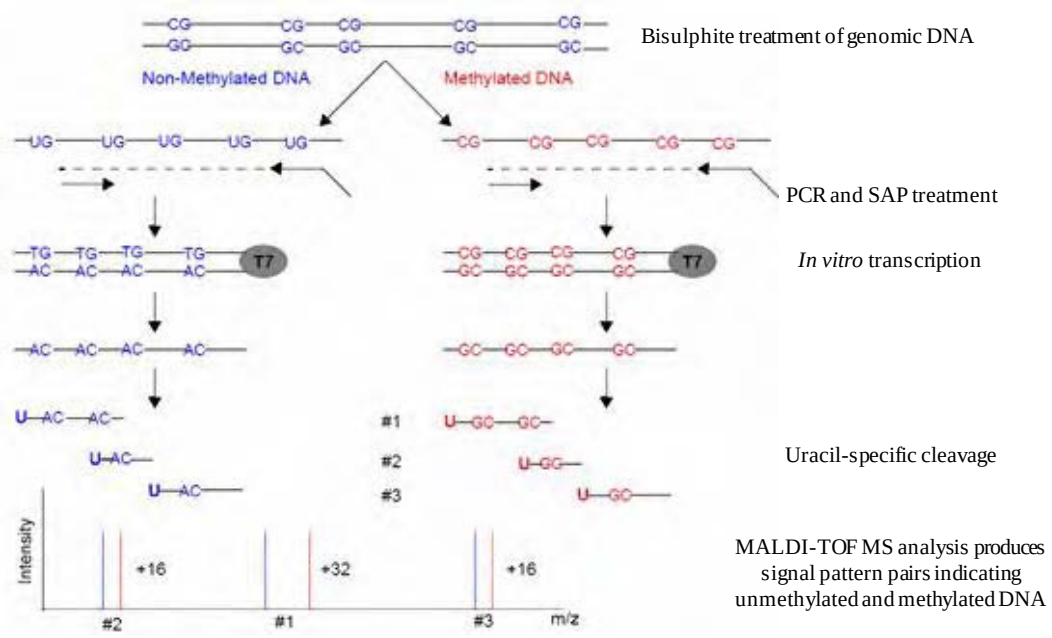


Figure 1.8_Schematic of the EpiTYPER technique
This figure illustrates the EpiTYPER technique. Bisulphite treated DNA is PCR amplified by using primers located around the CpGs of interest with one primer tagged with a T7 promoter sequence. The PCR product is transcribed into a RNA transcript and cleaved in a base specific manner. Cleavage products are analyzed by matrix-assisted laser desorption/ ionization – time of flight (MALDI-TOF) mass spectrometry (MS) and a characteristic mass signal pattern can be obtained, here the blue signal represents unmethylated DNA and red represents methylated DNA. This image was taken from Ehrich *et al*, 2005.

1.3.7.2 Methylated DNA enrichment assays

Following an initial experiment in 1994 that demonstrated that methylated CpG island DNA could be immunoprecipitated and enriched for by using an antibody specific to the methyl binding protein, MECP2 (Cross *et al*, 1994), further techniques have been developed that also allow for either the pull-down or immunoprecipitation of methylated DNA from a pool of short DNA fragments produced by whole genomic DNA sonication. These methods use either antibodies directly against the 5mC itself or methyl binding proteins, such as MBD1. When coupled with microarrays, these techniques offer a way of identifying differential methylation between samples without bisulphite modifying the DNA.

Methylated DNA immunoprecipitation (MeDIP)

The most widely used commercially available approach is MeDIP (methylated DNA immunoprecipitation) which utilises an antibody against 5-methylcytosine to pull down methylated DNA fragments (Weber *et al*, 2005). The approach is then coupled with microarray analysis to compare the immunoprecipitated, methylation enriched DNA to a control sample, most commonly used are the CpG island specific microarrays. The technique has been used for many studies; including assessing methylation levels in several cancers, such as colorectal cancer (Keshet *et al*, 2006), testicular cancer (Chueng *et al*, 2010) and RCC where the MeDIP technique was used on 9 RCC tumours and 3 normal kidney tissue samples to identify novel methylated genes in RCC (Morris *et al*, 2011). A recent study has also used the MeDIP technique to assess the methylation differences between familial *BRCA1*, *BRCA2* and *BRCAn* mutated breast cancer samples (Flanagan *et al*¸ 2010).

Methylated CpG island recovery assay (MIRA)

The main alternative to MeDIP is the methylated CpG island recovery assay (MIRA). This method involves the pull-down of fragmented methylated DNA through the use of methyl binding domain (MBD) proteins. The assay relies on the affinity of the MBD protein, MBD2b, for methylated DNA. The affinity of MDB2b for methylated DNA is enhanced by the MBD-like protein, MBD3L1 and therefore the MIRA assay uses a complex of full length glutathione-S-transferase (GST)-tagged MBD2b and His tagged MBD3L1 (Rauch *et al*, 2005).

The fragmentation of genomic DNA was initially performed by sonication, but now is more commonly achieved via restriction endonuclease digestion with an enzyme with an AT rich cut-site such as MseI (5'-TTAA-3') to avoid cleavage within CpG rich regions. DNA is incubated with the MBD2b/MBD3L1 protein complex and subsequently pulled-down with glutathione coated GST beads. The resultant methylation enriched DNA has been both PCR screened for the presence/absence of a particular gene (Rauch *et al*, 2005) and compared to normal control DNA by hybridisation to CpG island specific microarrays (Rauch *et al*, 2006) (figure 1.9). Analysis has shown the pull-down of methylated fragments is sensitive to a small number of methylated CpGs (Rauch *et al*, 2005).

To date, the MIRA technique has been used to successfully assess tumour acquired methylation in lung cancer (Rauch *et al*, 2007; Rauch *et al*, 2006), ductal carcinoma in situ (DCIS) (Tommasi *et al*, 2009), childhood acute lymphocytic leukaemia (Dunwell *et al*, 2010), breast cancer (Chapter 3; Hill *et al*, 2010) and astrocytomas/gliomas (Wu *et al*, 2010). In each case, novel methylated genes in these cancers have been identified, showing MIRA to be a useful technique when trying to identify novel methylated targets in cancers.

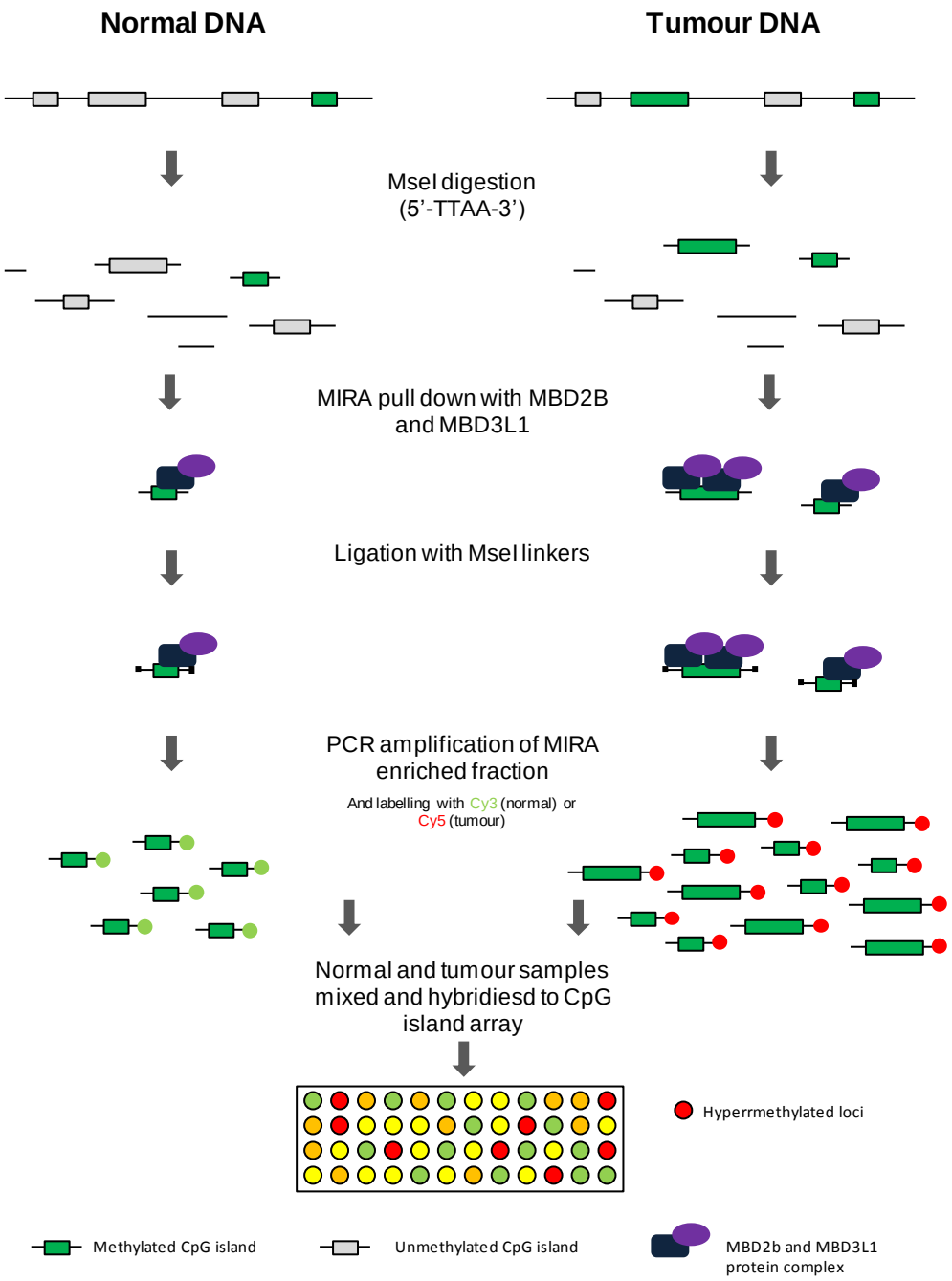


FIGURE 1.9_Schematic of the MIRA coupled with microarray technique
This figure shows a schematic of the MIRA assay. Briefly, genomic DNA is digested with MseI to fragment the DNA outside of CpG islands. Fragmented DNA is then incubated with an MBD2b and MBD3L1 protein complex to pull down methylated DNA fragments. Samples are ligated with MseI linkers and then PCR amplified and labelled with either Cy3 (normal) or Cy5 (tumour). Hybridisation to a CpG island microarray determines which fragments are hypermethylated in tumour DNA.

1.3.7.3 Direct hybridisation of sodium bisulphite modified DNA to microarrays

The Illumina methylation bead chip assays currently offer the only possibility of direct array hybridisation of sodium bisulphite modified DNA. There are currently three bead chip assays available from Illumina; GoldenGate, Infinium HumanMethylation27 and Infinium HumanMethylation450. All assays use the same Bead Chip technology originally designed to assess single nucleotide polymorphisms (SNPs), which has been adapted to assess the methylation of a single CpG locus based on the C or T status following bisulphite modification and PCR of 5mC and C respectively. Bead probe design is based much like an MSP or USP primer, with any upstream CpG dinculeotides of the one being assessed altered to correspond with the last nucleotide of the probe, which is either C (methylated) or T (unmethylated). Correct binding of methylated or unmethylated DNA to the bead chip results in single base extension with the addition of a fluorescently labelled nucleotide. For each probed CpG locus, numerous methylated and unmethylated probes are scattered throughout the chip and measurement of the fluorescence of all the probes for each locus bead type gives a quantitative representation of the amount of methylated and unmethylated alleles in the analysed sample. A schematic of the assay is shown in figure 1.10. The three types of methylation bead chip assays differ only in terms of array content. GoldenGate covers 1536 CpGs across a relatively small selection of genes (n=807) which were selected based on either being useful for cancer research, being X-linked or being imprinted. Infinium HumanMethylation27 covers 27, 578 CpGs across 14,475 consensus coding sequence (CCDS) genes, the majority of assessed CpGs being within Illumina defined CpG islands. The most recently available, Inifnium HumanMethylation450, covers a more comprehensive selection of the genome, with >450 000 CpG sites across a large number of genes and non-

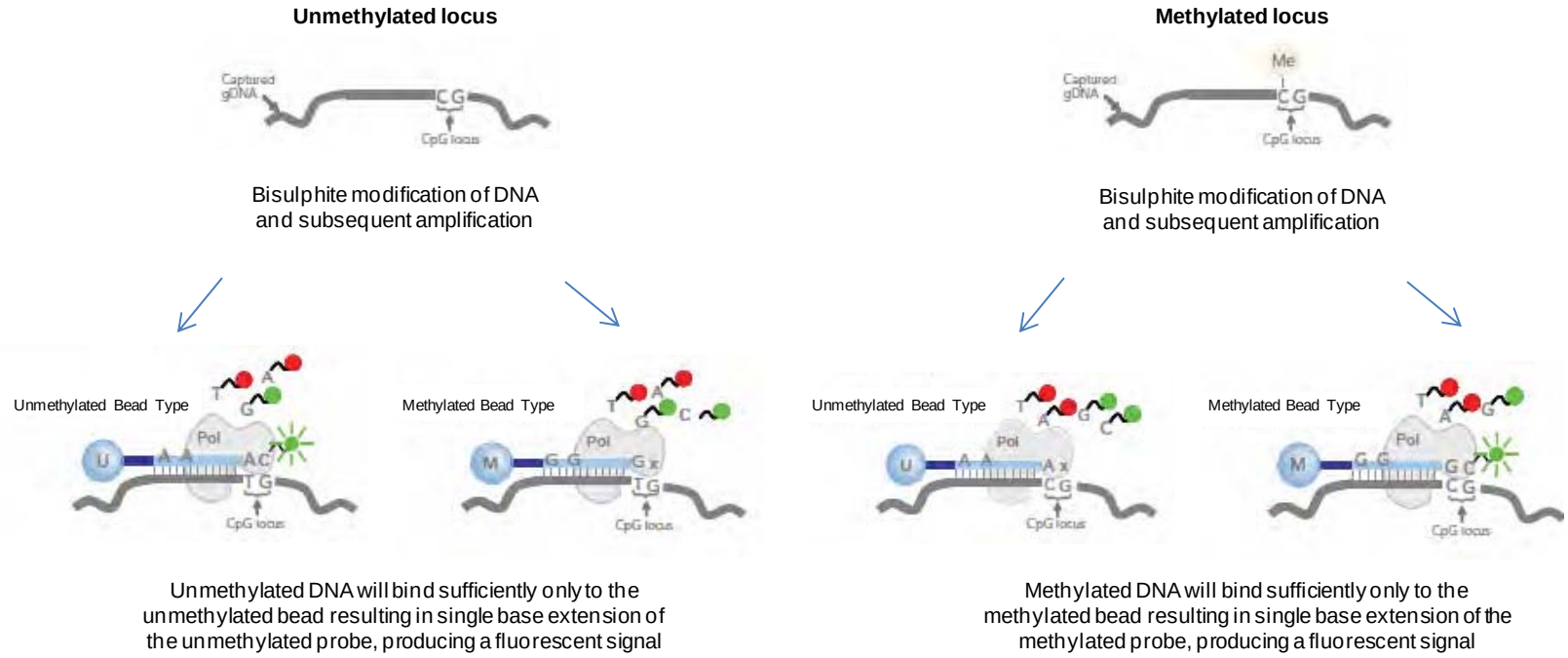


Figure 1.10_Schematic of the Illumina Infinium Methylation Assay

This figure shows a schematic of the Illumina Infinium methylation assay. Binding of unmethylated and methylated alleles to the two different bead types for each locus on the array are shown. This figure has been adapted from the Illumina website (www.illumina.com)

coding regions, CpG island shore regions, miRNA promoter regions, and CpG sites outside of CpG islands. The GoldenGate technique has been used to successfully identify differing pathways of cancer development in invasive and non-invasive urothelial cancer (Wolff *et al*¸ 2010), methylation hypermethylation profiles associating with *IDH* mutation in gliomas (Christensen *et al*, 2011) and a potential CIMP phenotype in renal cell carcinoma (McRonald *et al*¸ 2009). The Illumina Infinium HumanMethyation27 array, providing more coverage than the GoldenGate assay and a more global genomic profile, has been used to identify a CIMP phenotype in gliomas (Noushmehr *et al*, 2010) and lymphoma subtype specific DNA hypermethylation events and an enrichment for hypermethylation of polycomb group targets in lymphomas (Martín-Subero *et al*, 2009a), among many other studies. The Illumina Infinium HumanMethylation450 array has only been available for a short time and currently the only published study using this technique is an array validation study in a colorectal cancer cell line (Sandoval *et al*, 2011).

1.3.8 DNA methylation analysis methods and 5-hydroxymethylcytosine

The recent discovery of 5-hydroxymethycytosine (5hmC) within the human genome has had a big effect on this field and the perceived effectiveness of the existing techniques used to investigate methylation. Of the methods discussed for DNA methylation analysis, only the affinity enrichment based approaches (MeDIP and MIRA) differentiate between 5mC and 5hmC, binding only to 5mC (Jin *et al*, 2010). Bisulphite modification approaches are currently inadaptable due to sodium bisulphite converting both 5mC and 5hmC to uracil (Huang *et al*, 2010). The presence of 5mC and 5hmC is therefore undetermined by the majority of methods currently used to assess the methylation status of DNA. Companies are now producing a number of products allowing for the specific investigation of just 5mC or just 5hmC and these will undoubtedly prove useful in the future.

1.4 BREAST CANCER

Breast cancer is the most commonly diagnosed cancer in women both worldwide and in the UK. During 2008, 1,383,500 women were diagnosed globally (Jemal *et al*, 2011) of which, 47,693 were diagnosed in UK women with an additional 341 cases in UK men (CRUK cancer statistics). Mortality rates associated with breast cancer are the highest for any type of cancer in women worldwide and second highest in the UK, accounting for 14% of the total worldwide cancer deaths and 6.8% of the total UK cancer deaths (Jemal *et al*, 2011; CRUK cancer statistics). Higher worldwide breast cancer mortality rates are attributable to the large discrepancies between developing and developed countries. Whilst economically developed areas such as Western Europe, Australia/New Zealand and Northern America have some of the highest incidence rates, mortality rates are on a par with less economically developed countries such as Northern and Western African regions where breast cancer incidence rates are among some of the lowest (figure 1.11). The unfortunate discrepancies between breast cancer mortality rates in both developed and undeveloped countries highlights the importance of access to quality healthcare, both in order to detect cancers earlier and treat more effectively when diagnosed. The stage at which breast cancer is diagnosed and the age of the patient at diagnosis are hugely influential factors on survival (figure 1.12). Breast cancer incidence has been increasing over the past few decades and continues to increase, however the death rates are falling, largely as a result of the increase in screening techniques and therefore earlier detection of the disease (figure 1.13).

The overall lifetime risk of developing breast cancer for women is estimated at 1 in 9 in the UK (CRUK cancer statistics). Factors associated with increased risk include a family history of breast cancer, incidence of benign breast disease, increased age (Feuer *et al*, 1993), post-

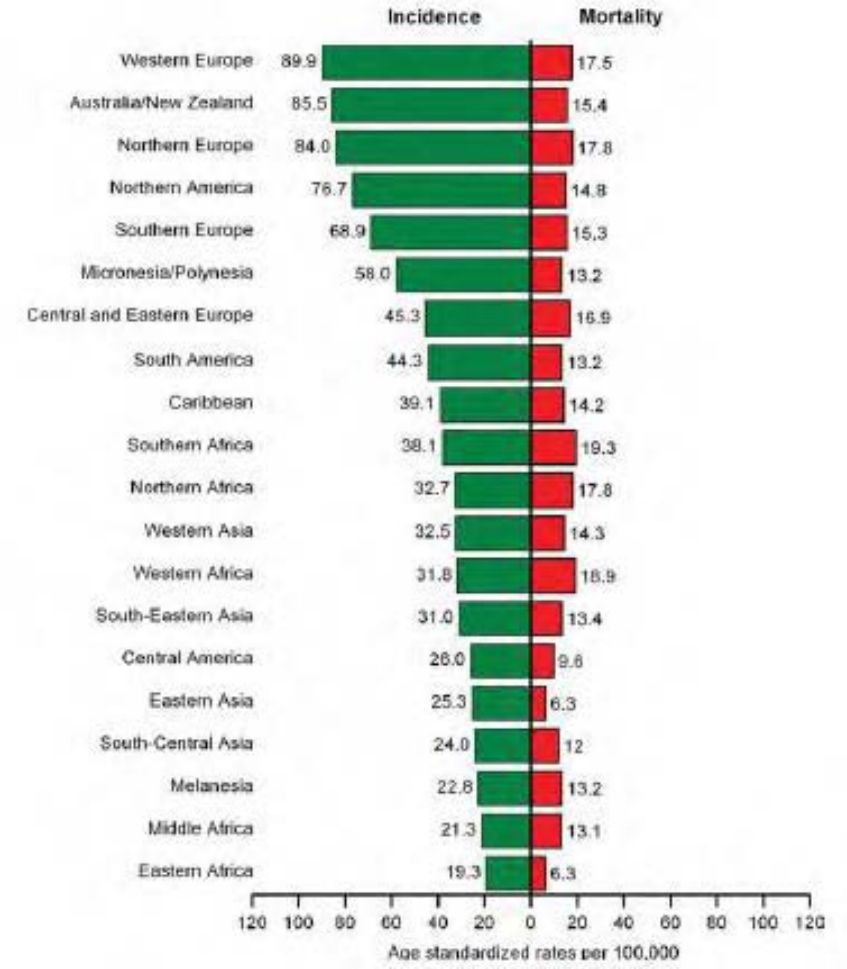
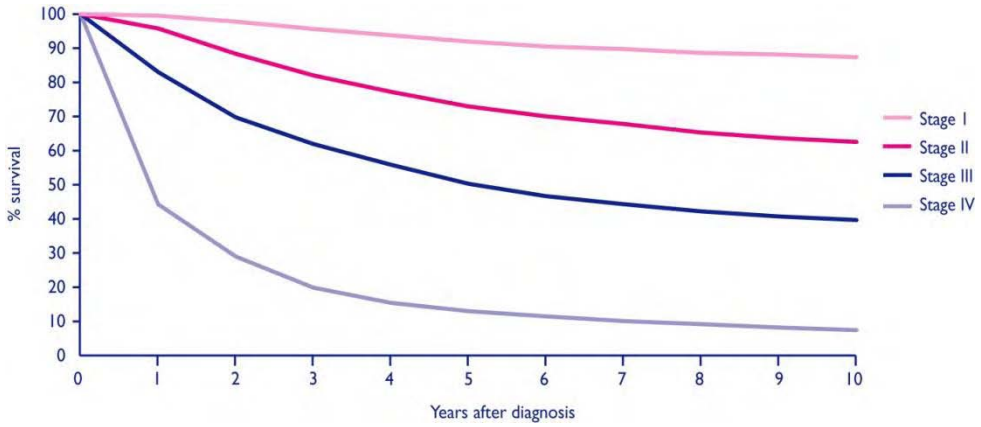


Figure 1.11_Worldwide breast cancer incidence and death rates
This graph shows worldwide age standardised breast cancer incidence and death rates split by geographical region. This graph has been taken from Jemal *et al*, 2011.

(a)



(b)

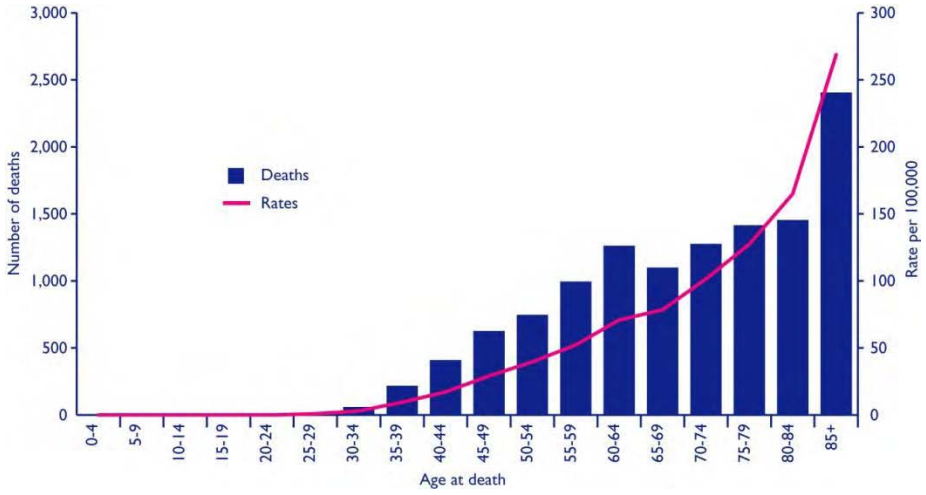


Figure 1.12_Influence of cancer stage and patient age at diagnosis on survival/mortality rates in the UK

(a) The graph shows the influence of cancer stage at diagnosis on survival rates of breast cancer patients in the UK diagnosed between 1990 and 1994. (b) The graph shows the influence of patient age at diagnosis on the mortality rates of patients from breast cancer in 2007. Both graphs are taken from the CRUK website (www.cancerresearchuk.org).

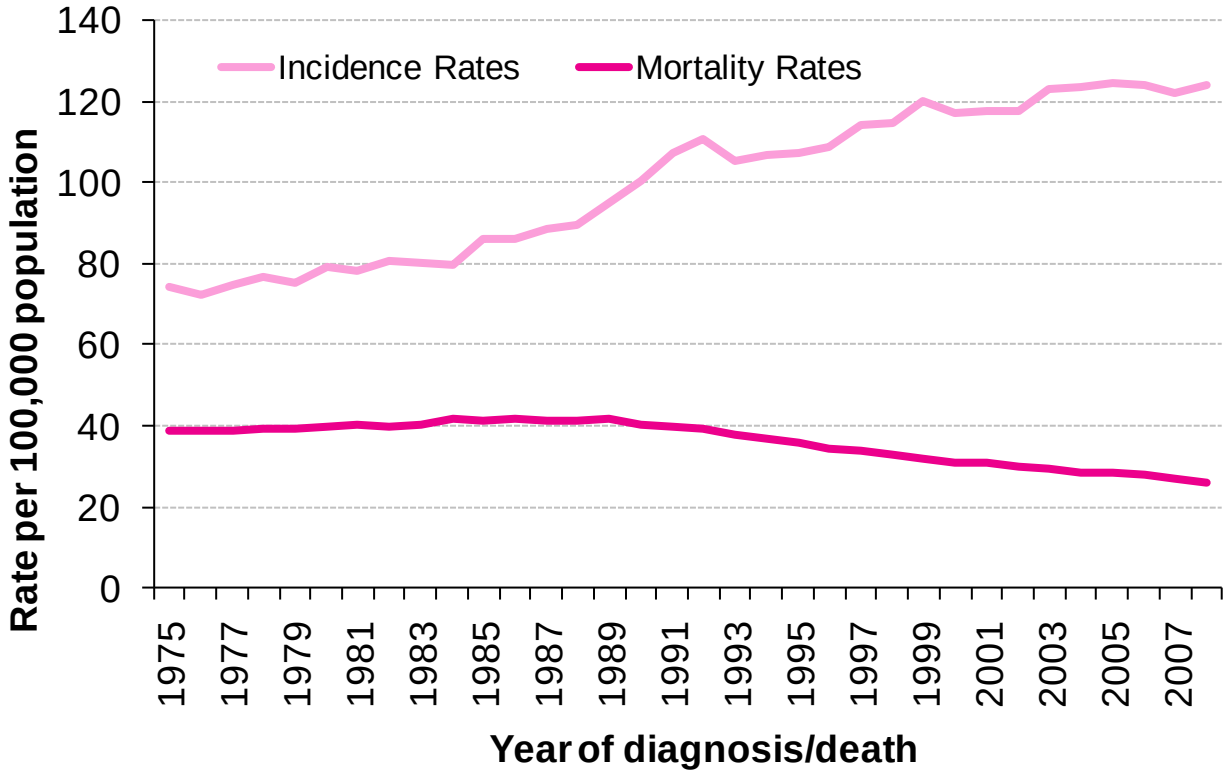


Figure 1.13_ Breast cancer European age-standardised incidence and mortality rates in the UK 1975-2008
The figure shows a graph of European age-standardised incidence and mortality rates in the UK for breast cancer between 1975 and 2008. This graph has been taken from the Cancer Research UK website (www.cancerresearchuk.org).

menopausal hormone therapy with HRT (hormone replacement therapy) (ICRF, 1997), long term use of oral contraceptives (although this is believed to be a short term effect) (Milne *et al*¸ 2005), early menarche and late menopause (Brinton *et al*, 1988), increased age at first full term pregnancy (Chie *et al*, 2000; Lambe *et al*, 1996), reduced parity (Lambe *et al*, 1996), increased breast tissue density (Chen *et al*, 2006), being overweight (Wolin *et al*, 2010) increased exposure to radiation (Helzlsouer *et al*, 1995) and increased alcohol intake (Hamajima *et al*, 2002).

1.4.1 Breast cancer subtypes and classification

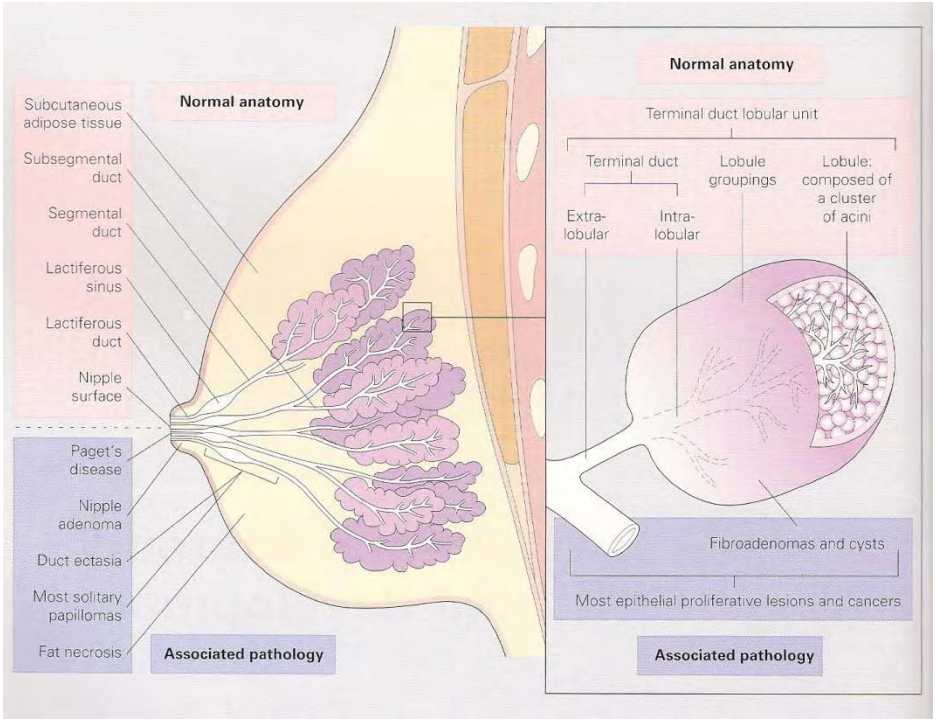
The normal breast consists largely of stromal and epithelial tissue, stromal tissue consisting of fat and fibrous tissue that supports the epithelial component of the breast which contains the functional and structural elements connected by multiple branching ducts (Hayes *et al*¸ 2000). The roles and overall contribution of these tissue types to breast anatomy vary greatly throughout a woman's lifetime. At birth, there is very little stromal component and only rudimentary branching ducts, however, through puberty and adulthood the stromal component has massively expanded and the epithelial ducts have given rise to sacular buds ready for expansion of the epithelial component during pregnancy. Post-lactation changes result in glandular atrophy and the stromal component again comprising most of the breast volume. Post-menopause, the stromal component loses its fibrous connective tissue and becomes mainly adipose tissue whilst the epithelial component begins to atrophy resulting in a decrease in lobular compartments whilst ductal regions remain (Hayes, 2000).

The majority of breast cancers form within the ductal regions of the breast although other regions can also give rise to cancers, the next most common region being the lobular component. Breast cancer disease can be classified as either in situ carcinomas or invasive

(infiltrating) carcinomas. In situ carcinomas are often considered as pre-cancerous lesions and a form of benign breast disease despite their name containing 'carcinoma'. They can be further sub-classified as ductal or lobular based on location within the breast. The rarer lobular carcinoma in situ (LCIS) demonstrates low histological variation whilst the more common ductal carcinoma in situ (DCIS) can be classified as either Comedo, Cribiform, Micropapillary, Papillary or Solid (Malhotra *et al*, 2010). These classifications are based entirely on histological findings rather than molecular markers. Invasive or infiltrating carcinomas can be split into six types; tubular, ductal lobular, invasive lobular, infiltrating ductal, mucinous (colloid) or medullary (Malhotra *et al*, 2010). Of these, infiltrating ductal is the most common; accounting for 70-80% of all invasive breast cancers, the next major subtype is invasive lobular carcinoma (Li *et al*, 2005). The work in this thesis concentrates upon the common infiltrating ductal carcinomas. Figure 1.14 illustrates the normal anatomy of a breast and the disruption of normal breast architecture during carcinoma.

Further characterisation of invasive carcinomas is based on gene expression signatures characterised during studies in the early 2000s. The epithelial component of the breast consists of basal and luminal cells. An expression study on a cohort of 86 breast samples (consisting primarily of invasive ductal carcinomas - some of which were paired before and after treatment with doxorubicin - and a mixture of lobular carcinomas, breast cancer cell lines and normal breast tissue samples) demonstrated that gene expression signatures are able to classify breast carcinomas as either luminal, basal, *ERBB2* gene (encoding HER2/neu – Human Epidermal growth factor Receptor 2) overexpressing or normal-like (Perou *et al*, 2000). Overexpression of the *ERBB2* oncogene is an important event in breast cancer and is thought to be present in approximately 30% of cases (Slamon et al, 1987). *ERBB2* overexpression forms the basis of treatment with the drug trastuzumab (Herceptin).

(a)



(b)

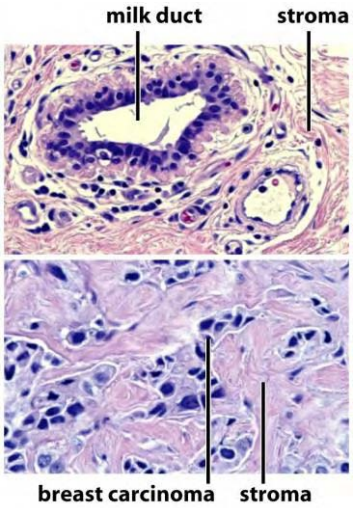


Figure 1.14_Normal and cancerous breast anatomy

(a) A simple schematic of normal breast anatomy is shown indicating stromal and epithelial components (left) and a more detailed schematic of epithelial terminal duct lobular units (right). The locations of breast diseases, both benign and malignant are also indicated. This figure has been taken from 'Atlas of Breast Cancer Second Edition' edited by Daniel F Hayes. (b) A high magnification histology section of the distinct architecture of normal breast anatomy indicating the milk duct and surrounding stromal region is shown (top). Below this, a high magnification histology section of breast carcinoma is shown, note the loss of distinctive architecture around the duct and epithelial cell invasion of the stroma. This figure has been taken from 'The Biology of Cancer' by Robert A Weinberg.

(Ménard *et al*, 2003). Luminal and basal subtypes both express genes consistent with known markers for these cell types, luminal tumours also express the oestrogen receptor (ER) whilst basal tumours do not. Normal-like cells express genes consistent with normal basal epithelial breast cells whilst *ERBB2* over-expressing tumours have low levels of ER and overexpression of *ERBB2* (Perou *et al*, 2000). Subsequent validation of this work in separate cohorts confirmed the subtypes described plus an additional splitting of the luminal subtype into luminal A and luminal B sub-classifications. Luminal A tumours have the highest expression of ER and ER related genes whilst luminal B tumours remain ER positive but have lower expression of ER and ER related genes (Sørlie *et* al, 2001; Sørie *et al*, 2003). Overall survival analysis of subtype specific cancers has shown that basal-like and *ERBB2* overexpressing cancers have the worst overall and metastases free survival of all subtypes whilst luminal A tumours have the best outcomes (Sørlie *et al*, 2001). Recently, a further subtype, claudin-low, has been described which is characterised by low expression of claudin 3, claudin 4, claudin 7 and E-cadherin (Herschkowiitz *et al*, 2007; Prat *et al*, 2010). Whilst the previous classifications of breast cancer demonstrated markers of epithelial cells, the claudin-low subtype demonstrates markers of a class of cells called TICs (tumour initiation cells). The detection of TICs that are unlike basal or luminal epithelial cells adds to the theory that breast cancers can arise as a result of cancer stem cells. Despite these studies the Claudin-low subtype classification is yet to be widely accepted.

Other expression studies on breast cancer have enabled the development of microarray tests to aid in identifying patients who will benefit most from particular treatments. For example, a study in 2002 on 98 breast cancers identified a 70 gene signature capable of predicting poor prognosis and therefore enabling prediction of patients who would benefit from more severe treatments (van t'Veer *et al*, 2002). The study has led to the development of MammaPrint, a

microarray test developed by Agendia that is now approved by the FDA in the USA for the detection of breast cancers that are most likely to metastasise and therefore would benefit from more severe treatments (Glas *et al*, 2006). Other, similar studies, have developed a gene expression signature associated with survival (van de Vijver *et al*, 2002) and a reduced gene signature of 50 genes from the original breast cancer subtyping studies (Perou *et al*, 2000; Sørlie *et* al, 2001; Sørie *et al*, 2003) that can predict clinical outcome (Parker *et al*, 2009). A second commercially developed microarray test, Onco*type* DX, has been produced that uses a 21 gene expression panel to determine the likelihood of relapse (Cobleigh *et al*, 2005). These studies, whilst extremely beneficial to patients, rely on mRNA expression studies which present more inherent problems than DNA based studies and therefore identification of further predictive markers would be useful.

1.4.2 Breast cancer genetics

The genetics of breast cancer patients, like the genetics of most forms of cancer, can be split into two distinct groups. By far the largest group consists of individuals who acquire somatic mutation and genetic/epigenetic alterations throughout their lifetimes, possibly affected by numerous environmental influences, which combine to initiate and enable the tumourigenic process. A smaller group consists of individuals who inherently carry a pre-existing genetic fault that increases their risk of developing breast cancer; these individuals are usually found as members of breast cancer predisposed families.

1.4.2.1 Familial/Inherited Predisposition to Breast cancer

Typically, breast cancer families exhibit autosomal dominant inheritance of a predisposition to breast and ovarian cancers with the affected allele demonstrating incomplete penetrance. Familial breast cancers often occur at an earlier age of onset than sporadic cancers, occurring

primarily before the age of 50, and are more commonly bilateral than sporadic cancers. Breast cancer families can also sometimes have affected males. These families, although unfortunate, have allowed for the identification of important genes associated with breast cancer susceptibility.

The most commonly mutated genes conferring an inherited risk of breast cancer are *BRCA1* (*breast cancer 1*) and *BRCA2* (*breast cancer 2*), both of which were identified through linkage analysis of breast cancer families. The first breast cancer susceptibility gene, *BRCA1*, was identified at 17q21 through linkage analysis of 23 affected families (Hall *et al*, 1990; Miki *et al*, 1994) and *BRCA2* was identified at 13q12.3 through linkage analysis and sequencing of single individuals from 46 affected families (Wooster *et al*, 1995). *BRCA1* is a 81 kb gene with 23 exons encoding a 207 kDa protein and *BRCA2* is a 84 kb gene with 27 exons encoding a 390 kDa protein. Although both genes do not have similar functions they have been shown to be involved in similar pathways. In normal cells, BRCA1 protein is able to act as a scaffold protein for p53 phosphorylation by ATM. BRCA1 deficient cells are therefore lacking active p53, resulting in a block on p53 mediated induction of p21 and therefore p21 induced G1/S arrest. BRCA1 deficient cells are therefore unable to activate the G1/S checkpoint, allowing cell cycle progression in the presence of DNA damage (Fabbro *et al*, 2004). The S phase check point is also affected by BRCA1 deficiency, due to loss of regulation of Chk1 kinase activity and specific Ser1387 phosphorylation of BRCA1 by ATM following ionizing radiation (Xu *et al*, 2001). In a similar fashion to the S phase checkpoint, BRCA1 regulates Chk1 activity at the G2/M checkpoint and is itself activated by phosphorylation at Ser1423 by ATM (Xu *et al*, 2001). A role for BRCA1 has also been implicated in the non-homologous end-joining (NHEJ) and homologous recombination (HR) methods of DNA repair (Fu *et al*¸2003; Greenberg *et al*, 2006). Whilst the role of BRCA1 in

NHEJ is somewhat controversial the role of BRCA1 in HR has been well defined. During the repair process, BRCA1 interacts directly with the MRN (Mre11/Rad50/NBS1) complex, which, along with ATM, is essential for converting double stranded broken ends into single stranded DNA, the first step of HR (Greenberg *et al*, 2006). In addition, BRCA1 has been shown to interact with PALB2 protein (Sy *et al*¸ 2009) and BRCA2 protein (Zhang *et al*, 2009a, b) at DNA damage sites, both PALB2 and BRCA2 are functional partners of Rad51, a member of the MRN (MRE11, RAD50 and NBS1) complex, and have been shown to aid the formation of single stranded DNA from double strand breaks. BRCA1 deficiency reduces the stability of PALB2 and BRCA2 at DNA damage sites, affecting Rad51 isolation at DNA lesions, ultimately impairing HR repair (Zhang *et al*, 2009a, b).

Due to their roles in DNA repair, both *BRCA1* and *BRCA2* deficient cells are sensitive to PARP (poly (ADP-ribose) polymerase) inhibitors. PARP is involved in the repair of single-strand breaks by inducing double strand breaks at the point of damage which are then repaired via HR. Inhibition of PARP results in an accumulation of single strand breaks, which, when combined with an already DNA repair compromised state in *BRCA1* and *BRCA2* deficient cells, results in cell death due the sheer amount of unrepaired DNA damaged sites. PARP inhibitors are routinely used in treatment of *BRCA1* and *BRCA2* mutated breast cancers (Bryant *et al*, 2005; Farmer *et al*, 2005).

Other germline mutated genes in breast cancer are *ATM* (*ataxia telangiectasia mutated*), *PALB2* (*partner and localizer of BRCA2*), *CDH1* (*cadherin 1, type 1, E-cadherin*), *TP53* (*tumour protein p53*), *BRIP1* (*BRCA1 interacting protein C-terminal helicase 1*), *CHEK2* (*checkpoint kinase 2 homolog*), *RAD54L* (*RAD54 like (S.cerevisiae)*) and *RAD51A* (*RAD51 homolog (RecA homolog, E. coli) (S. cerevisiae*)).

PALB2 is a binding partner of *BRCA2* and as such was assessed for mutations within *BRCA1/2* mutation negative breast cancer families. Truncating mutations were identified in 1% (10/923) of assessed families, conferring a 2-3 fold higher risk of breast cancer in *PALB2* mutated individuals (Rahman *et al*, 2007). *PALB2* frameshift mutations have also been identified in Finnish breast cancer families (Erkko *et al*, 2007). Germline mutations of *CDH1* have been identified in a gastric cancer and early-onset breast cancer family (Guilford *et al*, 1999) in addition to a lobular breast cancer patient with an affected mother but otherwise unaffected family (Masciari *et al*, 2007). Additionally, somatic mutations in lobular breast cancer patients have also been observed in *CDH1* (Berx *et al*, 1997). A further BRCA1 interacting partner, *BRIP1* (*BRCA1 interacting protein 1),* has been shown to be mutated in *BRCA1/2* unmutated breast cancer families (Cantor *et al*. 2004; Seal *et al*, 2006). *CHEK2* mutations were demonstrated to cause an alternative form of LiFraumeni syndrome (usually caused by inherited mutation of the *TP53* gene), of which breast cancer is a common feature (Bell *et al*, 1999). Subsequent analysis of *CHEK2* in *BRCA1/2* unmutated families identified truncating *CHEK2* mutations in 5.1% of affected individuals compared to 1.1% of unaffected controls (Meijers-Heijboer *et al*, 2002). The increased risk of breast cancer in specific *CHEK2* mutation carriers was estimated to be 2-fold greater than in *CHEK2* unmutated individuals (The CHEK2 Breast Cancer Case-Control Consortium, 2004).

In some cases, very infrequent mutation has been identified, for example, only one *RAD54L* mutation has been identified in 1 out of 95 analysed patients (Matsuda *et al*, 1999) and one specific *RAD51A* mutation has been documented in two unrelated patients with familial breast cancer (both patients exhibited bilateral breast cancer) (Kato *et al*, 2000). Interestingly, a SNP (singe nucleotide polymorphism) in *RAD51A* has also been implicated in increased breast cancer susceptibility. The CC homozygous allele of the 135G-C SNP in the *RAD51A* 5'UTR

has been shown to increase breast cancer susceptibility in BRCA2 mutation carriers (Wang *et al*, 2001; Levy-Lahad *et a*, 2001; Antoniou *et al*, 2007).

Additionally, breast cancer susceptibility can occur as a component of more generic cancer susceptibility syndromes such as Li-Fraumeni syndrome, Ataxia Telangiectasia, Peutz-Jehgers syndrome and Cowden syndrome. Li-Fraumeni syndrome (MIM ID 151623) is caused by germline mutations of the p53 gene, *TP53* (Malkin *et al*, 1990). The syndrome exhibits autosomal dominant inheritance and predisposes affected individuals to multiple, early-onset tumours. The most common cancers forming part of Li-Fraumeni syndrome are soft tissue sarcomas and osteosarcomas however breast cancer, brain tumours, adrenocortical carcinomas and leukaemias are also common. The estimated risk of developing breast cancer by age 45 as a Li Fraumeni affected individual is 50-60% (Chompret *et al*, 2000). Cowden syndrome (MIM ID 158350) is caused by mutation of the *PTEN* tumour suppressor in about 80% of cases (Liaw *et al*, 1997). Phenotypes include hamartomatous polyps of the gastrointestinal tract, mucocutaneous lesions and increased risk of developing cancers, the most common being gastrointestinal, breast and thyroid. Benign breast disease, including ductal hyperplasia, intraductal papillomatosis, adenosis, lobular atrophy, fibroadenomas and fibrocystic change are common features of Cowden syndrome in women and many go on to develop malignant tumours (Schrager *et al*, 1998). The estimated risk of Cowden syndrome patients developing breast cancer by the age of 70 years is 30-50% (Guénard *et al*¸ 2007). Peutz-Jeghers syndrome (MIM ID 175200) is an autosomal dominantly inherited disorder caused by mutations of the *STK11* gene (Jenne *et al*, 1998). Phenotypic manifestations include melanocytic macules of the lips, buccal mucosa and digits, gastrointestinal hamartomatous polyps and various cancers including breast. The risk of Peutz-Jeghers syndrome patients developing breast cancer by the age of 70 has been estimated at 45%

(Hearle *et al*, 2006). Ataxia telangiectasia is an autosomal recessive disorder caused by mutations of *ATM* gene (Savitsky *et al*, 1995). The disease is characterised by increased sensitivity to ionizing radiation, progressive cerebellar ataxia, frequent infections and an increased susceptibility to cancer. Leukaemias are common in childhood ataxia telangiectasia patients and epithelial cancers, including breast cancer, are common in adulthood. ATM is required for the phosphorylation of CTIP, a BRCA1 associated protein, which, when phosphorylated, dissociates from BRCA1, allowing BRCA1 to repress GADD45, a decrease in ATM levels can therefore affect the amount of active BRCA1 (Li *et al*, 2000). Despite this, screening of breast cancer patients for *ATM* truncating mutations has not detected inactivating heterozygous *ATM* mutation as a common event. However, missense mutations have been identified that have been shown to affect biological activity of ATM (Scott *et al*, 2002) and that are associated with ataxia telangiectasia. The relative risk of breast cancer associated with these *ATM* mutations has been estimated at 2.37 (Ahmed *et al*, 2006). Table 1.1 shows breast cancer susceptibility genes.

1.4.2.2 Somatic Genetic Alterations associated with Breast cancer

Despite the number of genes where germline mutations have been described in breast cancer, the overwhelming majority of sporadic breast cancers do not have these germline events and instead genetic and/or epigenetic events are tumour acquired. The Catalogue of Somatic Mutations (COSMIC) database (http://www.sanger.ac.uk/genetics/CGP/cosmic/), which contains sequencing results for specifically selected genes, lists 1346 genes that have been shown to be mutated out of 25578 breast cancer samples. However, it is the frequency of mutation of each gene that is more relevant than the total number of genes, for instance only the top three genes, *PIK3CA* (*phosphoinositide-3-kinase, catalytic, alpha polypeptide*), *TP53* and *CDH1*, demonstrated mutations in greater than 8% of samples (26%, 23% and 22%

PENETRANCE	GENE/LOCUS	RELATIVE RISK OF BREAST CANCER	CARRIER FREQUENCY	BREAST CANCER SUBTYPE	OTHER CANCERS IN MONOALLELIC CARRIERS	SYNDROME IN MONOALLELIC CARRIERS	SYNDROME IN BIALLELIC CARRIERS
High	BRCA1	>10	0.10%	Basal (ER-negative)	Ovarian	Breast and ovarian cancer predisposition	-
	BRCA2	>10	0.10%	-	Ovaran, prostate	Breast and ovarian cancer predisposition	Fanconi anaemia D1
	TP53	>10	<0.1%	-	Sarcomas, adrenal, brain	Li Fraumeni syndrome	-
Uncertain	PTEN	2-10	<0.1%	-	Thyroid, endometrium	Cowden syndrome	
	STK11	2-10	<0.1%	-	Gastric	Peutz-Jeghers syndrome	-
	CDH1	2-10	<0.1%	lobular	Gastric	-	-
Intermediate	ATM	2-3	0.40%	-	-	-	Ataxia telangiectasia
	CHEK2	2-3	0.40%	-	-	-	-
	BRIP1	2-3	0.10%	-	-	-	Fanconi anaemia J
	PALB2	2-4	<0.1%	-	-	-	Fanconi anaemia N
Low	10q26; 16q12; 2q35; 8q24; 5p12	1.08-1.26	24-50%	ER-positive	-	-	-
	11p15; 5q11	1.07-1.13	28-30%	-	-	-	-
	2q33	1.13	0.87	-	-	-	-

Table 1.1_Breast cancer susceptibility genes

This table lists genes known to provide an inherited predisposition to breast cancer. Genes are split into allelic penetrance of high, uncertain, intermediate and low. The estimated relative risk of developing breast cancer is given for each gene as well as the carrier frequency. If mutations are associated with a particular breast cancer subtype this is also provided. Inherited cancer syndromes caused by monoallelic inheritance are given alongside other cancers that these syndromes confer susceptibility to. Syndromes caused by biallelic inheritance are also shown. This table has been modified from Turnbull *et al*, 2008.

respectively). *BRCA1* and *BRCA2* are both mutated in 2%, whilst the twentieth gene, *HRAS* (*v-Ha-ras Harvey rat sarcoma viral oncogene homolog*), is only mutated in 1% of samples analysed. The top twenty most significantly mutated genes are shown in figure 1.15. They include well known tumour suppressor genes such as *TP53*, *RB1*, *PTEN*, *CDKN2A*, *APC*, *ATM*, *BRCA1* and *BRCA2* all of which have already been mentioned, and all except *CDKN2A* and *APC* are involved in inherited breast cancer susceptibility syndromes (section 1.4.2.1). The remaining genes in this list include well known oncogenes such as *KRAS* (*v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog*), *BRAF* (*v-raf murine sarcoma viral oncogene homolog B1*) and *NRAS* (*neuroblastoma RAS viral (v-ras) oncogene homolog*) (COSMIC - http://www.sanger.ac.uk/genetics/CGP/cosmic/).

Recent genome wide sequencing studies have identified an average of 80-90 genes mutated per genome in breast and colorectal cancers (Sjöblom *et al*, 2006; Wood *et al*, 2007). Also, sequencing of a primary basal-like breast cancer, associated brain metastasis sample and xenograft model produced from the primary cancer identified 50 novel somatic point mutations and small insertions and deletions, 28 large deletions, 6 inversions and 7 translocations within this single patient (Ding *et al*, 2010). Although many of the mutations observed in the localised primary breast tumour were maintained throughout progression to the distant brain metastasis, many were also lost, providing a smaller range of mutations, suggesting only cells carrying a subset of the original mutations metastasize (Ding *et al*, 2010). A similar study in a lobular breast cancer sample and associated metastasis sample produced similar results (Shah *et al*, 2009). The spectrum of mutations in breast cancer appears to be slightly different to prostate and colorectal cancers, which has been suggested to indicate a different DNA repair mechanism in normal breast tissue (Jones *et al*, 2008).

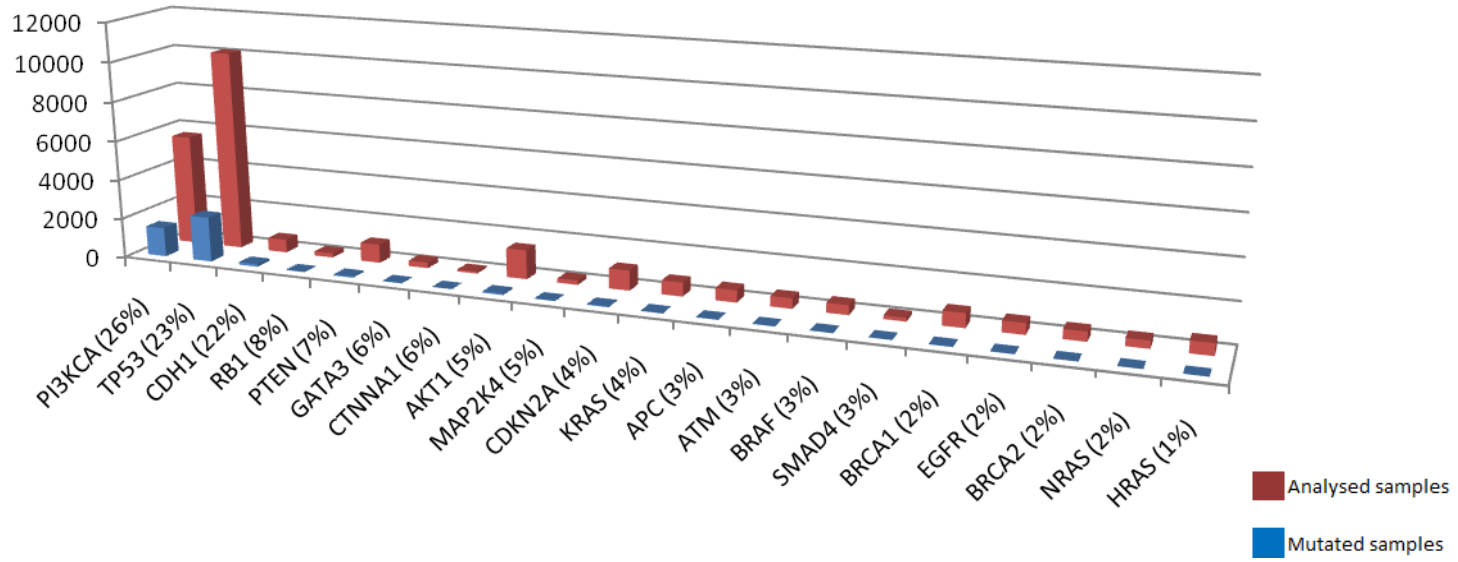


Figure 1.15_ The top 20 most significant genes somatically mutated in breast cancer
This graph illustrates the number of samples analysed (red bars at the back) and the number of samples found to be mutated (blue bars at the front) for each of the top 20 most significantly somatically mutated genes in breast cancer. This graph has been re-drawn from COSMIC.

The COSMIC data and results from recent large sequencing studies illustrate the high frequency of genetic alteration in breast cancer genomes that contribute to tumourigenesis in sporadic breast cancers. In addition to this, multiple chromosomal rearrangements occur in breast cancer genomes, as exemplified by the recent study by Stephens *et al* (2009). Nevertheless, a large number of these mutations may turn out to be carrier mutations, unimportant in regards to tumourigenesis, and numerous tumours still lack mutation or chromosomal rearrangements involving major tumour suppressor genes. This seems indicative of other forms of genomic and epigenetic alterations being heavily involved in breast cancer; in the case of epigenetics this has been shown to be true.

1.4.3 Breast cancer epigenetics and epigenomics

Although cancer genomes are generally demethylated, hypomethylation events are generally associated with repeat regions rather than specific genes (Ehrich, 2009). Despite this, a number of recent studies have detected gene specific hypomethylation in breast cancers. *IL-10* (*Interleukin 10*) methylation has been demonstrated to be present in 59% of normals and 30% of tumours; this decreased methylation in tumours was also shown to correlate with higher expression levels (Son *et al*, 2010) and the *MDR1* (*multidrug reistance 1*) gene promoter has been shown to be hypomethylated in 47% of tumours and 44% of associated patient serum (Sharma *et al*, 2010). Additional genes include *FEN1* (flap *structure-specific endonuclease 1*), *NAT1* (*N-acetyltransferase1*) and *CDH3* (*cadherin 3*) (Singh *et al*, 2008; Kim *et al*, 2008; Paredes *et al*, 2005). A recent study has illustrated one of the potential implications of genome wide hypomethylation whereby the *TLR9* (*Tol-like receptor 9*) gene binds to unmethylated DNA and so becomes increasingly expressed in cells showing greater overall hypomethylation levels. *TLR9* gene expression was greatest in hormone receptor negative cancers suggesting that these cancers may have the greatest level of hypomethylation. It was

also demonstrated that *TLR9* expression enhanced the migratory ability of cells (Berger *et al*, 2010).

Nonetheless, the number of genes demonstrated to be hypermethylated in breast cancer is substantially larger and constantly increasing. Use of the search term 'breast cancer methylation' in PubMed produces over 1500 hits and searching the PubMeth website (http://www.pubmeth.org/ - a slightly outdated website aiming to catalogue published methylated genes in various cancers) for genes methylated in breast cancer gives what is likely to be a large underestimate of 100 genes. Combined with the plethora of recent genome wide studies, which often publish lists of tens or hundreds of candidate methylated genes at a time, it is not feasible to discuss or even provide a table of all methylated genes in breast cancer. However, classical tumour suppressor genes and commonly methylated genes such as *CCND2* (*cyclin D2*), *CDKN2A* (*cyclin dependent kinase inhibitor 2A*), *BRCA1* (*breast cancer 1*), *APC* (*adenomatous polyposis coli*), *RASSF1A* (*Ras association (RalGDS/AF-6) domain family member 1 isoform A*) and *RARβ* (*retinoic acid receptor β*) have all been shown to be hypermethylated in breast cancer. *CDKN2A* is a cell cycle regulator and potent tumour suppressor, causing familial melanoma when mutated in the germline (Hussussian *et al*, 1994). As previously mentioned, methylation of *CDKN2A* was one of the first tumour suppressor genes to be shown to be inactivated by hypermethylation in numerous cancers, of which breast cancer was one (Herman *et al*, 1995). Since there is no evidence of *CDKN2A* methylation in DCIS samples, it is likely to be an event occurring during the progression of invasive ductal carcinomas (Muggerud *et al*, 2010). Another cell cycle regulator, *CCND2*, is frequently methylated in breast cancer; methylation has been observed at a frequency of 46% in breast cancers and correlates with low expression at the mRNA and protein levels (Evron *et al*, 2001). In addition, frequent methylation has also been observed in DCIS (Evron *et al*,

2001), suggesting *CCND2* methylation is an early event that could contribute towards the initiation of breast carcinogenesis. The role of *BRCA1* in breast cancer has already been discussed, but it is worth noting that in sporadic cases, *BRCA1* inactivation is sometimes achieved via promoter hypermethylation, however this a rare event, occurring in approximately 13% of cases, although there seems to be some bias for *BRCA1* methylation occurring in medullary breast carcinomas (Dobrovic *et al*, 1997; Esteller *et al*, 2000b). *APC* is a tumour suppressor gene involved in Wnt signalling inhibition which causes familial adenomatous polyposis (FAP) when mutated (Groden *et al*, 1991), and has been shown to be tumour-specifically methylated in 36% of breast cancers (Jin *et al*, 2001). *RASSF1A*¸which will be discussed in more detail in section 1.5, has been shown to be methylated in 62% of breast cancers (Dammann *et al*¸ 2001). *RARβ* is a commonly methylated gene in many cancers, including invasive and in situ breast carcinomas (Fackler *et al*¸ 2003). The most heavily analysed genes for methylation in breast cancer (according to PubMeth) are shown in figure 1.16.

In breast cancer, numerous recent studies have started to investigate the importance of the epigenome and to give insight into the role of global methylation differences in association with breast cancer subtypes and clinical features. A study in 2009 used a microarray approach to identify methylation events in 241 breast tumours, identifying a set of 14 methylated genes that associated with a shortened time before the occurrence of distant metastases (Hartmann *et al*, 2009). A set of 4 of these methylated genes showed promise as biomarkers of prognosis and identification of the patients that could benefit from most severe treatments (Hartmann *et al*, 2009). In a recent study by Bediaga *et al* (2010) using the Illumina GoldenGate Methylation Cancer Panel I (Illumina, San Diego, CA, USA) array, genes were identified where methylation status associated with the basal-like subtype, including *HOXA9* (*homeobox*

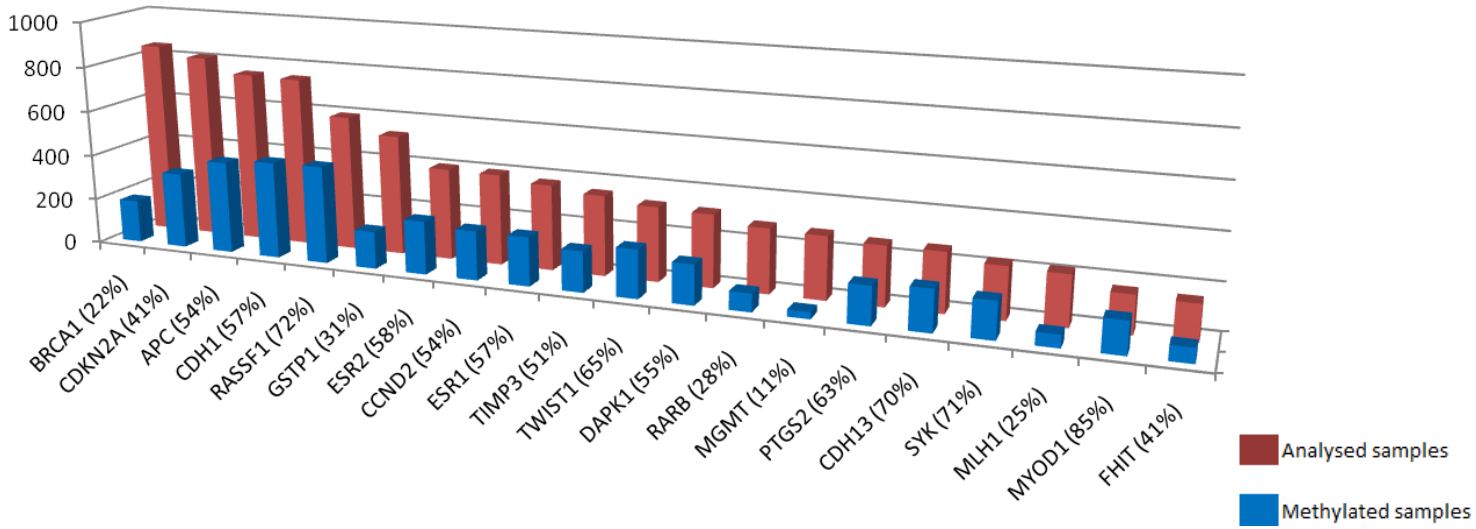


Figure 1.16_Methylation of the top 20 most analysed genes in breast cancer

This graph shows the top 20 most analysed genes in breast cancer for DNA methylation. The data for this graph was obtained from PubMeth (www.pubmeth.org). The number of samples analysed is represented by red bars at the back and the number of samples where methylation was detected is indicated by blue bars at the front. The frequency of methylation for each gene is shown as a percentage. These percentages are given by PubMeth as a average of that observed in numerous studies therefore the percentages given may differ from that quoted elsewhere in the text.

A9) and *SOX1* (*SRY (sex determining region Y)-box 1*), and genes where methylation status associated with the luminal B subtype, including *RASSF1A*, *FGF2* (*fibroblast growth factor 2 (basic)*), and *NPY* (*neuropeptide Y*) (Bediaga *et al*, 2010). The authors suggest that novel methylation profiles of relatively small numbers of genes would be of use in classifying tumours within a particular subtype. Another study using the Illumina GoldenGate method identified individual genes associated with particular clinical features and that the DNA methylation profiles were influenced by tumour size and alcohol and folate intake (Christensen *et al*, 2010). Further evidence for subtype specific methylation alteration was demonstrated in Holm *et al* (2010), where the authors again assessed quantitative methylation levels using the Illumina GoldenGate method in 189 breast tumours. In this panel of cancer related genes, approximately 31% of all analysed CpGs (approximately two per gene) were affected by tumour specific changes. Methylation based cluster groups organised samples primarily based on ER status, with subgroups sitting together within these ER based clusters. The greatest increases in methylation were observed in luminal B subtype tumours and the lowest in basal-like tumours. The authors were also able to demonstrate correlation with increased gene methylation and decreased mRNA expression (Holm *et al*, 2010). A recent genome-wide familial study of 33 *BRCA1*/*BRCA2* or *BRCAn* breast cancer cases clustered samples based on mutation status, with *BRCA1* mutated samples clustering together and *BRCA2*/*BRCAn* mutated samples clustering together. *BRCA1* mutated cancers were significantly more methylated that *BRCA2* or *BRCAn* mutated samples. In these familial samples, mutation status was the overriding factor associating with differing methylation levels (Flanagan *et al*, 2010). These results suggest that *BRCA1* mutated samples may be more prone to developing changes in methylation although mechanisms behind this are not yet clear. The field of epigenomics is expanding at a rapid rate with more and more platforms

available for the investigation of the global levels of CpG island and genome methylation which will aid in further elucidation of methylation changes in breast cancer.

1.5 THE RASSF FAMILY

The RASSF gene family currently consists of ten members (*RASSF1-10*) and represents a family of genes with tumour suppressor properties that are often downregulated in cancers as a result of hypermethylation. They encode proteins of the RASSF protein family which is characterised by the presence of a Ras-association (RalGDS/AF-6) domain (RA domain), although it should be noted that not all RA-domain proteins are RASSF family members. The family was originally defined as having both an RA domain and a SARAH (Sav-RASSF-Hpo) domain. However, a new subsection of the RASSF family, described as an evolutionary distinct group of RASSF proteins, have recently been described that lack the SARAH domain (Sherwood *et al*, 2008). The family is now split into two groups; the classical RASSFs (RASSF1-6) and the N-terminal RASSFs (RASSF7-10), defined by the position of the RA-domain and the presence/absence of a SARAH domain. Classical RASSFs contain an RA-domain and a SARAH domain at their C-terminus. RASSF1 and RASSF5 also contain a C1 (phorbol esters/diacylglycerol binding) domain (Richter *et al*, 2009). The N-terminal RASSFs differ from the classical RASSFs by the lack of a SARAH domain and positioning of the RA-domain at the N-terminus (Sherwood *et al*¸ 2008) (figure 1.17). RA domains are characteristic of Ras-effectors and Ras-related-GTPases (Ponting *et al*¸1996). SARAH domains are involved in homo/heterodimerisation (Scheel *et al*, 2003). No enzymatic activity is known for the RASSFs, they are considered to most likely act as scaffolds for larger multi-protein complexes (Pfeifer *et al*, 2010).

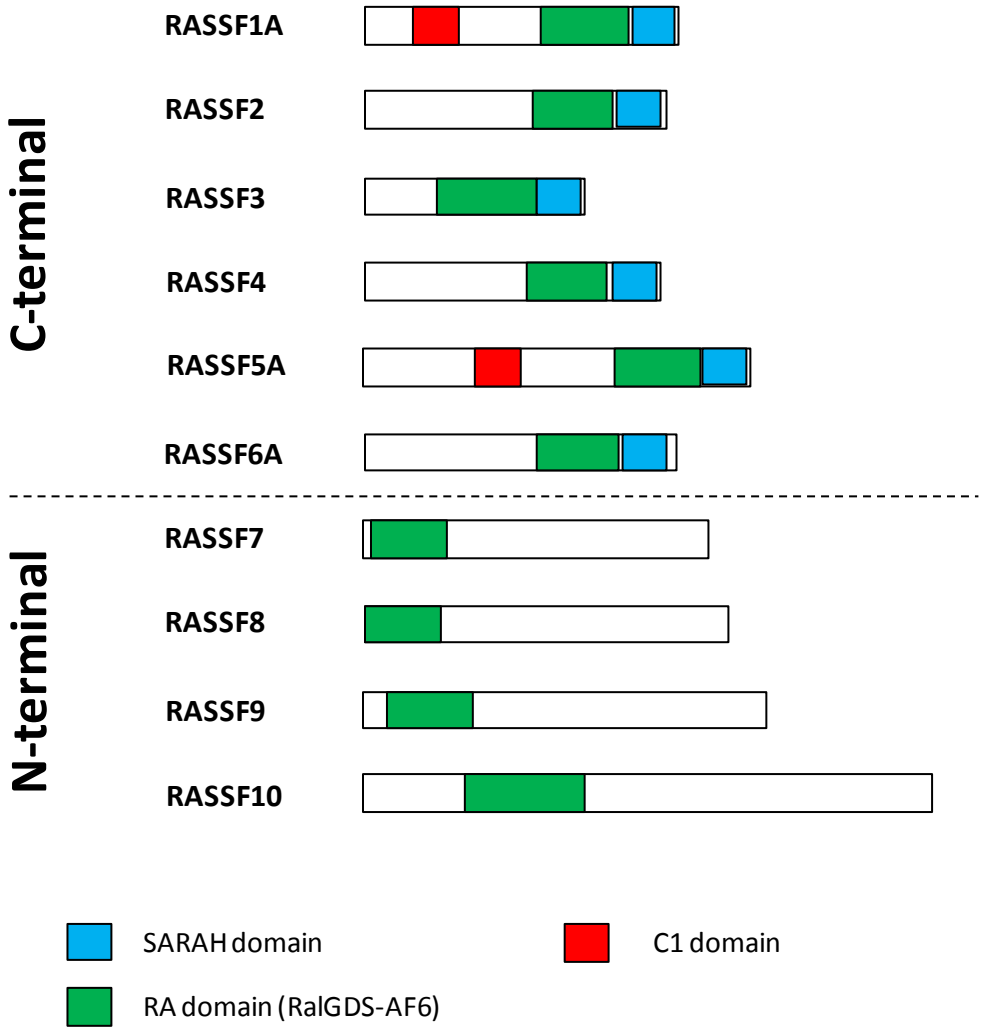


Figure 1.17 Major isoforms of RASSF family members
This figure shows the relative positions and sizes of functional domains within the major isoforms of RASSF proteins 1-10. This figure was redrawn and modified from Richter *et al*, 2009.

1.5.1 Mechanisms of tumour suppressor effects

Multiple functions of the RASSFs contribute to their tumour suppressor protective effects. The presence of an RA-domain means that RASSFs have the potential to act as RAS effectors, although functioning as such has not been confirmed for all members of the family. RAS genes are a family of oncogenes that when active in their GTP-bound form are responsible for transcriptional upregulation of certain genes, often resulting in cell proliferation (Malumbres *et al*¸ 2003). In addition, RAS genes have an additional mechanism of apoptotic protection by interacting with the PI3K (phosphoinositide 3-kinase) pathway (Malumbres *et al*¸ 2003). Binding of RASSFs to RASs has been shown to have an inhibitory effect on RAS activity and RASSFs 1A, 1C, 2, 4 and 5 have been shown to use this mechanism to mediate some of their apoptotic effects (Ortiz-Vega *et al*¸ 2002; Vos *et al*¸ 2000; Vos *et al*, 2003a; Eckfield *et al*¸ 2004; Vos *et al*, 2003b). RAS genes are frequently mutated in multiple tumour types and in pancreatic, colorectal and endometrial cancers, *RASSF1A* methylation has been shown to be lower in samples with *K-RAS* mutations (Dammann *et al*, 2003; van Engerland *et al*, 2002; Kang *et al*, 2006), suggesting dual mechanisms for increased cell proliferation of cancer cells by this mechanism.

As mentioned, classical RASSFs contain a SARAH (SALVADOR-RASSF-HIPPO) domain. SALVADOR and HIPPO are part of the SWH (Salvador/Warts/Hippo) pathway in *Drosophila melanogaster*. This pathway is involved in cell and tissue size and is capable of tumour suppressor effects when activated by inhibiting the transcriptional activity of the downstream member of the pathway, *Yorkie* (human homologue *YAP1*) (see section 6.2 for more information on the SWH pathway). The *D.melanogaster* orthologue of *RASSF* has been shown to inhibit *Hippo* activity by competing with *Salvador* (Polesello *et al*, 2006). Human homologues of *Salvador*, *Warts* and *Hippo* are *SAV1*, *LATS1/2* and *MST1/2* respectively.

RASSF members 1A, 1C, 2, 3, 4 and 5 have all been shown to bind with MST1 and in some cases, MST2 also, upon binding MST1/2 pro-apoptotic signals are activated (Praskova *et al*, 2004; Guo *et al*, 2007; Eckfield *et al*, 2004; Khokhlatchev *et al*, 2002).

RASSFs have been associated with inducing/promoting apoptosis by many mechanisms. Results from various studies have implicated interactions of RASSF1A with multiple proteins to bring about apoptosis including CNK1, C19ORF5, MOAP-1, MST1, RAS, RASSF5 (Vos *et al*, 2006; Rabizadeh *et al*, 2004; Liu *et al*, 2005). RASSF1C, RASSF2, RASSF4, RASSF5 and RASSF6 have also been shown to promote apoptosis (Kitagawa *et al*, 2006; Vos *et al*, 2003b; Akino *et al*, 2005; Eckfield *et al*, 2004; Khokhlatchev *et al*, 2002; Ikeda *et al*, 2007). RASSF1A has also been implicated in cell cycle control (Agathanggelou *et al*, 2003; Shivakumar *et al*, 2002; Deng *et al*, 2008).

1.5.2 The Classical RASSFs and promoter hypermethylation

RASSF1 is located on chromosome 3p (3p21.3) and was discovered due to it being within a commonly deleted region of lung cancer (Dammann *et al*, 2000). Although seven *RASSF1* isoforms have been described (*RASSF1A – RASSF1G*), only *RASSF1A* and *RASSF1C* have known biological activity (Richter *et al*, 2009). *RASSF1A* and *RASSF1C* contain 6 and 5 exons respectively and are transcribed by separate promoters, both of which contain CpG islands. The initial study by Dammann *et al* (2000) identified hypermethylation of *RASSF1A* in 40% primary lung tumours. Since then, a large number of studies have described frequent methylation of *RASSF1A* in numerous cancer types including primary breast tumours (Dammann *et al*, 2001), primary renal cell carcinoma (Morrissey *et al*, 2001), primary colorectal tumours (van Engerland *et al*, 2002) and primary glioma tumours (Hesson *et al*, 2004). Many subsequent studies have confirmed these results and shown frequent methylation

of *RASSF1A* in many other tumour types and found diagnostic and prognostic potential of *RASSF1A* methylation status (see Hesson *et al*, 2007 for a review and references therein). *RASSF1A* is ubiquitously expressed in normal tissue but frequently downregulated in multiple tumour types (Dammann *et al*, 2000; Agathanggelou *et al*, 2005). Hypermethylation of *RASSF1* appears to be specific of the *RASSF1A* CpG island (Dammann *et al*, 2000; Burbee *et al*, 2001).

RASSF2, located at 20q13, exists as two isoforms, *RASSF2A* and *RASSF2C*, of which only *RASSF2A* is transcribed from a CpG island region. Promoter hypermethylation of the *RASSF2A* CpG island and subsequent downregulation of *RASSF2A* has been associated with multiple tumours, including colorectal (Hesson *et al*¸2005; Akino *et al*, 2005; Park *et al*, 2007), gastric (Endoh *et al*, 2005), nasopharyngeal (Zhang *et al*, 2007) and thyroid cancers (Schagdarsurengin *et al*, 2010).

RASSF3 is located at 12q14.2, has five exons and is transcribed as a single isoform from a CpG island region. Promoter hypermethylation of *RASSF3* has not been detected in any of the cancers studied to date nor has downregulation been observed in cancer cell lines (Hesson *et al*, 2005).

As with the other classical RASSFs, *RASSF4* is transcribed from a CpG island region. It is located on chromosome 10 (10q11.21) and contains 11 exons. *RASSF4* mRNA has been detected in most normal tissues but downregulated in many cancer types (Eckfield *et al*¸ 2004). Hypermethylation has been demonstrated for *RASSF4* in kidney, lung and breast cancer cell lines and in the respective primary tumours (Eckfield *et al*, 2004). In addition, downregulation resulting from hypermethylation can be reversed using 5-aza-2'-deoxycitidine treatment and over-expression of *RASSF4* has been shown to be capable of inhibiting growth

and inducing apoptosis in cancer cell lines, suggesting biological relevance of *RASSF4* hypermethylation (Eckfeld *et al*, 2004).

There are three *RASSF5* isoforms, *RASSF5A-RASSF5C*, of which all are transcribed from CpG island regions. *RASSF5A* and *RASSF5B* are transcribed from a 1208 bp CpG island region containing 135 CpGs whilst *RASSF5C* is transcribed from a smaller, 511bp CpG island region containing 60 CpGs. *RASSF5* was originally named *NORE1* (*Novel Ras Effector 1*) and is sometimes still referred to as such. Confusingly, *RASSF5A* is known as *NORE1A* whilst *RASSF5C* is known as *NORE1B*. Both *RASSF5A* and *RASSF5C* are expressed in most normal tissues (Tommasi *et al*¸ 2002) and have been extensively studied for promoter hypermethylation. It appears to be a rare event in the cancers analysed to date and any methylation detected appears to be predominantly in *RASSF5A* rather than *RASSF5C* (Tommasi *et al*, 2002; Hesson *et al*, 2003). The highest frequency of *RASSF5A* methylation detected in primary tumours to date is in lung cancer, in which methylation was present in 24% of samples and specific for the NSCLC subtype (Hesson *et al*, 2003). Whilst methylation is uncommon, downregulation has been demonstrated in some cancer cell lines (Tommasi *et al*, 2002).

RASSF6 is located at 4q13.3 and is transcribed from a small 214bp CpG island region containing 20 CpGs. *RASSF6* promoter hypermethylation is a rare occurrence, having been only observed frequently in childhood acute lymphocytic leukaemias (ALL) but none of the solid tumours analysed to date (Hesson *et al*, 2009).

1.5.3 The N-terminal RASSFs

1.5.3.1 RASSF7

RASSF7 is located at 11p15.5 and is transcribed from a promoter region containing a 1015bp CpG island with 115 CpGs. Originally named *HRC1* (*HRAS1 related cluster protein 1*), the *RASSF7* gene produces three transcripts from alternative splicing resulting in varying C-terminal ends. All three protein variants contain an RA domain at their N-termini. Hypermethylation of the *RASSF7* CpG island has been searched for in a large number of various cancer cell lines but no methylation has been identified (Recino *et al*, 2010). Selected cancer cell lines were also sequenced for mutations but to no avail (Recino *et al*, 2010). However, unusually for a RASSF family member, upregulation of *RASSF7* has been observed in several cancers, including pancreatic ductal carcinoma, endometrial cancer and clear cell ovarian cancer (Brandt *et al*¸ 2007; Friess *et al*¸ 2003; Logsdon *et al*, 2003; Lowe *et al*¸ 2007; Mutter *et al*¸ 2001; Tan *et al*, 2009). RASSF7 has been shown to have roles in cell growth, apoptosis and mitotic progression (Sherwood *et al*, 2008; Recino *et al*, 2010; Hitomi *et al*¸ 2008; Goshima *et al*¸ 2007).

1.5.3.2 RASSF8

Located at 12p12.3, *RASSF8* contains 6 exons that are alternatively spliced to form four predicted variants, giving rise to two isoforms of 419 and 392 amino acids, each containing an RA domain at its N-terminus. *RASSF8* is unbiquitously expressed in normal tissue (Lock *et al*, 2010). The UCSC genome browser does not recognise a CpG island at the 5' end of *RASSF8*, however use of the EBI (European Bioinformatics Institute) CpG plot program detects a CpG island of 1201bp containing 182 CpGs situated upstream of and including the transcritption start site of *RASSF8*. CoBRA analysis of this region has been carried out in

numerous solid tumour cancer cell lines but no methylation has been found (Lock *et al*, 2010). Methylation was observed in 33% of leukaemia cell lines, although this proved to be infrequent when analysed in leukaemia patient samples (10% in childhood T-ALL and 9% in childhood B-ALL) (Hesson *et al*, 2009). Despite the lack of methylation detected in *RASSF8*, significant downregulation has been observed in lung cancer, and, when re-introduced to *RASSF8* deficient cells, *RASSF8* is able to inhibit anchorage-independent growth (Falvella *et al*, 2006). Knockdown of *RASSF8* in expressing lung tumour cell lines results in increased anchorage-independent growth and loss of contact-dependent growth inhibition (Lock *et al*, 2010). *RASSF8*, therefore, behaves much like a classical RASSF family member in terms of tumour suppressor properties despite the lack of hypermethylation in cancers. The mechanisms by which this loss occurs remain to be elucidated.

1.5.3.3 RASSF9

RASSF9 was previously known as *P-CIP1* (*PAM COOH-terminal interactor-1*) and is located at 12q21.31. It is a two exon gene producing a 435 amino acid protein. Neither the UCSC genome browser nor the EBI CpG plot program were able to detect a CpG island located at or near the 5' end of *RASSF9*. Little is known about *RASSF9* function, but it is predicted to be widely expressed and the RASSF9 protein has shown preferential binding to N-RAS and K-RAS (Rodriguez-Viciana *et al*, 2004).

1.5.3.4 RASSF10

RASSF10 (previously known as *LOC644943*), located at 11p15.2, is the latest member of the RASSFs to be described and was identified through sequence homology to *RASSF9* (Sherwood *et al*, 2008). Human genome browsers described *RASSF10* (*LOC644943*) as a two exon gene producing a 615 amino acid protein. Previous work in this laboratory identified an

alternative transcription start site that would result in a single exon gene producing a 507 amino acid protein along with 5' and 3' UTRs (untranslated regions) which were not described in human genome browsers (Hesson *et al*, 2009). Expression of *RASSF10* has since been shown in multiple tissues (Hesson *et al*, 2009; Schagdarsurengin *et al*, 2009) and, as already mentioned, promoter region hypermethylation has been shown in childhood ALL and thyroid cancers (Hesson *et al*, 2009; Schagdarsurengin *et al*, 2009). Little else is known about *RASSF10* but the diverse nature of leukaemia and thyroid cancer suggests *RASSF10* may play a general role in tumour suppression like classical RASSF family members.

1.6 THE SWH PATHWAY

The Salvador/Warts/Hippo (SWH) pathway interacts with the classical RASSF family members via their SARAH domain. Deregulation of the SWH pathway by multiple mechanisms, RASSF downregulation included, is common in many cancers (Pan, 2010). A new member of the family, *KIBRA* (*kidney and brain protein*) has recently been described (Yu *et al*, 2010; Genevet *et al*, 2010) and presents a new and interesting target for investigation. There are many facets to the SWH pathway that could, and have been, discussed in great depth (Pan, 2010; Grusche *et al*, 2010), however a brief summary of the pathway should elucidate its importance and the importance of finding a new member.

The Salvador/Warts/Hippo (SWH) pathway was discovered in *Drosophila melanogaster* as a modifier of growth (Tapon *et al*, 2002). Loss of function of the originally identified core SWH members results in an overgrowth phenotype exhibiting high proliferation and low levels of apoptosis (Justice *et al*¸1995; Tapon *et al*, 2002; Wu *et al*, 2003; Lai *et al*, 2005). This function has since been shown to be conserved in mammals. As a result, the SWH pathway is capable of affecting tissue and organ size, making disruption of the pathway a target for tumourigenesis. Members of the pathway can be loosely classified as either core components, upstream members or downstream members Although the classification of genes into these groups sometimes differs, for the purposes of this introduction, the core components will be considered as *Sav*, *Warts*, *Hippo* and *Mats*.

1.6.1 Core SWH components

There are four members of the pathway in *Drosophila melanogaster* considered core components; two serine/threonine kinases, *Warts (Wts)* and *Hippo (Hpo)* and two adapter proteins, *Mob as tumour suppressor (Mats)* and *Salvador (Sav)* (Justice *et al*¸1995; Tapon *et*

al, 2002; Wu *et al*, 2003; Lai *et al*, 2005). These four members exist as six respective homologues in *Homo sapiens*; two paired serine/threonine kinases, *LATS1/2* and *MST1/2*, and two adapter proteins, *MOBKL1A/B* and *SAV1* (Tao *et al*, 1999; Yabuta *et al*, 2000; Harvey *et al*, 2003; Tapon *et al*, 2002; Lai *et al*, 2005).

Core components are involved in control of cell growth, proliferation and apoptosis via inactivation of the downstream transcriptional co-activator *Yorkie (Yki)*, human homologues *YAP1* and *TAZ* (Grusche *et al*, 2010).

1.6.2 Downstream SWH pathway members

The main downstream member of the SWH pathway, through which all regulation is currently known to occur, is *Yki* (human homologues *YAP1* and *TAZ*). *Yki* is a transcriptional co-activator and mediates its effects by binding to DNA-binding transcription factors, such as *Scalloped* (human homologues *TEAD1-4*) (Wu *et al*¸ 2008). Potential targets of *Yki* include *Diap1* (human homologue *BIRC2*), *Cyclin E* (human homologue *CCNE1*) and *Bantam (Bam*) (human homologue unknown). *Diap1* is thought to confer resistance to apoptosis, *Cyclin E* is thought to mediate ectopic proliferation and the *Bantam* locus encodes a miRNA capable of inhibiting apoptosis and promoting growth; however little is currently known about its target genes (Harvey *et al*, 2007).

1.6.3 Upstream SWH pathway members

There are many upstream members of the SWH pathway known to regulate the core SWH components and novel modifiers are frequently being identified. Currently, upstream members can be split into those that signal via *Fat* (human homologue *FAT1-4*), those that are part of the *Kibra-Expanded-Merlin* complex and those that are apico basal polarity proteins (Grusche *et al*, 2010). Activation of these members/pathways results in activation of *Hpo* by

phosphorylation, which in turn phosphorylates *Sav, Mats* and *Warts.* This results in inactivation of *Yki* by phosphorylation culminating in repression of its transcriptional targets and subsequent reduction in cell number, hence controlling tissue and organ size.

Figure 1.18 shows a schematic of the SWH pathway. It can be seen from this diagram that the pathway is more intricate than discussed here, particularly in drosophila, and that there are multiple ways in which the activity of pathway members can be modified.

1.6.4 The SWH pathway and cancer

The core components of the SWH pathway were identified through drosophila screening experiments for tumour suppressor genes. Since then, deregulation of the human/mammalian SWH pathway has been observed in several cancer types.

Mechanisms include mutations (*Mer/NF2* in neurofibromatosis 2 and *MOBKL11* in melanoma and mouse mammary carcinoma), deletions (*SAV1* in RCC cell lines), hypermethylation (*MST1/2* in soft tissue sarcoma and *LATS1/2* in astrocytoma and breast cancer) and general downregulation (*MOBKL1* in colorectal and lung and *Mst1/2* in colorectal and prostate cancers) (Rouleau *et al*, 1993; Lai *et al*, 2005; Tapon *et al*, 2002; Seidel *et al*, 2007; Jiang *et al* 2006; Takahashi *et al*, 2005; Kosaka *et al*, 2007; Sasaki *et al*, 2007). All of the above mechanisms result in inactivation of the relevant genes. In contrast, the *YAP1/TAZ-TEAD* transcription factor complex is considered to have oncogenic properties. Locus amplification has been observed in many types of mouse and human cancers and *YAP1* overexpression has been observed in many cancers including lung, medulloblastomas and esophageal (Dai *et al*, 2003; Fernandez-L *et al*¸2009; Imoto *et al*, 2001). In addition, both *TAZ* and *TEAD* have shown either locus amplification and/or gene upregulation in a smaller variety of cancers; *TAZ* overexpression has been observed in breast cancer (Chan *et al*, 2008)

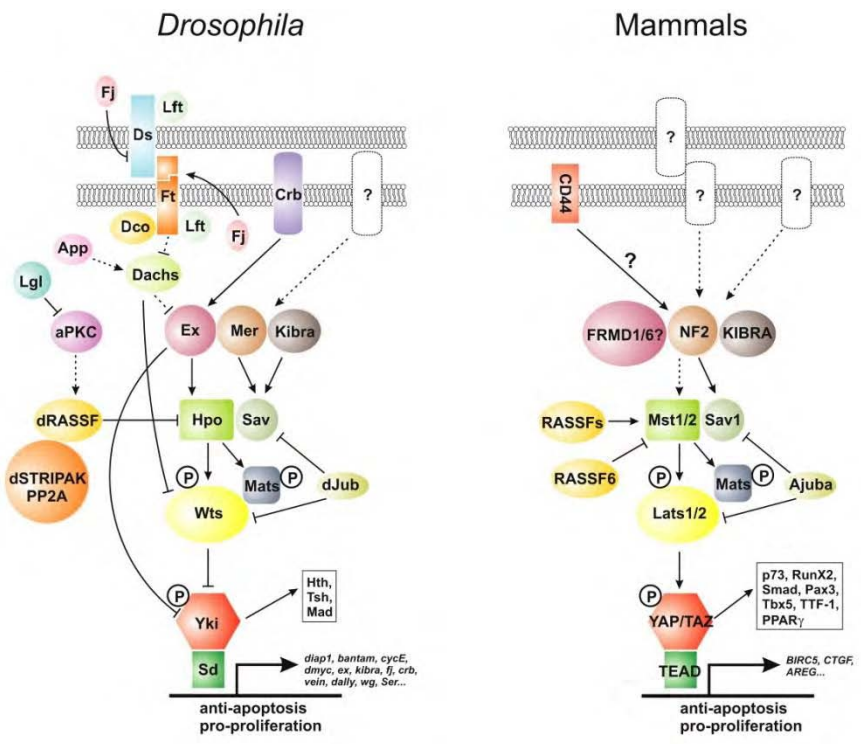


Figure 1.18 SWH pathway in *drosophila* and mammals
This figure has been taken from Pan, 2010. It nicely shows the complex interactions in the SWH pathway for both *drosophila* and mammals. Due to the complexity of the pathway, not all genes in the *drosophila* pathway have been discussed.

whilst *TEAD4* amplification has been observed in ovarian and fallopian tube carcinomas (Nowee *et al*, 2007). Both downregulation of SWH pathway members upstream of *YAP1* and upregulation of the YAP1/TAZ-TEAD complex would result in similar phenotypic activity of *YAP1* transcriptional targets. However, there is also evidence to suggest *YAP1* is capable of acting as a tumour suppressor by promoting apoptosis via p73 (Yuan *et al*, 2008; Basu *et al*, 2003). It is thought these conflicting effects of *YAP1* may be reliant on cellular context (Pan, 2010).

KIBRA (*kidney and brain protein*), also known as *WWC1* (*WWC and C2 domain containing 1*), was only recently described as an SWH pathway member and has been shown to function in a tumour suppressor capacity in *drosophila melanogaster* (Yu *et al*, 2010; Genevet *et al*, 2010). *KIBRA* therefore represents a good candidate gene for genetic analysis, including hypermethylation analysis, in various cancers.

1.7 RESEARCH AIMS

Hopefully this introduction has highlighted the reasons why the identification of hypermethylated genes in cancer can be of use both scientifically and clinically within the cancer research field and how identification of such genes can be achieved. Broadly, the aims of this thesis are to identify hypermethylated genes in cancer to aid in this research. The main focus will be on sporadic breast cancer, using two approaches;

(i) To use the Methylated CpG Island Recovery Assay (MIRA), a genome wide affinity pull-down approach, coupled with CpG island microarrays, on 5 breast cancer cell lines. To analyse the data from this approach to identify a list of frequently hypermethylated genes in cell lines that can then be used to selectively screen sporadic breast cancer patient samples for novel frequently hypermethylated genes.

(ii) To use the Illumina Infinium HumanMethylation27 BeadChip array, a direct quantitative microarray based assessment of DNA using single CpG specific probes within or within close proximity to known CpG islands, to screen a clinically well characterised cohort of breast cancer patient DNA samples. Clinical comparisons can then be made to identify differences in overall methylation between differing clinical features and single genes that show an association between methylation and particular clinical features.

This section of this research will hopefully identify genes where the demonstration of hypermethylation will either be of use to those investigating the cellular basis of breast (and/or other cancers) or those investigating molecular markers that can be of use clinically for breast cancer patients.

The final section of this research aims to analyse potentially good candidate genes that are part of families or pathways already known to be involved in various cancers. Recently identified members of the RASSF family of tumour suppressors (*RASSF10*) and the SWH signalling pathway (*KIBRA*) will be analysed for hypermethylation in a range of cancers to enhance knowledge of members of this family and pathway that are involved in cancer. This research will be of use to those working in the RASSF family and SWH pathway fields and hopefully those working in particular cancer fields both scientifically and clinically.

CHAPTER 2

MATERIALS AND METHODS

2.1 MATERIALS

All primers were supplied from either Invitrogen or Alta Biosciences. All materials were acquired using the suppliers and product numbers cited within the text. Suppliers addresses are listed below:

Alta Bioscience

Alta Bioscience, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK.

AMS Biotechnology (Europe) Ltd.

184 Milton Park, Abingdon, OX14 4SE, UK.

Bioline

16 The Edge Business Centre, Humber Road, London, NW2 6EW, UK.

Fermentas

Sheriff House, Sheriff Hutton Industrial Park, York, YO60 6RZ, UK.

Invitrogen Ltd.

3 Fountain Drive, Inchinnan Business Park, Paisley, UK, PA4 9RF, UK.

National Diagnostics USA

National Diagnostics USA, 305 Patton Drive, Atlanta, Georgia, 30336, UK.

New England Biolabs

75/77 Knowle Piece, Wilbury Way, Hitchin, Hertfordshire, SG4 0TY, UK.

PAA Laboratories Ltd.

Termare Close, Houndstone Business Park, Yeovil, Somerset, BA22 8YG, UK.

Promega Corporation

2800 Woods Hollow Road, Madison, WI 53711-5300, USA.

Roche Diagnostics Ltd.

Charles Avenue, Burgess Hill, West Sussex, RH15 9RY.

Sigma-Aldrich

Sigma-Aldrich Company Ltd., Dorset, England, UK.

Qiagen

Qiagen House, Fleming Way, Crawley, West Sussex, RH10 9NQ, UK.

2.2 DNA SAMPLES

2.2.1 Cell lines

2.2.1.1 Breast cell lines

A total of nine breast cancer cell lines were analysed in this study; HCC1806, HCC1937, HCC1395, HCC1419, HCC1143, MCF-7, T-47D, MDA-MB-231 and HTB19. All HCC lines were kindly donated by Professor John D. Minna (Hamon Center for Therapeutic Oncology Research, The University of Texas Southwestern Medical Center at Dallas, Dallas, Texas, USA). T-47D and MCF-7 were donated by Dr. Feodor Berdichevsky (Cancer Research UK Institute of Cancer Studies, Birmingham University, Birmingham, UK). MDA-MB-231 and HTB19 were already available within the laboratory. Two normal human mammary epithelial cell lines (NHMEC) were also used in this study and were also donated by Professor John D. Minna (Hamon Center for Therapeutic Oncology Research, The University of Texas Southwestern Medical Center at Dallas, Dallas, Texas, USA).

2.2.1.2 Lung cell lines

Fourteen lung cancer cell lines were used in this study (H1993, H157, H1792, H143, H1648, H1155, H1299, H2122, A549, H2171, H460, H187, H838, H1395). All lung cancer cell lines were kindly donated by Professor John D. Minna (Hamon Center for Therapeutic Oncology Research, The University of Texas Southwestern Medical Center at Dallas, Dallas, Texas, USA).

2.2.1.3 Glioma cell lines

Seven glioma cell lines were analysed in this study (Hs683, A172, H4, U343, T17, U87 and U373), all kindly provided by Prof. Dietmar Krex (Klinik und Poliklinik für Neurochirurgie, Universitätsklinikum Carl Gustav Carus, Technische Universität Dresden, 01307 Dresden, Germany).

2.2.1.4 Kidney cell lines

Twelve renal cell carcinoma (RCC) cell lines were used in this study (SKRC18, SKRC39, SKRC47, SKRC54, UMRC3, UMRC2, SKRC45, 786-O, KTCL140, KTCL26, RCC4 and CAKI) all kindly provided by Professor Eamonn R. Maher (Medical and Molecular Genetics, School of Clinical and Experimental Medicine, College of Medical and Dental Sciences, University of Birmingham, Birmingham, B15 2TT, UK).

2.2.1.5 Prostate cell lines

Five prostate cancer cell lines (DU-145, PC3, LNCaP, 22Rv1 and VCaP) were kindly donated by Dr. Tapio Visakorpi (Professor of Cancer Genetics, Institute of Medical Technology, University of Tampere, Biokatu 6, FI-33520 Tampere, Finland) for use in this study.

2.2.1.6 Leukaemia cell lines

Seven leukaemia lines (JURKAT, NALM6, U937, SUP-T1, CEM, DND41 and THP1) were used in this study, all donated by Dr. Raffaella Chiaramonte (Department of Medicine, Surgery and Dentistry, Università degli Studi di Milano, via Di Rudinì 8, 20142 Milan, Italy).

2.2.1.7 Colorectal cell lines

Seven colorectal cancer (CRC) cell lines were used in this study (LS411, DLD1, SW60, SW480, 174T, HVT116 and LOVO), all kindly provided by Professor Dion Morton (Medical and Molecular Genetics, University of Birmingham, Birmingham, UK).

Appendix A provides details of major cancer cell lines used in this study.

2.2.2 Patient samples

All primary tumour samples have been anonymised and collected following ethical approval.

2.2.2.1 Breast tissue DNA

Forty breast ductal carcinoma samples were provided by Dr. Ivan Bieche (Oncogenetic Laboratory, INSERM U 735, Centre René Huguenin, Saint Cloud, France). These samples are listed in table 4.1 along with the corresponding clinical features. A further twenty pairs of breast ductal carcinoma samples along with corresponding non-cancerous breast tissue DNA were also used in this study, kindly donated by Dr. Cheryl Lewis (Department of Surgery, University of Texas Southwestern Medical Center, Dallas, Texas, USA). Normal breast tissue DNA and RNA samples were purchased from AMS biotechnology (Cat.Nos: D1234086-50 and R1234086-50 respectively). Samples were not microdissected.

2.2.2.2 Glioma samples

A total of 68 primary glioma tumour DNA samples consisting of 10 WHO grade I astrocytomas, 10 WHO grade II astrocytomas, 10 WHO grade III astrocytomas, 20 WHO grade IV primary glioblastoma multiforme (pGBM) and 18 WHO grade IV secondary glioblastoma multiforme (sGBM), 11 of which also had available DNA from the corresponding earlier grade lesion. All glioma tumour DNA samples were kindly donated by Prof. Dietmar Krex (Klinik und Poliklinik für Neurochirurgie, Universitätsklinikum Carl Gustav Carus, Technische Universität Dresden, 01307 Dresden, Germany). Prof. Krex also generously provided 15 non-cancerous brain samples for use as a control.

2.2.2.3 Childhood acute lymphoblastic leukaemia samples

Seventy childhood acute lymphoblastic leukaemia (ALL) samples were provided by Dr. Daniel Catchpoole (The Children's Hospital at Westmead, Locked Bag 4001, Westmead, NSW, 2145. Australia). A further five sets of paired childhood ALL DNAs, with DNA samples extracted from both diagnosis stage tissue and the disease remission stage tissue, were obtained from a local tissue bank. Normal bone marrow was purchased from AMS biotechnology.

2.2.2.4 NSCLC tumours

Twenty pairs of NSCLC cancerous tissue DNA and corresponding non-cancerous lung tissue DNA were kindly provided by Professor John D. Minna (Hamon Center for Therapeutic Oncology Research, The University of Texas Southwestern Medical Center at Dallas, Dallas, Texas, USA). Samples were not microdissected.

2.2.2.5 RCC tumours

Sixteen RCC tumour samples were kindly provided by Professor Eamonn R. Maher (Medical and Molecular Genetics, School of Clinical and Experimental Medicine, College of Medical and Dental Sciences, University of Birmingham, Birmingham, B15 2TT, UK) for use in this study. Samples were not microdissected.

2.2.2.6 Ovarian tumours

Twenty ovarian cancer DNA samples were kindly provided by Dr. Massimo Broggini (Istituto di Ricerche Farmacologiche "Mario Negri", Milan, Italy).

2.3 NUCLEIC ACID EXTRACTION

2.3.1 DNA Extraction

DNA was extracted from cell line pellets (see section 2.8.6 for cell pellet preparation) using the Qiagen DNeasy Blood and Tissue Kit (Cat.No. 69506). Manufacturers' instructions were followed using the Spin-Column protocol for 'Total DNA extraction from Animal blood or Cells'. Briefly, following pellet resuspension in 200µl PBS (phosphate buffered saline), 20µl proteinase K and 200µl buffer AL were added and mixed by vortexing. Samples were incubated at 56˚C for 10 minutes. Following this incubation period, 200µl 100% ethanol was added and samples were vortexed until a homogenous solution was produced. After transfer to a DNeasy Mini spin column with a 2ml collection tube, samples were centrifuged for 1 minute at 11,200 x g at room temperature. Flow-through was discarded and a clean 2ml collection tube attached to the spin column. After the addition of 500µl buffer AW1, samples were centrifuged for 1 minute at 11200 x g at room temperature and again the flow-through

was discarded and a new 2ml collection tube attached to the spin column. 200µl buffer AW2 was added to the column and samples centrifuged for 3 minutes at 11200 x g at room temperature. This final flow-though was discarded carefully to ensure that it did not come into contact with the membrane and the 2ml collection tube replaced with a labelled 1.5ml microcentrifuge tube. DNA was eluted from the membrane by adding 200µl of buffer AE directly to the membrane, allowing samples to incubate at room temperature for 1 minute followed by centrifugation at 11,200 x g for 1 minute. The extracted DNA was stored at -20°C.

2.3.2 RNA extraction

RNA was extracted from cell pellets (see section 2.8.6 for cell pellet preparation) using RNA-Bee reagent (AMS Biotechnology; Cat. No: CS-105B). Cells were first homogenised by resuspension of the cell pellet in 1ml RNA-Bee reagent. Samples were incubated at room temperature for 5 minutes prior to the addition of 200µl chloroform (200µl per 1ml RNA-Bee reagent), samples were then shaken vigorously for 15-30 seconds and incubated on ice for 5 minutes. Subsequent centrifugation at 12,000 x g for 5 minutes at 4˚C separated the homogenate into 3 distinct phases; a lower blue phenol-chloroform phase, an opaque, white interphase and an upper colourless aqueous phase. The aqueous phase (containing the entirety of the RNA) was transferred, being careful not to disrupt the interphase layer, into a clean RNase free 1.5ml microcentrifuge tube. To precipitate the RNA, 500µl isopropanol (500µl per 1 ml RNA-Bee reagent) was added, mixed and samples were allowed to incubate at room temperature for 5-10 minutes prior to centrifugation at 16,300 x g for 5 minutes at 4˚C. A visible pellet was produced and, taking care not to disturb the pellet, the supernatant was removed and 1ml 75% ethanol was added to the pellet. Samples were vortexed and centrifuged at 7,500 x g for 5 minutes at 4˚C. The supernatant was removed, again taking care

not to disrupt the pellet, and samples allowed to briefly air-dry before resuspension in 50µl RNase free water. The extracted RNA was stored at -80°C.

2.3.3 Quantification of nucleic acid preparations

A NanoDrop 1000 Spectrophotometer from NanoDrop Technologies was used to quantify DNA and RNA preparations. Initially, 2µl of the solution used to elute nucleic acids during preparation (i.e. buffer AE or RNase free water) was used as a reference for zero absorbance. 2µl of undiluted nucleic acid preparation was then added to the spectrophotometer. The NanoDrop was set to read absorbance at 260nm and 280nm. NanoDrop output readings consisted of a 260/280 ratio and a ng/µl concentration. Purity of the DNA and RNA preparations was assessed using the 260/280 ratio; DNA was considered pure if the ratio was ~1.8 and RNA considered pure if the ratio was ~2.0. The spectrophotometer calculated the ng/µl concentrations using a modification of the Beer-Lambert equation that correlates absorbance with concentration: **c = (A*e)/b**

Where **c** is the nucleic acid concentration in ng/µl, **A** is absorbance at 260nm in AU, and **e** is the wavelength dependent extinction coefficient in ng-cm/µl for either double stranded DNA (50ng-cm/µl) or RNA (40ng-cm/µl).

2.4 DNA MODIFICATION FOR METHYLATION ANALYSES

2.4.1 Bisulphite modification of DNA

The studies in this thesis used one of two methods to bisulphite modify DNA; (i) a traditional in-house method or, (ii) by use of a Qiagen EpiTect Bisulphite Kit (48) (Cat.No.59104). Work from chapter 3 (MIRA array study) was achieved using the traditional method and chapters 4,

5 and 6 (Illumina Inifinium array, *RASSF10* and *KIBRA* studies respectively) using the Qiagen kit.

2.4.1.1 Bisulphite modification of DNA using the traditional method

Preparation of sodium bisulphite reagent

Firstly, fresh sodium bisulphite reagent (3.12M sodium metabisulphite, 0.3M NaOH, 5mM hydroquinone, pH 5.0) was made. For 20mls, 11.86g sodium metabisulphite (Sigma; Cat.No.243973) was dissolved in 15mls of distilled water by heating at 100˚C with intermittent vortexing to ensure complete solubilisation. Once dissolved, the solution was made up to a total volume of 20mls with distilled water allowing for the addition of 2mls 3M sodium hydroxide (Sigma; Cat.No. 255556). Finally, 100µl freshly prepared 1M hydroquinone (Sigma; Cat.No H9003) was added in a fume hood. As both the sodium metabisulphite mix and hydroquinone are light sensitive all solutions were covered with foil at all times.

Bisulpite conversion reaction

1µg DNA was prepared for bisulphite conversion by the addition of 1µl 3M NaOH and the necessary amount of distilled water to produce a total volume of 10µl with water. Following incubation of the samples at 37˚C for 15 minutes, 500µl freshly prepared bisulphite reagent was added to each sample in a fume hood and overlaid with mineral oil to prevent evaporation during the subsequent steps. Samples were then incubated for 20 cycles of [99˚C for 15 seconds followed by 50˚C for 15 minutes] in a Hybaid Omn E thermocycler.

Clean up of modified DNA

Bisulphite converted DNA was purified using a Promega Wizard DNA clean-up system (Promega Cat.No.7280). The bisulphite conversion reactions were added to 1ml DNA clean-up resin in a 1.5ml microcentrifuge tube, taking care to transfer as little mineral oil as possible. Samples were mixed by inversion and then transferred to syringe barrels attached to spin columns, which in turn were attached to a vacuum manifold. The DNA/resin mix was drawn through the spin columns using the vacuum manifold. Subsequently, 2mls 80% (v/v) isopropanol was added to each syringe barrel and drawn through to wash the spin columns. The spin columns were removed and placed in 1.5ml microcentrifuge tubes and centrifuged at room temperature for 2 minutes at 16,300 x g. The spin columns were placed in fresh 1.5ml microcentrifuge tubes and 50µl of distilled water was added to each spin column membrane followed by a 5 minute incubation at room temperature. Samples were centrifuged at room temperature for 1 minute at 16,300 x g to elute the DNA from the column and 5µl 3M sodium hydroxide was added to each sample and allowed to incubate for 10 minutes at room temperature. To precipitate the DNA, 5µl sodium acetate and 125µl 100% ethanol was added to each sample followed by an overnight incubation at -20˚C. Samples were centrifuged at room temperature for 30 minutes at 16,300 x g, the supernatant was discarded and samples were washed with 100µl 70% (v/v) ethanol. Samples were again centrifuged at room temperature for 1 minute at 16,300 x g followed by removal of the supernatant prior to air drying and resuspension in 50µl distilled water. Eluted DNA was ready for use at a concentration of approximately 20ng/µl and was stored at -20°C.

2.4.1.2 Bisulphite modification of DNA using the Qiagen EpiTect Kit

Aliquots of bisulphite mix were dissolved in 800µl RNase-free water and vortexed until completely dissolved. The bisulphite reaction was prepared for 1µg DNA by addition of RNase-free water to a total volume of 20µl, followed by addition of 85µl bisulphite mix. 35µl of green DNA protect buffer was added to each sample and pipetted up and down until a homogenous blue colour change was produced, indicating sufficient mixing and correct pH. The bisulphite reaction was carried out in a Perkin Elmer Gene Amp PCR system 9700 (Applied Biosystems) thermal cycler using the conditions shown in table 2.1.

Following the bisulphite conversion the reactions were transferred to 1.5ml microcentrifuge tubes and 560µl freshly prepared buffer BL containing 10µg/ml carrier RNA was added to each sample. Samples were then vortexed and centrifuged briefly before transfer to an EpiTect spin column. Samples were then centrifuged at room temperature for 1 minute at 16,300 x g, flow-through was discarded and samples were washed with 500µl buffer BW. A further centrifugation step was carried out at room temperature for 1 minute at 16,300 x g. Flow-through was discarded and samples desulphonated by incubation at room temperature for 15 minutes in 500µl buffer BD, being careful to keep the lids closed. Samples were centrifuged at room temperature for 1 minute at 16,300 x g. Flow-through was again discarded and samples washed twice with 500µl buffer BW with 1 minute centrifugation steps (room temperature, 16,300 x g) each time. Samples were then spun again in clean collection tubes at room temperature for 1 minute at 16,300 x g to remove any residual liquid prior to elution. Elution of bisulphite modified DNA was achieved by adding 20µl buffer EB directly to the column and centrifugation at room temperature for 1 minute at 16,300 x g into clean 1.5ml microcentrifuge tubes. This elution step was repeated to increase the total yield of

Step	Temperature (˚C)	Time (minutes)
Denaturation	95	5
Incubation	60	25
Denaturation	95	5
Incubation	60	85
Denaturation	95	5
Incubation	60	175
Hold	20	Indefinite

Table 2.1_Bisulphite modification conditions for use in EpiTect kit.
Conditions are shown for bisulphite modification when using Qiagen EpiTect kit.

bisulphite modified DNA, to produce a final volume of 40µl. Bisulphite modified DNA was stored at -20°C.

2.5 COMBINED BISULPHITE RESTRICTION ANALYSIS (CoBRA) AND METHYLATION SPECIFIC PCR (MSP)

2.5.1 CoBRA primer design

In chapters 3, 5 and 6 (MIRA array, *RASSF10* and *KIBRA* studies respectively), CoBRA primers were designed to be within a UCSC defined CpG island and situated upstream of, and in some cases, incorporating the transcription start site. However, for the work in chapter 4 (Illumina Infinium array study) primers were specifically designed to be around the methylated probe within the array (where possible), irrelevant of its position within the CpG island.

All CoBRA primers used in this thesis were designed manually for this study unless otherwise stated. The CoBRA PCR was initially carried out using semi-nested primers in all cases unless the PCR yield was poor or non-specific, in which case fully nested primers were designed and used. To maximise the effectiveness of the CoBRA primers two important factors were incorporated into the design of the primers; (i) the most specific region, the 3'end, of forward and reverse primers, were designed to be situated at ACC or CAC genomic sites, becoming ATT or TAT in forward primers and TAA or ATA in reverse primers, and (ii) the central primer sequence ideally contained a string of genomic C nucleotides, becoming T in forward primers and A in reverse primers. This should provide a strong initial binding point and a high specificity to bisulphite modified DNA. It was important to ensure that CoBRA primers were unbiased towards methylation status. For this reason, the incorporation of CG

dinucleotides within the primer were avoided wherever possible, however if this was not possible C nucleotides were ordered as Y (designating a 50:50 ratio of C or T nucleotide incorporation) in forward primers, and R (designating a 50:50 ratio of G or A nucleotide incorporation) in reverse primers. The incorporation of BstUI restriction sites (CGCG) or TaqαI restriction sites (TCGA) within primers were strenuously avoided. Annealing temperatures of primers were calculated using the following equation: [(41 x GC)/N] + [64.9-(600/N)] where GC is the total number of C and G nucleotides in the primer sequence (Y and R counted as 0.5) and N is the total number of nucleotides within the primer. CoBRA primer sequences used in chapters 3, 4, 5 and 6 of this thesis are shown in appendices B.I, C.I, D.I, and E.I, respectively.

2.5.2 CoBRA PCR

All CoBRA assays consisted of using semi or fully nested primers and therefore two rounds of PCR were carried out per assay. Both primary and secondary reactions contained 1x 10X buffer (500mM Tris-HCL (pH 8.3), 100Mm potassium chloride, 50mM ammonium sulphate), 2mM magnesium chloride, 2.5mM dNTP mix, 1x 5X GC-RICH solution, 0.8μM forward primer, 0.8μM reverse primer and either 0.5U (primary reaction) or 1.0U (secondary reaction) FastStart Taq DNA polymerase (Roche; Cat.No.12158264001). Reactions were made up to 25μl or 50μl for primary and secondary reactions respectively with distilled water. Primary reactions used 2μl (0.04μg) bisulphite modified DNA as starting material and secondary reactions used 5μl primary product as starting material.

Primary CoBRA PCR programs were either a standard PCR program or touchdown PCR, see tables 2.2a and b for program details. All primer sequences and annealing temperatures are shown in appendices B.I, C.I, D.I and E.I.

(a)

		Primary CoBRA	Secondary CoBRA
95	5 minutes		
95	1 minute	35 cycles	40 cycles
annealing	1 minute		
72	2 minutes		
72	10 minutes		

(b)

		Primary CoBRA	Secondary CoBRA
95	5 minutes		
95	45 seconds	5 cycles with annealing temperature decreasing by 1˚C each cycle	5 cycles with annealing temperature decreasing by 1˚C each cycle
annealing	45 seconds		
72	45 seconds		
95	45 seconds	37 cycles	42 cycles
annealing	45 seconds		
72	45 seconds		
72	10 minutes		

Table 2.2_Standard and touchdown CoBRA PCR programs

(a) A table showing the PCR conditions used for primary and secondary CoBRA reactions done using the standard CoBRA PCR program. (b) A table showing the PCR conditions used for primary and secondary CoBRA reactions done using the touchdown CoBRA PCR program.

2.5.3 Agarose gel electrophoresis

PCR products were visualised with 2% (w/v) agarose (Bioline; Cat. No. BIO-41025) gels containing ethidium bromide (Sigma Aldrich; Cat. No: E8751). Gels were made by adding 4grams agarose to 200ml of 1X TBE (Tris-Borate-EDTA) diluted from 10X TBE (National Diagnostics Cat. No.EC-860) and heated until dissolved. Once cooled sufficiently, 5µl of 10mg/ml ethidium bromide was added to the gel to a final concentration of 0.25mg/ml and then poured. Gels were allowed to set at room temperature. Samples were loaded with 5% (v/v) loading buffer (50% glycerol, 48.5% distilled water, 1% EDTA and 0.5% Orange G). When loaded, DNA was migrated through the gel at 180 volts until ready to be visualised under UV light using a SynGene InGenius GelDoc system.

2.5.4 Restriction digest

A suitable volume of PCR product, between 5 and 20µl, depending on the intensity of the PCR band as determined by agarose gel electrophoresis (section 2.5.3) was digested with either BstUI (Fermentas;Cat.No ER0921) or TaqαI (New England Biolabs; Cat.No: R0149L). Due to the greater prevalence of CGCG sites over TCGA sites within CpG islands BstUI digestion was used as the primary screening method.

BstUI digests were carried out with 1u BstUI, 1x 10X buffer R, the relevant amount of PCR product and made up to a total volume of 25µl with distilled water. Samples were incubated overnight at 37˚C.

Samples being digested with TaqαI were incubated at 65˚C for three hours with 20u TaqαI, 0.01µl BSA, 1x 10X NEB buffer 4, the relevant amount of PCR produce and made up to a total volume of 25µl with distilled water.

Undigested and digested PCR products were run simultaneously, alongside each other on a 2% (w/v) agarose gel (section 2.5.3) to visualise the results.

2.5.5 Methylation Specific PCR (MSP)

Methylation specific PCR (MSP) and unmethylated specific PCR (USP) reactions were carried out with 1x 10X buffer (500mM Tris-HCL (pH 8.3), 100Mm potassium chloride, 50mM ammonium sulphate), 3mM magnesium chloride, 2.5mM dNTP mix, 1x 5X GC-RICH solution, 0.8µM forward primer, 0.8µM reverse primer and 1.0U FastStart Taq DNA polymerase (Roche; Cat.No.12158264001). Reactions were made up to 50µl with distilled water. MSP and USP reactions used 3µl (0.06µg) and 4µl (0.08µg) bisulphite modified DNA as starting material respectively. PCR reactions were performed using the standard PCR program listed in table 2.2a with annealing temperatures described in appendix C.II.

2.6 BISULPHITE SEQUENCING ANALYSIS

2.6.1 PCR product purification

PCR products selected for cloning were first purified from agarose gels using a Gel Extraction Kit (Qiagen; Cat.No. 28704). Once PCR products were resolved using standard agarose gel electrophoresis (see section 2.5.3) the bands were cut out of the gel using a scalpel and transferred to 1.5ml microcentrifuge tubes. Gel slices were weighed and 3 x mass (mg) in volume (µl) of buffer QG was added to samples prior to incubation at 50˚C for 10 minutes to dissolve the agarose gel. To ensure the agarose gel was completely dissolved intermittent vortexing was carried out during the 10 minute incubation. Once completely dissolved, 1 x mass (mg) in volume (µl) of 100% isopropanol was added to the solution and mixed. The solution was transferred to a QIAquick spin column and centrifuged at room temperature for 1 minute at 16,300 x g. The flow-through was discarded and a further 500µl buffer QG added to

the column before a second centrifugation step as before. Flow-through was again discarded and 500µl buffer PE was added to the column. Samples were centrifuged once as before and the flow-through discarded and then centrifuged again to ensure no residual ethanol remained. After placing the column in a fresh 1.5ml microcentrifuge tube, 30µl distilled water was added directly to the column membrane and left to incubate at room temperature for 1 minute prior to a 1 minute centrifugation step at room temperature to elute the purified DNA from the column. The purified PCR product was either used immediately or stored at -20°C.

2.6.2 Ligation of PCR products into plasmid vector

The pGEM-T Easy vector system (Promega; Cat.No. A1360) was used to introduce PCR products into bacterial cells for cloning. The ligation reaction was set up with 0.5µl pGEM vector, 1µl T4 DNA ligase and 5µl 2X ligation buffer. This was then added to 3.5µl purified PCR product. Samples were left to incubate at 4˚C overnight for maximum ligation efficiency. A plasmid map for the pGEM-T easy vector is shown in figure 2.1.

2.6.3 Transformation of bacterial cells with PCR product ligated into the pGEM-T easy vector

Prior to the transformation reaction, LB agar plates were prepared by mixing 35g/l LB agar (Sigma Aldrich; Cat. No: L2897) with distilled water before autoclaving. When sufficiently cooled, ampicillin (Sigma Aldrich; Cat. No:H0166) was added to achieve a final concentration of 100µg/ml. Agar was then poured into petri dishes and allowed to set before transfer to 4°C for storage.

To begin the transformation reaction, α-select silver efficiency chemically competent cells (Bioline; Cat.No. BIO-85026) were defrosted on ice and 40µl of the competent cells were aliquoted into 1.5ml microcentrifuge tubes. 4µl (10% volume of cells) of pGEM/PCR product

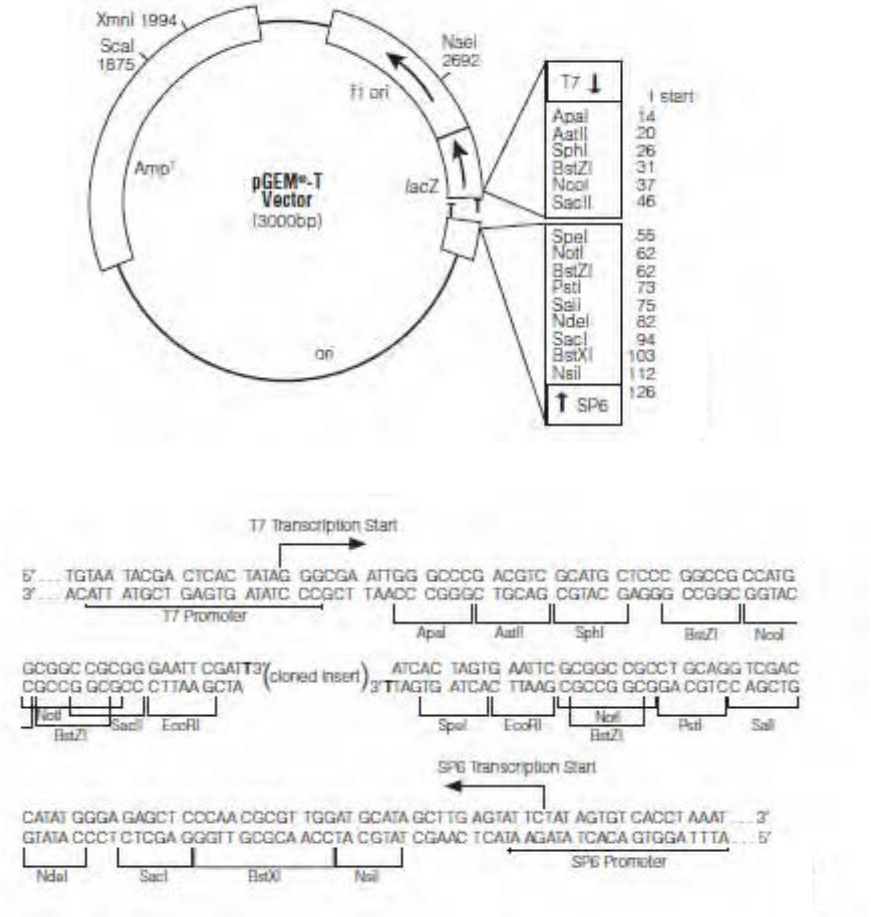


Figure 2.1_pGEM T-easy plasmid map

The above figure shows a plasmid map of the pGEM T-easy vector used for cloning of PCR products.

ligation was added to each aliqout defrosted competent cells and the solution was mixed by gentle swirling to avoid cell damage. Samples were incubated on ice for 20 minutes prior to a 30 second heat shock at 42˚C. Cells were then returned to ice for 2 minutes before adding 500µl SOC media (Invitrogen; Cat.No.15544034) and incubating for one hour on a rotating shelf at 37˚C. During this incubation, 20µl of 50mg/ml X-gal (Bioline; Cat. No: BIO37035) was added to each required LB plate and allowed to dry before 200µl of each culture was plated out and incubated for 16 hours at 37°C.

2.6.4 Single colony PCR

To ensure the greatest chance of picking colonies containing the insert, only white colonies were picked, preferentially those surrounded by blue colonies. This is due to an intrinsic feature of the pGEM-T easy vector that the cloning site exists within the coding β-galactosidase gene and successful insertion disrupts this gene. Undisrupted plasmids are capable of producing the β-galatosidase enzyme that can utilise the X-gal producing a blue by-product; thus blue colonies probably have empty plasmids while white colonies have inserts. Selecting a white colony surrounded by blue colonies merely suggests that plenty of X-gal was available to the colony therefore increasing the likelihood of picking successfully transformed cells. Colonies were transferred to 20µl distilled water and heated to 95˚C for 5 minutes to rupture all the cellular membranes. Once cooled, 7µl was then used as starting material in a PCR reaction containing 3µl 10x buffer (500mM Tris-HCL (pH 8.3), 100Mm potassium chloride, 50mM ammonium sulphate), 2mM magnesium chloride, 2.5mM dNTP mix, 1x 5X GC rich solution, 0.8µM forward primer, 0.8µM reverse primer and 0.5U FastStart Taq DNA polymerase (Roche; Cat.No.12158264001) and made up to a final volume of 30µl with distilled water. Forward and reverse primers were designed to anneal to the pGEM backbone; forward primer: 5'-TAATACGACTCACTATAGGG-3; and reverse primer: 5'-

ACACTATAGAATACTCAAGC-3'. Single-colony PCR reactions were carried out using a touchdown PCR as follows; 5 minutes at 95˚C, 3 cycles of [30 seconds at 95˚C, 30 seconds at 60˚C and 30 seconds at 72˚C], 3 cycles of [30 seconds at 95˚C, 30 seconds at 58˚C and 30 seconds at 72˚C], 34 cycles of [30 seconds at 95˚C, 30 seconds at 56˚C and 30 seconds at 72˚C] and finally 10 minutes at 72˚C.

2.6.5 PCR product clean-up for sequencing

Prior to sequencing, single colony PCR products were cleaned-up using a combination of (a) FastAP thermosensitive alkaline phosphatase (Fermentas; Cat. No: EF0651) to remove excess dNTPs, and (b), ExonucleaseI (New England Biolabs; Cat. No: M0293S) to remove excess single stranded primer. A reaction was set up containing 10µl PCR product, 1U FastAP, 0.01U ExoI and 1x 10x FastAP buffer (100mM Tris-HCL (pH8), 50mM magnesium chloride, 1M potassium chloride, 0.02% Triton X-100 and 1mg/ml bovine serum albumin (BSA)). Samples were incubated for 30 minutes at 37˚C followed by an enzyme inactivation step for 20 minutes at 80˚C.

2.6.6 Sequencing reaction

Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems; Cat. No: 4336917) was used for all sequencing. Following PCR clean up (see section 2.6.5) 5µl PCR product was added to a reaction of 20 pmol forward or reverse primer, 0.5µl Big Dye, 2µl 5X buffer and made up to a total volume of 10µl with distilled water. The sequencing reaction was then carried out using the following program: 94˚C for four minutes followed by 35 cycles of [94 ˚C for 25 seconds, 50˚C for 25 seconds and 60˚C for four minutes]. Samples were wrapped in foil following the sequencing reaction and stored at -20˚C until precipitated and prepared to be read on the sequencer.

2.6.7 Ethanol precipitation following sequencing reaction

Following the sequencing reaction, 3.5µl precipitation buffer (1.5M sodium acetate, 1.5M EDTA) and 100µl of 95% (v/v) ethanol were added to each sample prior to centrifugation at 2,254 x g, 4˚C for 30 minutes. The supernatant was carefully removed and the samples were then briefly pulsed at 23 x g upside down before washing with 200µl 70% (v/v) ethanol and a further centrifugation step at 2,254 x g, 4˚C for 30 minutes. The supernatant was again carefully removed and samples pulsed at 23 x g upside down before a final wash with 200µl 70% (v/v) ethanol and centrifugation at 2,254 x g, 4˚C for 30 minutes. The supernatant was removed and samples pulsed at 23 x g upside down before being allowed to air dry.

To prepare precipitated samples for sequencing, pellets were resuspended in 10µl Hi-Di Formamide (Applied Biosystems; Cat. No: 4311320) and denatured at 95˚C for 5 minutes. Samples were run on a 3730 DNA Analyzer (Applied Biosystems) and output files analysed using sequencing analysis 5.2 (Applied Biosystems).

2.7 GENE EXPRESSION ANALYSIS

2.7.1 DNase treatment of RNA

For the analysis of single exon genes (such as *RASSF10*, chapter 5), RNA was DNase treated before cDNA synthesis to remove any DNA contaminant. DNase treatment was achieved using DNase I (Fermentas; Cat.No. EN0521). For 1µg total RNA, a 10µl reaction was set up with 1x 10xbuffer (10mM Tris-HCL, (pH 7.5), 25mM magnesium chloride and 1mM calcium chloride) and 1U DNase I, made up to 10µl with RNase free water. Samples were incubated at

37˚C for 30 minutes before the reaction was terminated by the addition of 1µl 25mM EDTA and subsequent incubation at 65˚C for 10 minutes.

2.7.2 cDNA synthesis

cDNA was synthesised from 1µg total RNA using Superscript III (Invitrogen; Cat. No. 18080-093) and random hexamer primers (Fermentas; Cat. No. S0142). An initial reaction was set up containing 1µg total RNA, 1µl of 100µM random hexamers and 1µl 10mM dNTP mix, made up to 13µl with RNase free water. Samples were incubated at 65˚C for 5 minutes. Samples were then transferred to ice for 1 minute prior to the addition of 4µl 5X First-Strand Buffer (250mM Tris-HCL (pH 8.3), 375mM potassium chloride, 15mM magnesium chloride) 1µl 0.1M Dithiothreitol (DTT), 1µl (40U) RNaseOUT™ and 1µl (200U) Superscript™ III RT. Samples were then incubated for 5 minutes at 25˚C followed by 60 minutes at 50˚C. The reaction was inactivated by incubating at 70˚C for 15 minutes. cDNA was stored at -20˚C.

2.7.3 Reverse transcript polymerase chain reaction (RT-PCR) primer design

For multi exon genes, primers for expression analysis were designed to encompass one or more intronic regions to produce a product too large to amplify using conventional PCR methods and therefore ensuring successful amplification from cDNA and not any contaminant DNA from the RNA extraction process. Since cDNA is synthesised from the 3' end, primers were preferentially designed to be located towards the 3' end to ensure amplification of the best quality template cDNA. In most cases, primers were designed using the freely available Primer 3 program (http://frodo.wi.mit.edu/primer3/) using settings to design primers between 250-400bp and not mis-match against repeats within the human genome. In some cases, primers were designed manually observing the following guidelines; making primers 20-25bp long, not having long strings of complimentary bases (i.e. As with Ts or Gs with Cs) to avoid

hairpin loops, an overall 50% GC content and a 3' end containing either CC, CG, GG or GC. Primer sequences were then compared against the human genome using the NCBI BLAST program to check for specificity. Expression primer sequences are shown in appendices B.II, C.II, D.II and E.III.

2.7.4 RT-PCR

RT-PCR reactions were set up using 1µl prepared cDNA, 2.5µl 10X buffer (500mM Tris-HCL (pH 8.3), 100Mm potassium chloride, 50mM ammonium sulphate), 2mM magnesium chloride, 2.5mM dNTP mix, 1x 5X GC rich solution, 0.8µM forward primer, 0.8µM reverse primer and 0.5u FastStart Taq DNA polymerase (Roche; Cat.No.12158264001) and made up to a final volume of 25µl with distilled water. RT-PCR reactions were carried out using a touchdown PCR as follows; 5 minutes at 95˚C, 3 cycles of [30 seconds at 95˚C, 30 seconds at 60˚C and 30 seconds at 72˚C], 3 cycles of [30 seconds at 95˚C, 30 seconds at 58˚C and 30 seconds at 72˚C], 34 cycles of [30 seconds at 95˚C, 30 seconds at 56˚C and 30 seconds at 72˚C] and finally 10 minutes at 72˚C. GAPDH RT-PCR was carried out using the same program with the exception of a lower total cycle number (28 cycles).

2.7.5 Real-time quantitative expression PCR

Real-time RT-PCR was carried out for selected genes to assess gene expression in primary tumours. This work was carried out by Sophie Vacher in the laboratory of Dr. Ivan Bièche (Oncogenetic Laboratory, INSERM U 735, Centre René Huguenin, Saint Cloud, France). Real-time PCR used a SYBR green based assay described in Tozlu-Kara *et al*, 2007. Briefly, cDNA was synthesised from primary breast tumour and breast normal tissue RNA and subsequently used in the real time assay. Since this is a quantitative assay and it can be difficult to accurately assess the amount and quality of RNA/cDNA being used as starting

material all results were normalised on the basis of an endogenous RNA control, in this case, the *TATA box-binding protein (TBP)* gene. Measurements for the target gene were calculated from the initial detection of *TBP* until the end of the reaction to produce a result of N-fold difference between target gene and *TBP*. N-fold differences were determined by 2ΔCt sample where ΔCt sample is the equivalent of subtracting the mean ΔCt of the target gene from the mean ΔCt of *TBP*. ΔCt is the difference in parameter threshold cycle. The SYBR green PCR Core Reagents Kit (Perkin Elmer, Applied Biosystems. Cat. No:4304886) was used for the assay and PCR conditions consisted of 95˚C incubation for 10 minutes followed by 50 cycles of [95˚C for 15 seconds and 65˚C for 1 minute].

2.8 TISSUE CULTURE

Tissue culture work was carried out in class II hoods using a sterile technique at all times. All equipment was wiped down with 70% (v/v) industrial methylated spirits (IMS) before use in the hood. Since cell pellets and/or nucleic acid preparations from leukaemia lines were already available, only adherent cells were grown for this study.

2.8.1 Preparation of growth media

All solid tumour lines grown for this study were maintained in Dulbecco's Modified Eagles Media (DMEM) (Sigma-Aldrich; Cat. No: D5671) supplemented with 10% (v/v) fetal bovine serum (FBS) (PAA Laboratories; Cat. No: A15-151), 2mM L-glutamine (Sigma Aldrich, Cat. No: G7513), penicillin (100u/ml) and streptomycin (100µg/ml) (Sigma Aldrich; Cat. No: P4333).

2.8.2 Resurrection of cell lines

Cell line frozen stocks were removed from liquid nitrogen and placed into a 37˚C water bath. Cells were removed from the water bath as soon as they were defrosted and carefully transferred to 10mls pre-warmed (37˚C) culture media. To remove any cryopreservant from the media, cells were then centrifuged at room temperature for 4 minutes, 280 x g. Supernatant was carefully removed and cells very carefully resuspended in 1ml culture media prior to transfer to a 25cm^2 tissue culture flask and the addition of a further 7mls growth media. Freshly revived cells were maintained at 37˚C, 5% carbon dioxide in a 25cm^2 flask until ready for transfer to a 75cm^2 flask.

2.8.3 Maintenance and passaging of cell lines

Cell lines were fed every two days by removing half the media volume and replacing with fresh pre-warmed media. Once the cell lines reached approximately 70%-90% confluency they were passaged into ratios of 1:3, 1:5 or 1:10 depending on the growth rate of the cell line. After the removal of media, cell lines were washed with 4mls pre-warmed phosphate buffered saline (PBS), followed by the addition of 3mls pre-warmed 1x 0.25% Trypsin-EDTA (Invitrogen; Cat.No: 25200056). Cells were then transferred back to 37˚C until the cells were dislodged from the flask surface. Cells were resuspended in 7mls media and split accordingly, with the desired volume of cells being returned to the flask and the unwanted cells discarded. Media was added to the cells to make up to 15mls and cells returned to the incubator.

2.8.4 Counting of cells using a haemocytometer

Cells were trypsinised as described in section 2.8.3 before being resuspended in 7mls fresh media prior to counting. The haemocytometer chamber and cover slip were first cleaned thoroughly with 70% (v/v) IMS before the cover slip was placed firmly onto the

haemocytometer chamber and an aliquot of 70µl cells was applied to each side of the haemocytometer underneath the coverslip. The cells in each side were counted under a microscope and an average calculated. Each millilitre of media contains $1x10^4$ of this average count.

2.8.5 Treatment of cell lines with 5-aza-2'deoxycitidine

5-aza-2'deoxycitidne (5-azaDC) was used to demethylate cell line genomic DNA to enable expression analysis of selected genes with and without methylation. Treatments were performed by myself and a laboratory technician, Dean Gentle. DNA demethylation by 5-azaDC only occurs during DNA replication, therefore cells need to be dividing for effective 5-azaDC treatment. For this reason, cell lines were grown to only 30-50% confluency for the start of 5-azaDC treatment; however this was dependent on particular cell line growth rates. Media supplemented with 5µM 5-azaDC (Sigma-Adrich; Cat. No: A3656) was added to cell lines and changed daily for five days. Cell pellets were then produced as described in section 2.8.6 and stored for later nucleic acid extraction. A non-treated control for each cell line was grown concurrently, pelleted and stored for later extraction.

2.8.6 Producing cell pellets

Cell line pellets were prepared for extraction of nucleic acids. Once cells reached 90% confluency or the end of their 5-azaDC treatment, they were washed with PBS and detached from the flask using 1x trypsin-EDTA as described in section 2.8.3. Cells were resuspended in culture media and transferred to a 15ml falcon prior to centrifugation at 1000 x g, at 4° for 4 minutes. Supernatant was removed and the cell pellet washed with 1ml PBS followed by a second centrifugation step as before and removal of the supernatant before being snap frozen in liquid nitrogen. Pellets were quickly transferred to -80˚C for storage until required.

2.8.7 Producing frozen aliquots

Cells were first trypsinised, diluted to 10ml and counted as described in sections 2.8.3 and 2.8.4. The cells were then transferred to a 15ml falcon, centrifuged at 1000 x g, at 4° for 4 minutes, after which all the supernatant was removed. The cell pellet was resuspended in standard DMEM media containing an additional 10% (v/v) dimethyl sulphoxide (DMSO), to act as a cryopreservant during storage, to a final concentration of $1x10^6$ cells per ml. One millilitre was then transferred to a cryo-vial and put into a cryopreservation pot to ensure freezing occurred at -1˚C per minute. The cryopreservation pot was then transferred to -80˚C overnight before transfer of the vials to liquid nitrogen for long term storage.

2.9 ARRAY ANALYSIS

This study used two genome wide methylation assays/arrays; (a) methylated CpG island recovery assay (MIRA) coupled with a CpG island array (chapter 3) and (b) Illumina Infinium HumanMethylation27 BeadChip array (chapter 4).

2.9.1 MIRA assay

The MIRA assay and subsequent array were carried out by Dr. Stella Tommasi from Professor Gerd P. Pfeifer's laboratory (Beckman Research Institute, City of Hope, 1500 E Duarte Road, Duarte, CA 91010. USA). Detailed descriptions of the MIRA assay can be found in Rauch *et al*, 2006 and Rauch *et al*, 2007. Briefly, with modifications as described in Tommasi *et al*, 2009, the MIRA procedure was carried out as follows: genomic DNA was double-digested with MseI (5'-TTAA-3') and HhaI (5'-GCGC-3') to produce small fragments allowing for collection of heavily methylated CpG island DNA and subsequently incubated overnight at 4˚C with a 1:1 mixture of MBD2b (GST-tagged) and MBD3L1 (His tagged) pre-bound to glutathione beads. Following the incubation, pelleted glutathione beads were washed

with high salt buffer three times. Guanidium hydrochloride containing buffer was then used to elute methylated DNA enriched genomic DNA. DNA was then purified using QIAquick PCR purification kit (Qiagen), according to manufacturers' instructions. Ligation of purified DNA fragments to MseI linkers was done prior to PCR amplification. Products were then labelled with either Cy5-dCTP (breast cancer cell lines) or Cy3-dCTP (NHMEC lines). Samples were then mixed and hybridized to a human CpG island microarray (Agilent Technologies). Raw data was provided in the form of ratio values of Cy5:Cy3 for each probe on the CpG island microarray.

2.9.2 Illumina Infinium HumanMethylation27 BeadChip array

The Illumina Infinium array was carried out by Ghazala Mirza at the Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, OX3 7BN, UK. The study utilised the HumanMethylation27 BeadChip array. DNA was first bisulphite modified and then prepared and hybridized to the chip according to manufacturers' instructions. Raw data was initially run through BeadStudio (Illumina) and provided as β-values that represented methylation levels for each probe on the array ranging from 0 (designating no methylation) to 1 (Designating complete methylation).

2.10 STATISTICAL METHODS

2.10.1 Fisher's exact test

Fisher's exact test was used to determine the significance of the frequency of methylated samples in a particular group. In single cases, GraphPad (http://graphpad.com) was used to determine significance and in multiple cases, as in chapter 4, a Macro was installed into Microsoft Excel. Both programs used the same Fisher's exact test equation. In every case of

using Fisher's exact test, $p<0.05$ was considered significant. However, when multiple tests were carried out p-values were subject to multiple correction, see section 2.10.5.

2.10.2 Student's t-test

When testing the significance of two means, student's t-test was utilised. Student's t-tests were carried out in Microsoft Excel and in all cases were two-tailed, two sample unequal variance. $P<0.05$ was considered significant.

2.10.3 ANOVA

One-way analysis of variance (ANOVA) analysis, carried out using MedCalc (http://medcalc.be), was used to assess significance between more than one mean. Significance was taken as $p<0.05$

2.10.4 Kaplan-Meier analysis

To determine whether DNA methylation of a particular gene associated with survival outcome, Kaplan-Meier graphs were produced using MedCalc (http://medcalc.be), as with other statistical test, $p<0.05$ was considered significant.

2.10.5 Multiple correction

Statistical significance was taken at $p<0.05$, however, when conducting multiple tests on the same group of data, the occurrence of $p<0.05$ becomes 1 in 20 by chance. Therefore, when carrying out numerous tests, as in chapter 4, the occurrence of false positives can become quite high. Either Bonferroni correction or False Discovery Rate (FDR) corrections were used in this thesis.

Bonferroni correction

Bonferroni correction involves adjusting the significance value by the number of tests carried out. For example, the significant value becomes p/n where p is the p-value considered significant and n is the number of tests carried out. Therefore, if carrying out 10 tests, a significance value of 0.05 becomes 0.005. Bonferroni correction is highly stringent and was used in this thesis on small datasets where it was most applicable.

False discovery rate (FDR)

This method is more suitable for large datasets than Bonferroni correction and is commonly used in microarray experiments. FDR correction works on a ranking basis, unadjusted p-values are ranked smallest to largest and then each p-value is multiplied by its rank within the list i.e. the third most significant p-value (X) would become 3 x X. Statistical significance remained $p<0.05$ once p-values had been adjusted.

2.11 BIOINFORMATICS

2.11.1 Gathering gene structure information

General gene structure information was obtained from the UCSC genome browser (http://genome.ucsc.edu/) using the most up to date human genome reference (either hg18 or hg19 for the work in this thesis).

CpG island information was gathered from UCSC genome browser (http://genome.ucsc.edu/) and the European Bioinformatics Institute (EBI) CpG plot program (http://www.ebi.ac.uk/Tools/emboss/cpgplot/).

2.11.2 Functional annotation of gene lists

The freely available Database for Annotation, Visualization and Integrated Discovery (DAVID) database was used for analysing large lists of genes to mine for functionally related groups of genes and genes within pathways.

2.11.3 Individual gene information

The National Centre for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/) and GeneCards (http://www.genecards.org/) were used to gather information on the functions, published work and general information for individual genes.

2.11.4 Gene mutation and methylation information

The Catalogue Of Somatic Mutations In Cancer (COSMIC) website (http://www.sanger.ac.uk/genetics/CGP/cosmic/) was used to partially assess the known levels of mutation in selected individual genes. While, the PubMeth: Reviewed methylation database in cancer website (http://www.pubmeth.org/) was used to partially assess the known levels of methylation in selected individual genes. Further assessment for both these aspects was performed by specific literature searches using the NCBI PubMed website (http://www.ncbi.nlm.nih.gov/pubmed/).

CHAPTER 3

USE OF THE METHYLATED CpG ISLAND RECOVERY ASSAY (MIRA) COUPLED WITH HUMAN CpG ISLAND ARRAYS TO IDENTIFY METHYLATED GENES IN BREAST CANCER

3.1 ABSTRACT

This study has used the methylated CpG island recovery assay (MIRA) coupled with human CpG island arrays on five breast cancer cell lines and two normal human mammary epithelial cell lines (NHMECs) to identify a large number of potentially methylated targets in breast cancer. Confirmation of 32 of these targets by combined bisulphite restriction analysis (CoBRA) identified 16 frequently methylated genes in breast cancer cell lines.

Analysis of 16 targets in sporadic breast cancer patient samples identified nine genes that showed frequent methylation. Further analysis confirmed five of these genes, *CIDE-A*, *DBC1*, *EMILIN2*, *FBLN2* and *SALL1* show tumour specific methylation and can be re-expressed in methylated cell lines following treatment with the demethylating agent, 5-azaDC. In addition, mRNA levels of *EMILIN2* in methylated patient samples were significantly lower than unmethylated samples ($p = 0.018$). This study has also demonstrated varying levels of methylation of the above five genes in other solid tumours. The most frequent being *FBLN2* in glioma tumours (45%).

Association of methylation with breast cancer clinical features demonstrated *FBLN2* methylation associated with PR positive tumours ($p = 0.013$) and *CIDE-A* methylation associated with ER positive tumours ($p = 0.016$). *EMILIN2* methylation associated with worse

disease free survival ($p = 0.041$), relapse ($p = 0.031$) and ER and PR receptor status ($p = 0.0009$ and $p = 0.0082$ respectively).

3.2 INTRODUCTION

Chapter one demonstrated that although frequently mutated genes in familial breast cancer, such as *BRCA1*, *BRCA2*, *PTEN*, *BRIP1*, and *CHEK2* have been identified (Hall *et al*, 1990; Phelan *et al*, 1996; Liaw *et al*, 1997; Rutter *et al*, 2003; Meijers-Heijboer *et al*, 2002), very few sporadic cases have been shown to have such alterations. It is likely to be a build up of many factors that contribute to sporadic cases, such as mutations, chromosomal alterations and downregulation or upregulation of genes by other mechanisms. Recent studies have demonstrated the large number of alterations per breast cancer genome illustrating the magnitude of molecular changes required for sporadic tumourigenesis (Stephens *et al*, 2010; Ding *et al*, 2010; Shah *et al*¸ 2009; Sjöblom *et al*, 2006; Wood *et al*¸ 2007). In addition to genetic events, epigenetic events such as histone modification and DNA methylation could also contribute to these molecular changes that initiate and enable tumourigenesis. DNA hypermethylation events have been shown to be common in many cancers, including breast, both through the inactivation of genes with tumour suppressor properties such as *RASSF1A*, *CCND2*, *VHL*, *PTEN* and *APC* (Dammann *et al*, 2000; Evron *et al*, 2001; Herman *et al*, 1994; Goel *et al*, 2004; Esteller *et al*, 2000c), the identification of methylation profiles in relation to breast cancer clinical features (Christensen *et al*¸ 2010; Bediaga *et al*, 2010; Holm *et al*¸2010) and the identification of methylation markers of disease progression (Hartmann *et al*, 2009). The determination of hypermethylation events in the sporadic breast cancer genome can be of potential use in further understanding of the disease and identifying useful markers that can aid patient diagnosis, treatment plans and prognosis. Recent advances in genome-wide methods and isolation of methylated CpG islands have resulted in new technologies becoming available to identify such genes through analysis of large swathes of the genome.

The work in this chapter uses one of these methods, the methylated CpG island recovery assay (MIRA) in conjunction with a human CpG island microarray to identify methylated genes in breast cancer cell lines with the aim of using this information to identify genes that are frequently methylated in sporadic breast cancer patient samples that may relate to clinical features or provide potential in-roads to better disease understanding.

3.3 AIMS

1. To use the MIRA assay (methylated CpG island recovery assay) and human CpG island microarrays to produce a list of hypermethylated genes in breast cancer cell lines

2. To analyse selected candidate genes in the assayed breast cancer cell lines plus additional breast cancer cell lines and further confirm positively methylated candidates in sporadic ductal breast tumour samples

3. To assess expression status of candidate genes pre and post treatment with 5-azaDC

4. To assess tumour specificity in a further sporadic ductal breast cancer cohort with matched normal breast tissue

5. To assess any association between positive methylation and clinical features in the sporadic ductal breast tumour cohort

6. To determine the methylation status of positive candidate genes in other common solid tumour types

3.4 RESULTS

3.4.1 Array analysis

This study used the methylated CpG island recovery assay (MIRA) coupled with Agilent Human CpG island microarrays on five breast cancer cell lines (HCC1806, HCC1419, HCC1395, HCC1143 and HCC1937) and two normal human mammary epithelial cell (NHMEC) lines to assess methylation changes across 27,800 CpG islands. See section 1.3.7.2 for more details on the MIRA assay.

To isolate hypermethylated probes, methylation enrichment factors were calculated for all probes and then screened so that only genes demonstrating a >5-fold increase in the MIRA-enriched fraction of tumour vs. normal in a single probe in 3 or more cell lines were retained. This resulted in 14,875 probes, representing 6,912 individual gene loci. From this were removed entries that did not represent a known, annotated gene opposed to a predicted mRNA, non-coding mRNA or genomic region (ORFs, chromosomal regions, miRNAs, LOCs, MGCs, FLJs, KIAs and Agilent plate related hits) and genes located on the X chromosome due to their intrinsic levels of methylation present in the inactivated X chromosome (all cell lines were female in origin as were subsequently analysed patient samples) and the confusion caused by loss of part or all of the active X chromosome, resulting in a list of 10,840 probes (3,347 genes). This low stringency list was further refined by the removal of all probes that were not located within a promoter or divergent promoter region, resulting in 3352 probes (1268 genes). Due to this large number of genes, a second, more stringent list was created, consisting only of probes that demonstrated a >10-fold increase in 4 or more cell lines, resulting in 895 probes (276 genes) (figure 3.1). The list of 276 genes from the high stringency list is provided in appendix A.III.

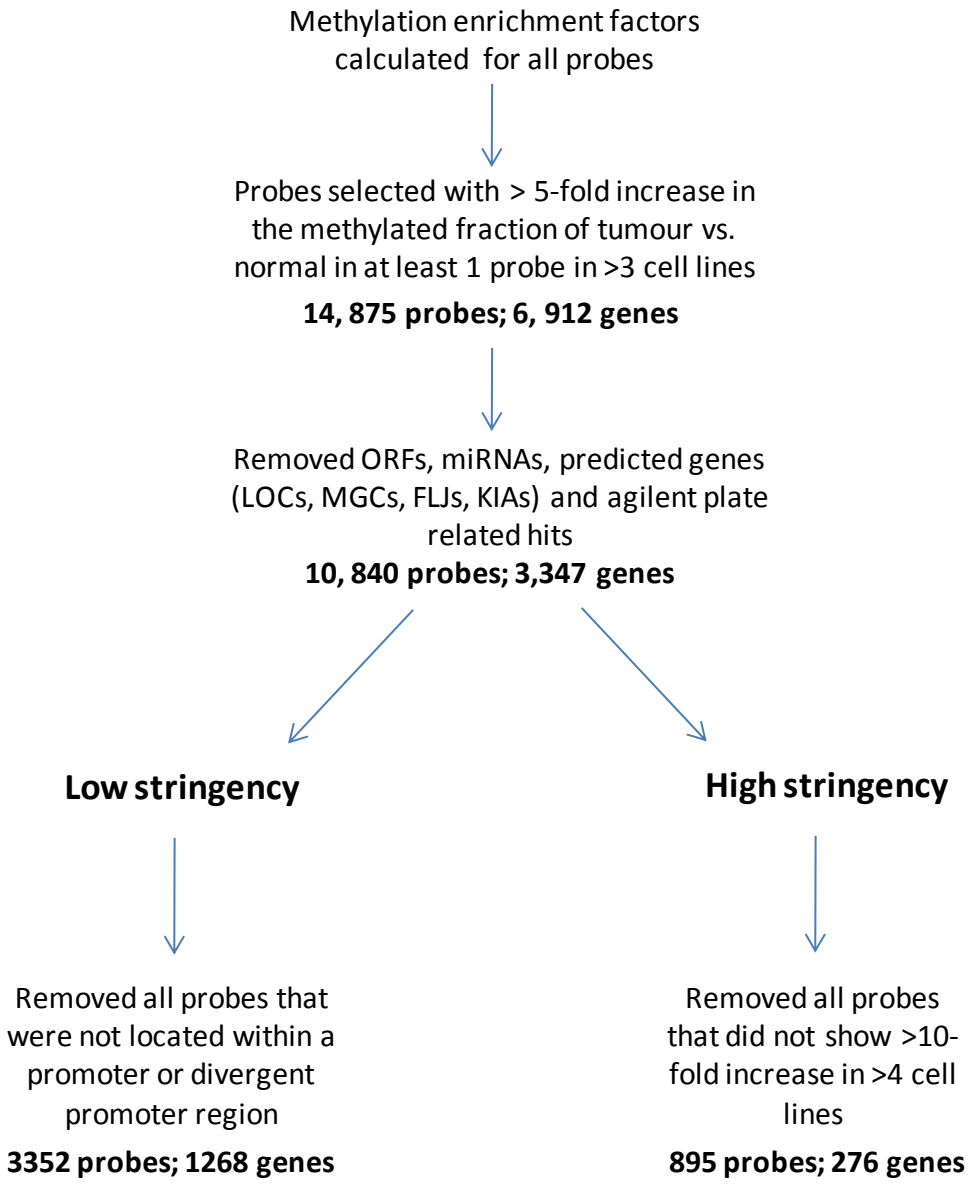


Figure 3.1 Flow-through of MIRA list analysis
This diagram illustrates the method used to generate a working list of candidate genes from the raw MIRA array data.

From these lists, target genes were chosen to be analysed further by the use of DAVID (Database for Annotation, Visualization and Integrated Discovery) functional analysis tool, literature searching, SAGE (serial analysis of gene expression) expression data and oncomine gene expression data. In total, 32 genes were selected for further analysis: *ABT1, BARHL2, BUB3, CCNJ, CD44, CDH24, CENPJ, CIDE-A, COMP, DBC1, DDEF2, DLGAP1, EMILIN2, EPSTI1, FBLN2, FOXF2, FOXQ1, GSC, KLF11, KLF13, LYPD5, MADL1L1, NRXN2, ONECUT1, PAX9, PHF2, POU4F1, SALL1, SALL3, SESN3, SIM2* and *TP53INP1* (table 3.1).

3.4.2 Analysis of selected genes in breast cancer cell lines

Targets identified from the array were analysed by CoBRA in the cancer cell lines used for the array, additional breast cancer lines and a normal breast DNA sample. This was done to confirm the frequent methylation in a wider range of breast cancer cell lines and the absence of methylation in normal breast tissue DNA. The aim of the array was to identify genes where methylation resulted in downregulation of the gene, therefore, where possible, primers were situated around the gene transcription start point regardless of array probe position as this is the most likely region for hypermethylation to result in gene inactivation and therefore have biological relevance.

In total, 32 genes were analysed in cell lines. Of these, 16 (50%) showed methylation in greater than 30% of samples and no methylation in the normal breast tissue DNA sample (*CD44, CIDE-A, COMP, DBC1, EMILIN2, EPSTI1, FBLN2, FOXF2, GSC, NRXN2, PAX9, POU4F1, SALL1, SESN3, SIM2, TP53INP1*) (table 3.2 and figure 3.2). Of the remaining 16 genes, 11 (34.4%) did not show methylation in the region analysed by CoBRA (*ABT1, BUB3, CCNJ, CDH24, CENPJ, DDEF2, KLF11, KLF13, MAD1L1,*

GENE SYMBOL	GENE NAME	RELEVANT BACKGROUND
ABT1	ACTIVATOR OF BASAL TRANSCRIPTION 1	Novel Target - Down regulated in MDA-MB-231 Breast cell line
BARHL2	BARH-LIKE 2 (DROSOPHILA)	Novel Target - Homeobox Gene
BUB3	BUB3 BUDDING UNINHIBITED BY BENZIMIDAZOLES 3 HOMOLOG (YEAST)	Involved in control of mitotic checkpoint
CCNJ	CYCLIN J	Regulator of Cell Cycle Progression
CD44	CD44 ANTIGEN (INDIAN BLOOD GROUP)	Methylated in Prostate Cancer, Connection to Breast cancer
CDH24	CADHERIN-LIKE 24	Under-expressed in Breast cancer (SAGE), Mutated in AML, Involved in Cellular Adhesion
CENPJ	CENTROMERE PROTEIN J	Microtubule cytoskeleton organization
CIDEA	CELL DEATH-INDUCING DFFA-LIKE EFFECTOR A	Involved in Apoptosis, Under-expressed in Breast cancer (SAGE)
COMP	CARTILAGE OLIGOMERIC MATRIX PROTEIN	Involved in TGF-beta signalling, cell adhesion and communication
DBC1	DELETED IN BLADDER CANCER 1	Associated with Bladder cancer, methylated in NSCLC
DDEF2	DEVELOPMENT AND DIFFERENTIATION ENHANCING FACTOR 2	Involved in differentiation, Regulation related to Oestrogen Receptor
DLGAP1	DISCS, LARGE (DROSOPHILA) HOMOLOG-ASSOCIATED PROTEIN 1	Novel Target
EMILIN2	ELASTIN MICROFIBRIL INTERFACER 2	Re-expressed after demethylation in ovarian cancer, Slightly under-expressed in Breast cancer (SAGE)
EPSTI1	EPITHELIAL STROMAL INTERACTION 1 (BREAST)	Differentially expressed in Breast tissue, Under-expressed in Breast cancer (SAGE)
FBLN2	FIBULIN 2	In-House candidate gene
FOXF2	FORKHEAD BOX F2	Involved in cell morphogenesis and blood vessel development
FOXQ1	FORKHEAD BOX Q1	Novel Target - Transcription factor
GSC	GOOSECOID	Implicated in lung cancer (Potential hit on another MIRA array)
KLF11	KRUPPEL-LIKE FACTOR 11	Negative regulator of cell cycle progression
KLF13	KRUPPEL-LIKE FACTOR 13	Suppresses K-RAS by down regulated of Cyclin B
LYPD5	LY6/PLAUR DOMAIN CONTAINING 5	Novel Target - Slightly under-expressed in Breast cancer (SAGE)
MAD1L1	MAD1 MITOTIC ARREST DEFICIENT-LIKE 1 (YEAST)	Mitotic spindle assembly checkpoint protein ,Mutated in multiple cancers
NRXN2	NEUREXIN 2	Associated with Prostate cancer, Cell Adhesion, under-expressed in Breast cancer (SAGE)
ONECUT1	ONE CUT DOMAIN, FAMILY MEMBER 1	Methylated in Cervical cancer,
PAX9	PAIRED BOX GENE 9	Involved in Development, Implicated in lung cancer (Potential hit on another MIRA array)
PHF2	PHD FINGER PROTEIN 2	Novel Target
POU4F1	POU DOMAIN, CLASS 4, TRANSCRIPTION FACTOR 1	Expression enhanced when BRCA1 reduced, Implicated in lung cancer (Potential hit on another MIRA array)
SALL1	SAL-LIKE 1 (DROSOPHILA)	Methylated in Leukaemia, Because SALL3 was also a hit
SALL3	SAL-LIKE 3 (DROSOPHILA)	Because SALL1 was also a hit
SESN3	SESTRIN 3	Methylated in Endometrial cancer, Induced by p53
SIM2	SINGLE-MINDED HOMOLOG 2 (DROSOPHILA)	Potential Tumour Suppressor Gene
TP53INP1	TUMOR PROTEIN P53 INDUCIBLE NUCLEAR PROTEIN 1	Involved in Apoptosis, Interacts with p53

Table 3.1_Reason for selecting the 32 genes for analysis in cell lines

Each of the 32 gene symbol and names are given along with the reason they were chosen to be analysed further.

GENE NAME	GENE PRODUCT	CELL LINE	NORMAL BREAST
SESN3	sestrin 3	6/6	0/1
SIM2	single-minded homolog 2 (Drosophila)	7/7	0/1
CIDE-A	cell death-inducing DFFA-like effector a	7/8	0/1
EPSTI1	epithelial stromal interaction 1 (breast)	5/6	0/1
PAX9	paired box 9	4/5	0/1
POU4F1	POU class 4 homeobox 1	6/8	0/1
COMP	cartilage oligomeric matrix protein	4/6	0/1
FOXF2	forkhead box F2	6/9	0/1
SALL1	sal-like 1 (Drosophila)	4/6	0/1
FBLN2	fibulin 2	3/5	0/1
EMILIN2	elastin microfibril interfacer 2	5/9	0/1
CD44	CD44 molecule	3/6	0/1
TP53INP1	tumor protein p53 inducible nuclear protein 1	3/6	0/1
DBC1	deleted in bladder cancer 1	2/6	0/1
GSC	goosecoid homeobox	3/9	0/1
NRXN2	neurexin 2	2/6	0/1
BARHL2	BarH-like homeobox 2	2/8	0/1
FOXQ1	forkhead box Q1	1/6	0/1
DLGAP1	discs, large (Drosophila) homolog-associated protein 1	5/5	1/1
SALL3	sal-like 3 (Drosophila)	3/4	1/1
LYPD5	LY6/PLAUR domain containing 5	3/6	1/1
ABT1	activator of basal transcription 1	0/5	0/1
BUB3	budding uninhibited by benzimidazoles 3 homolog (yeast)	0/9	0/1
CCNJ	cyclin J	0/6	0/1
CDH24	cadherin 24, type 2	0/9	0/1
CENPJ	centromere protein J	0/9	0/1
DDEF2	development and differentiation enhancing factor 21	0/6	0/1
KLF11	Kruppel-like factor 11	0/9	0/1
KLF13	Kruppel-like factor 13	0/6	0/1
MAD1L1	MAD1 mitotic arrest deficient-like 1 (yeast)	0/6	0/1
ONECUT1	one cut homeobox 1	0/9	0/1
PHF2	PHD finger protein 2	0/6	0/1

Table 3.2_CoBRA results for 32 selected genes in breast cancer cell lines

Gene names and gene products are shown for each of the 32 genes selected for CoBRA analysis in breast cancer cell lines. The methylation frequency by CoBRA is given for each gene along with the result from the normal breast tissue DNA sample. Genes highlighted in dark grey were excluded from further analysis because of methylation in the normal breast tissue DNA, genes highlighted in light grey were excluded from further analysis because no methylation was detected, genes highlighted in light blue were excluded from further analysis because of infrequent methylation and genes highlighted in blue were selected for further analysis.

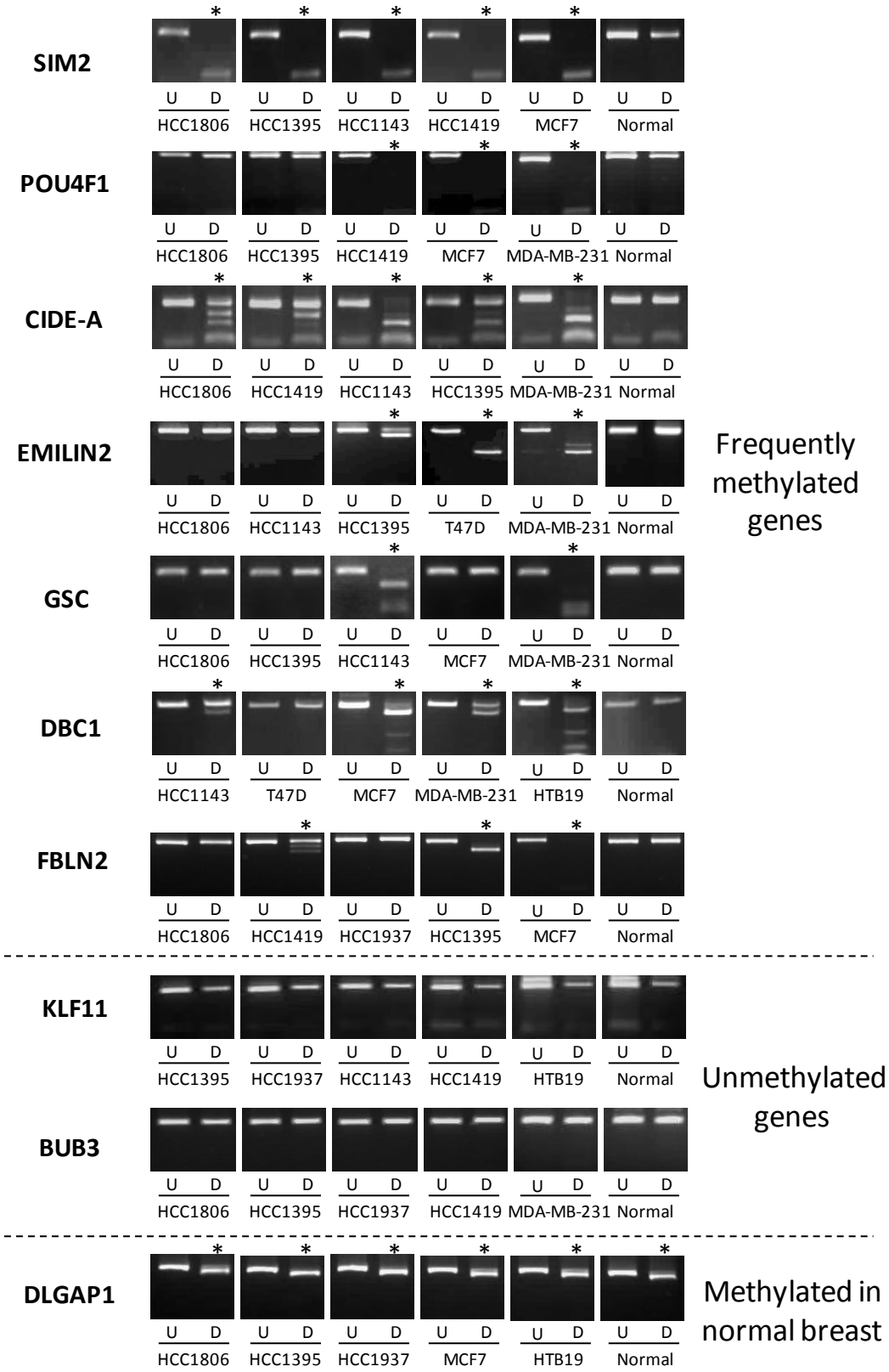

SIM2
U D U D U D U D U D U D
HCC1806 HCC1395 HCC1143 HCC1419 MCF7 Normal
POU4F1
U D U D U D U D U D U D
HCC1806 HCC1395 HCC1419 MCF7 MDA-MB-231 Normal
CIDE-A
U D U D U D U D U D U D
HCC1806 HCC1419 HCC1143 HCC1395 MDA-MB-231 Normal
EMILIN2
U D U D U D U D U D U D
HCC1806 HCC1143 HCC1395 T47D MDA-MB-231 Normal
GSC
U D U D U D U D U D U D
HCC1806 HCC1395 HCC1143 MCF7 MDA-MB-231 Normal
DBC1
U D U D U D U D U D U D
HCC1143 T47D MCF7 MDA-MB-231 HTB19 Normal
FBLN2
U D U D U D U D U D U D
HCC1806 HCC1419 HCC1937 HCC1395 MCF7 Normal
Frequently methylated genes
KLF11
U D U D U D U D U D U D
HCC1395 HCC1937 HCC1143 HCC1419 HTB19 Normal
BUB3
U D U D U D U D U D U D
HCC1806 HCC1395 HCC1937 HCC1419 MDA-MB-231 Normal
Unmethylated genes
DLGAP1
U D U D U D U D U D U D
HCC1806 HCC1395 HCC1937 MCF7 HTB19 Normal
Methylated in normal breast

Figure 3.2 CoBRA results for selected genes in cell lines
CoBRA results are shown for selected genes that demonstrated either frequent methylation, no methylation or additional methylation in normal breast. Undigested products (U) are run next to digested products (D). * indicates methylated samples.

ONECUT1 and *PHF2*), two (6.3%) showed infrequent methylation (<30%) in breast cancer cell lines (*BARHL2* and *FOXQ1*) and three (9.4%) showed methylation in both the cancer cell lines and normal breast tissue DNA (*DLGAP1*, *LYPD5* and *SALL3*) (table 3.2 and figure 3.2). Genes showing frequent methylation in the cell lines were selected for further analysis in primary breast tumour samples (n=16).

3.4.3 Analysis of selected genes in primary breast tumour samples

To determine whether methylation was present in primary tumours and not just cell lines specific, the 16 genes showing frequent methylation in cell lines were analysed by CoBRA in a cohort of 40 ductal breast tumours. Frequent tumour methylation (>25%) was observed in nine of the 16 genes (*CIDE-A*, *COMP*, *DBC1*, *EMILIN2*, *EPSTI1*, *FBLN2*, *SALL1*, *SESN3* and *SIM2*) (table 3.3 and figure 3.3). Methylation frequencies ranged from 26% (*DBC1*) to 63% (*SALL1*). The remaining seven genes were either unmethylated (*TP53INP1*, *NRXN2*) or infrequently (<25%) methylated (*CD44*, *FOXF2*, *GSC*, *PAX9*, *POU4F1*) (table 3.3 and figure 3.3). The nine frequently methylated genes were selected for further analysis.

3.4.4 Expression analysis of selected genes

To assess whether methylation of the nine genes showing frequent methylation in primary breast tumours may have functional relevance, breast cancer cell lines were treated with the demethylating agent 5-aza-2'-deoxycitidine (5-azaDC). Of the nine genes analysed for mRNA expression pre and post 5-azaDC treatment, eight (*CIDE-A*, *COMP*, *DBC1*, *EMILIN2*, *EPSTI1*, *FBLN2*, *SALL1* and *SIM2*) showed either increased expression or complete re-expression in methylated cell lines (figure 3.4). Only *SESN3* did not show either increased/complete re-expression in methylated samples (figure 3.4). *SESN3*

GENE NAME	GENE PRODUCT	PRIMARY TUMOURS
EPSTI1	epithelial stromal interaction 1 (breast)	13/36 (36%)
SIM2	single-minded homolog 2 (Drosophila)	24/29 (83%)
SESN3	sestrin 3	7/10 (70%)
SALL1	sal-like 1 (Drosophila)	25/40 (63%)
COMP	cartilage oligomeric matrix protein	20/33 (61%)
CIDE-A	cell death-inducing DFFA-like effector a	21/40 (53%)
EMILIN2	elastin microfibril interfacer 2	16/36 (44%)
FBLN2	fibulin 2	13/38 (34%)
DBC1	deleted in bladder cancer 1	8/31 (26%)
PAX9	paired box 9	8/39 (21%)
POU4F1	POU class 4 homeobox 1	4/27 (15%)
FOXF2	forkhead box F2	5/35 (14%)
CD44	CD44 molecule	1/24 (4%)
GSC	goosecoid homeobox	1/40 (3%)
NRXN2	neurexin 2	0/12 (0%)
TP53INP1	tumor protein p53 inducible nuclear protein 1	0/18 (0%)

Table 3.3 CoBRA results for selected genes analysed in primary breast tumour samples
CoBRA results are shown for each of the genes analysed in primary tumour samples. Genes highlighted in dark blue showed frequent methylation and have been selected for further analysis. Genes highlighted in light blue showed infrequent methylation in tumours and genes highlighted in grey were unmethylated in primary tumour samples.

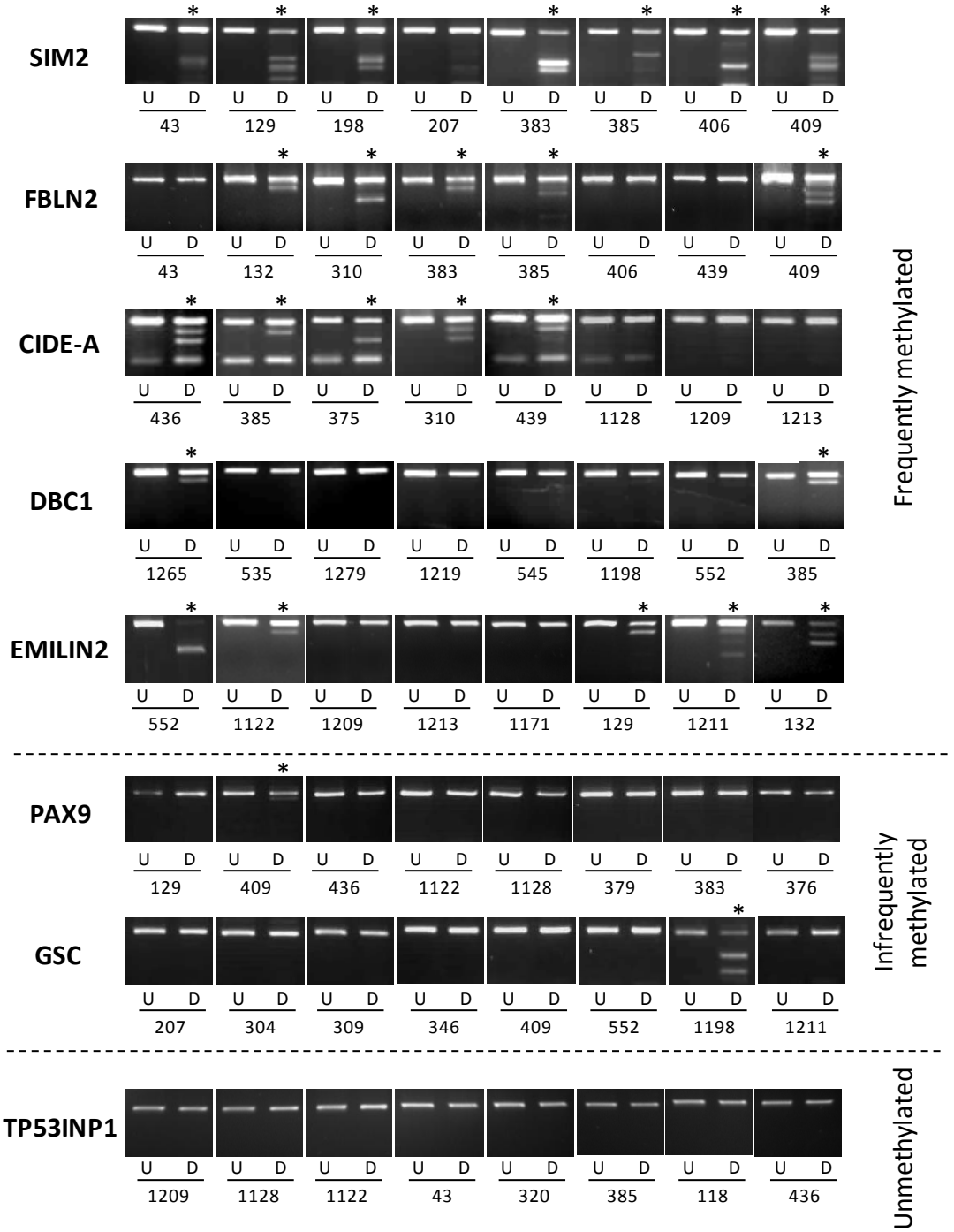


Figure 3.3 CoBRA results for selected genes in primary tumour samples
CoBRA results are shown for eight samples for eight genes representing those classified as frequently methylated, infrequently methylated or unmethylated. Undigested products (U) are shown adjacent to digested products (D). * indicates methylation.

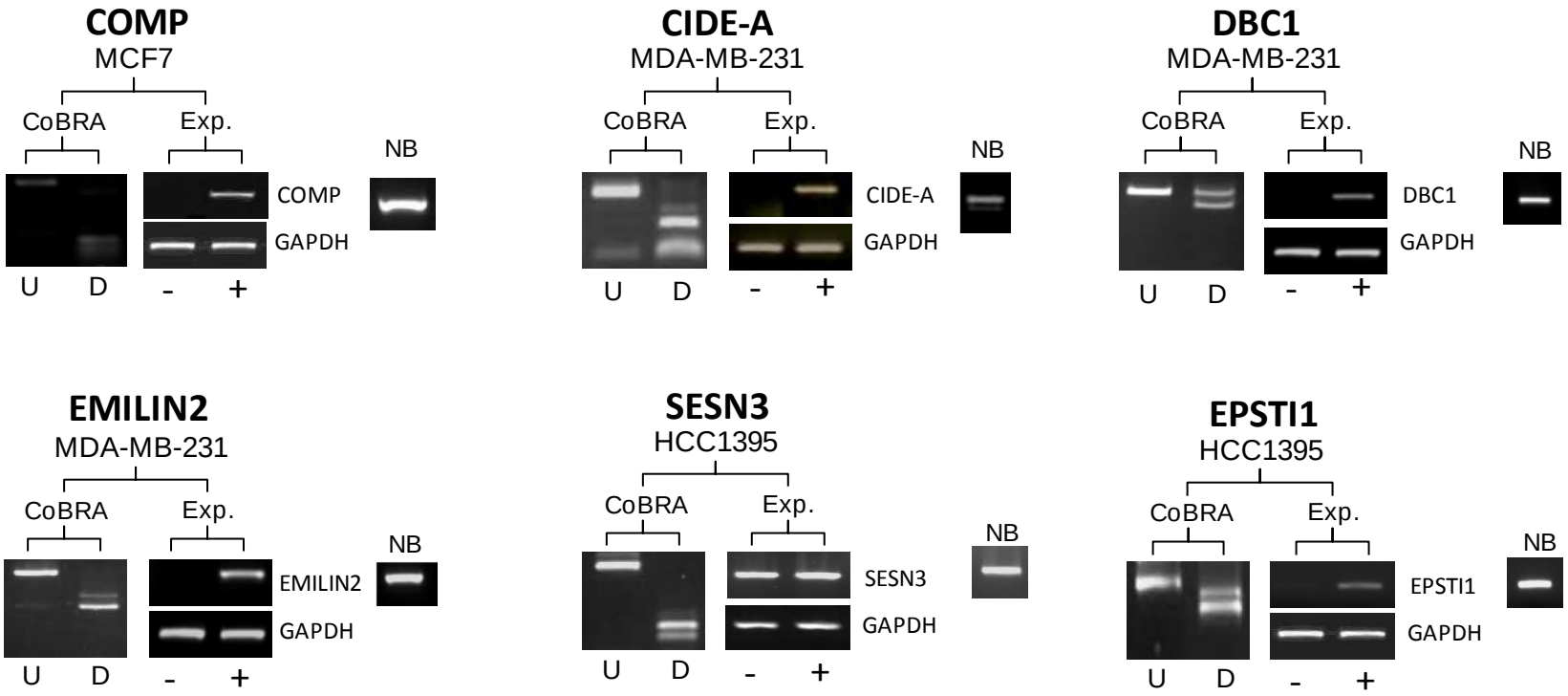


Figure 3.4 Expression analysis of selected genes.
Expression results are shown for *COMP*, *CIDE-A*, *DBC1*, *EMILIN2*, *SESN3* and *EPSTI1*. For each gene expression results are given for the selected gene and *GAPDH* with (+) and without (-) treatment with 5azaDC. Expression in normal breast (NB) is also shown. CoBRA results for the selected cell line is also shown alongside with undigested (U) products run next to digested products (D).

expression analysis failed to show any expression except for two cell lines, both of which were heavily methylated, and expression was observed both pre and post treatment with 5-azaDC, indicating methylation does not influence expression of *SESN3* in breast tissue. Expression was observed in normal breast tissue mRNA for all genes analysed. The methylation of the remaining eight genes was therefore considered to have some potential biological function, all eight genes were analysed further.

3.4.5 Analysis of selected genes in primary breast tumour/normal paired samples

Access to a second, independent cohort of 20 breast tumour samples with adjacent normal breast tissue DNA was available, therefore, to assess tumour specificity of DNA methylation all eight genes were analysed in this cohort of samples. Five of the eight genes (*CIDE-A*, *DBC1*, *EMILIN2*, *FBLN2* and *SALL1*) showed greater levels of methylation in tumour samples compared to adjacent normal tissue and in most cases, when methylation was observed in the tumour sample, none was observed in corresponding normal samples (figure 3.5). Tumour specific methylation was observed in 4 of the 8 genes at frequencies of 4/11 (36.4%) for *EMILIN2*, 6/19 (31.6%) for *FBLN2*, 4/16 (25%) for *DBC1*, and 7/14 (50%) for *CIDE-A*. *SALL1* was the only gene to demonstrate some methylation in adjacent normal tissue, 9/20 (45%) samples were methylated in the tumour only and 1/20 (5%) of samples showed a lower level of methylation in the adjacent normal tissue of a methylated tumour. For the remaining three genes (*COMP*, *EPSTI1* and *SIM2*), methylation was detected at a similar level in both tumour and normal tissue, demonstrating that methylation is a not a cancer-specific event in these genes.

Taken together, the cancer cell line and primary tumour methylation results, expression results pre and post 5-azaDC treatment and tumour specificity of methylation, five genes were

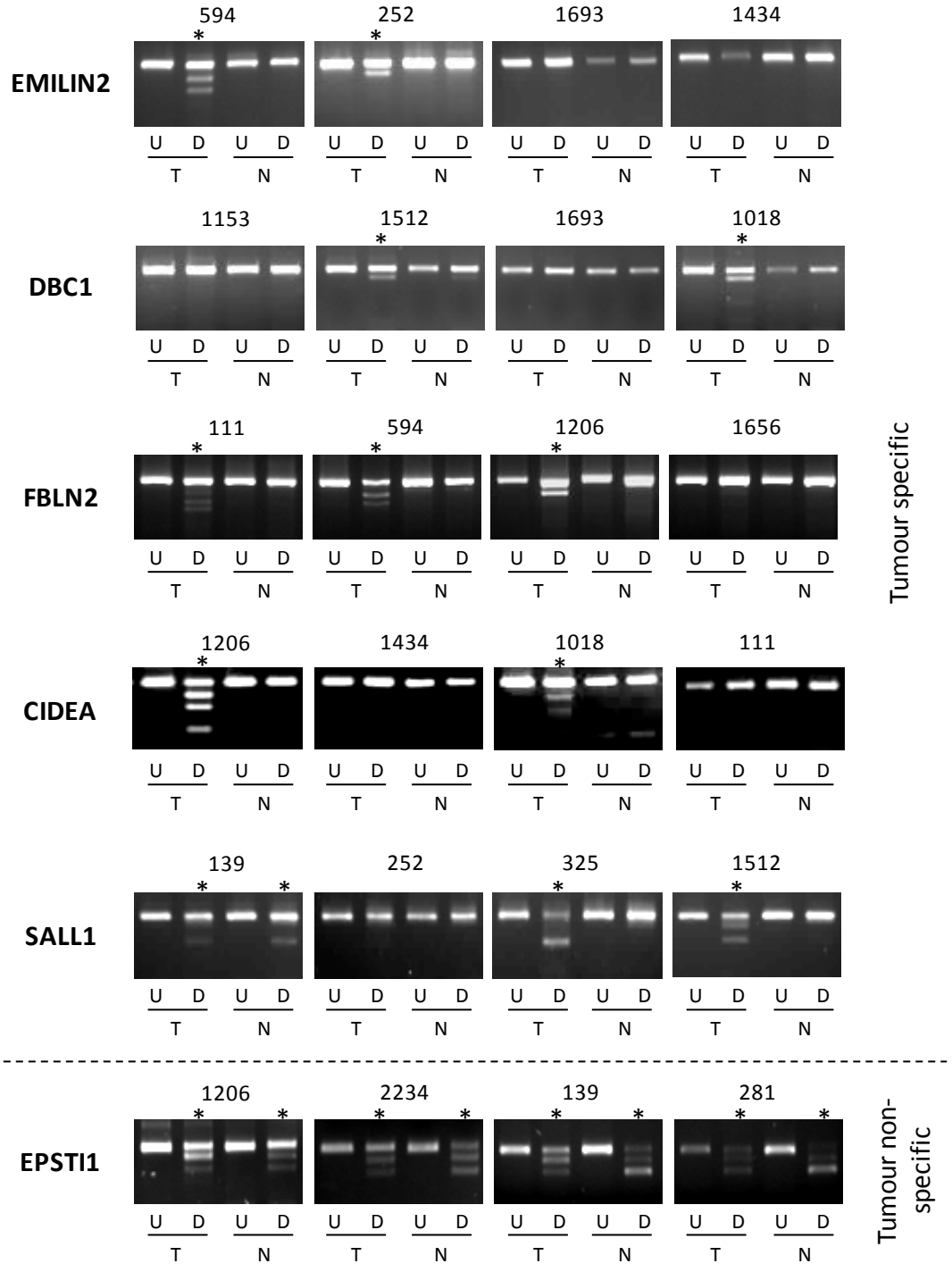


Figure 3.5 Tumour specificity analysis of selected genes
CoBRA results for 6 of the 8 genes analysed for tumour specificity. For each gene CoBRA results are shown for four tumour (T)/normal (N) pairs. Undigested products (U) are run next to digested products (D). * indicates methylation

identified that represented good candidates. For three of these five good candidate genes, *EMIILIN2*, *CIDE-A* and *SALL1*, clone sequencing was performed on methylated primary tumour samples (figures 3.6, 3.7 and 3.8 respectively). Results from this confirmed the methylation observed by CoBRA digest and demonstrated the extent of the methylation throughout the investigated region of each genes CpG island. Methylation indexes (MIs) were high for *CIDE-A* and *SALL1* demonstrating extensive methylation throughout whilst the MIs were relatively low for *EMILIN2* (34% and 36%).

3.4.6 Association of methylation and clinical features

Relapse, survival, ER (oestrogen receptor) and PR (progesterone receptor) status, presence of lymph node metastases, menopause status and age at diagnosis data were available for the cohort of 40 breast tumour samples. Methylation was significantly associated with ER positive tumours in *CIDE-A* (p=0.016) and PR positive tumours in *FBLN2* (p=0.013). In *EMILIN2*, methylation was significantly associated with ER and PR positive tumours (p=0.0009 and p=0.0082 respectively), relapse (p=0.031) and, most importantly, with poor disease free survival (p=0.041) (figure 3.9).

3.4.7 Expression of *EMILIN2* in primary tumour samples

Clinically, methylation of *EMILIN2* was determined to be the most interesting and therefore expression of *EMILIN2* was assessed in primary tumour samples by quantitative real-time RT-PCR. This work was done by Dr. Sophie Vacher in Dr. Ivan Bieche's laboratory (Oncogenetic Laboratory, INSERM U 735, Centre René Huguenin, Saint Cloud, France). Results demonstrated significantly decreased expression in methylated samples compared to unmethylated samples (p=0.018) (figure 3.10), suggesting

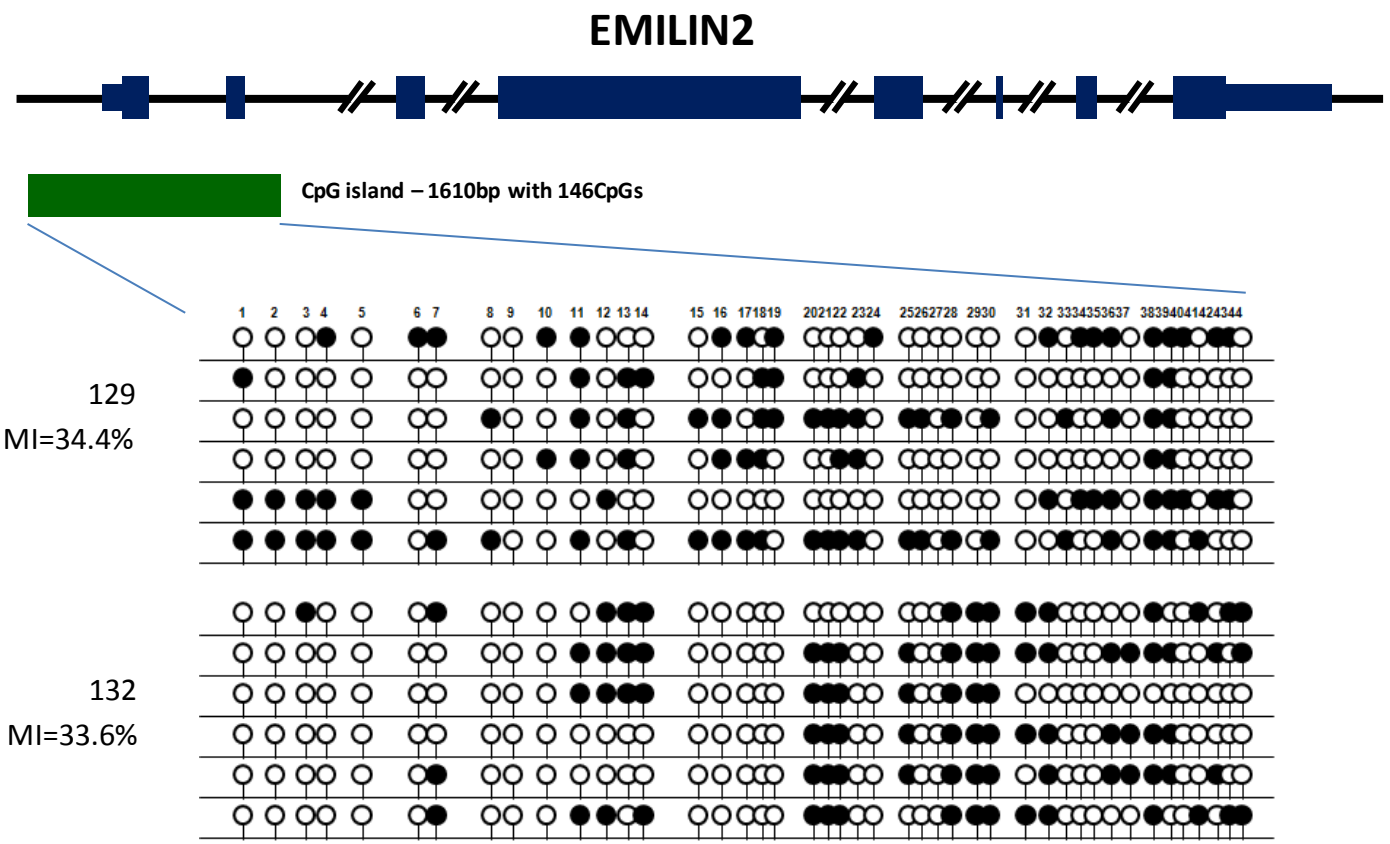


Figure 3.6_Sequencing results for EMILIN2

A gene schematic of *EMILIN2* is shown to demonstrate the region being analysed. Clone sequencing results are shown for two methylated samples, each line represents a single allele and black and white circles represent methylated and unmethylated CpGs respectively. Methylation indexes (MIs) are calculated as a percentage of methylated CpGs out of the total number of CpGs assessed.

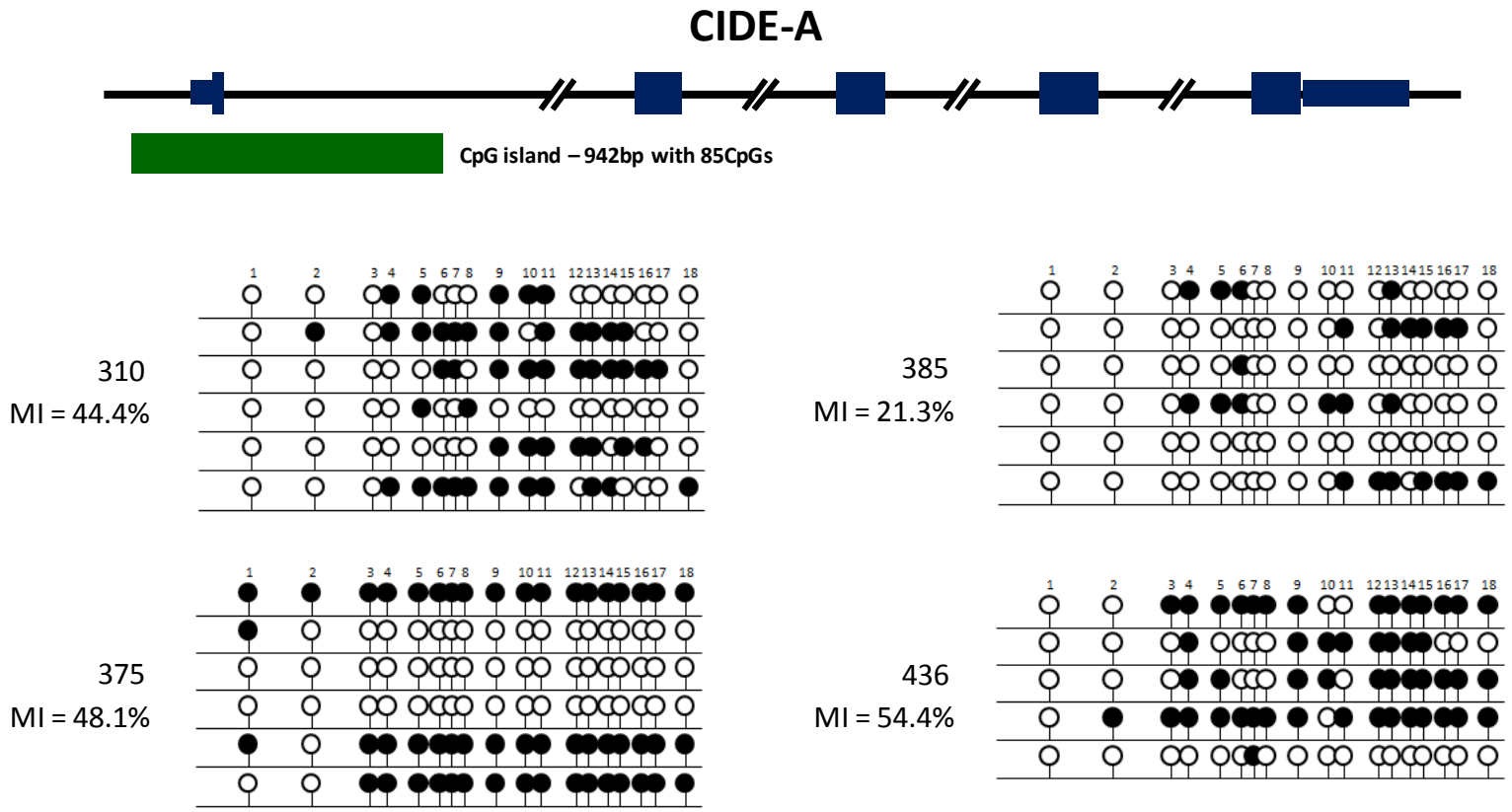


Figure 3.7_Sequencing results for CIDE-A

A gene schematic of *CIDE-A* is shown to demonstrate the region being analysed. Clone sequencing results are shown for four methylated samples, each line represents a single allele and black and white circles represent methylated and unmethylated CpGs respectively. Methylation indexes (MIs) are calculated as a percentage of methylated CpGs out of the total number of CpGs assessed.

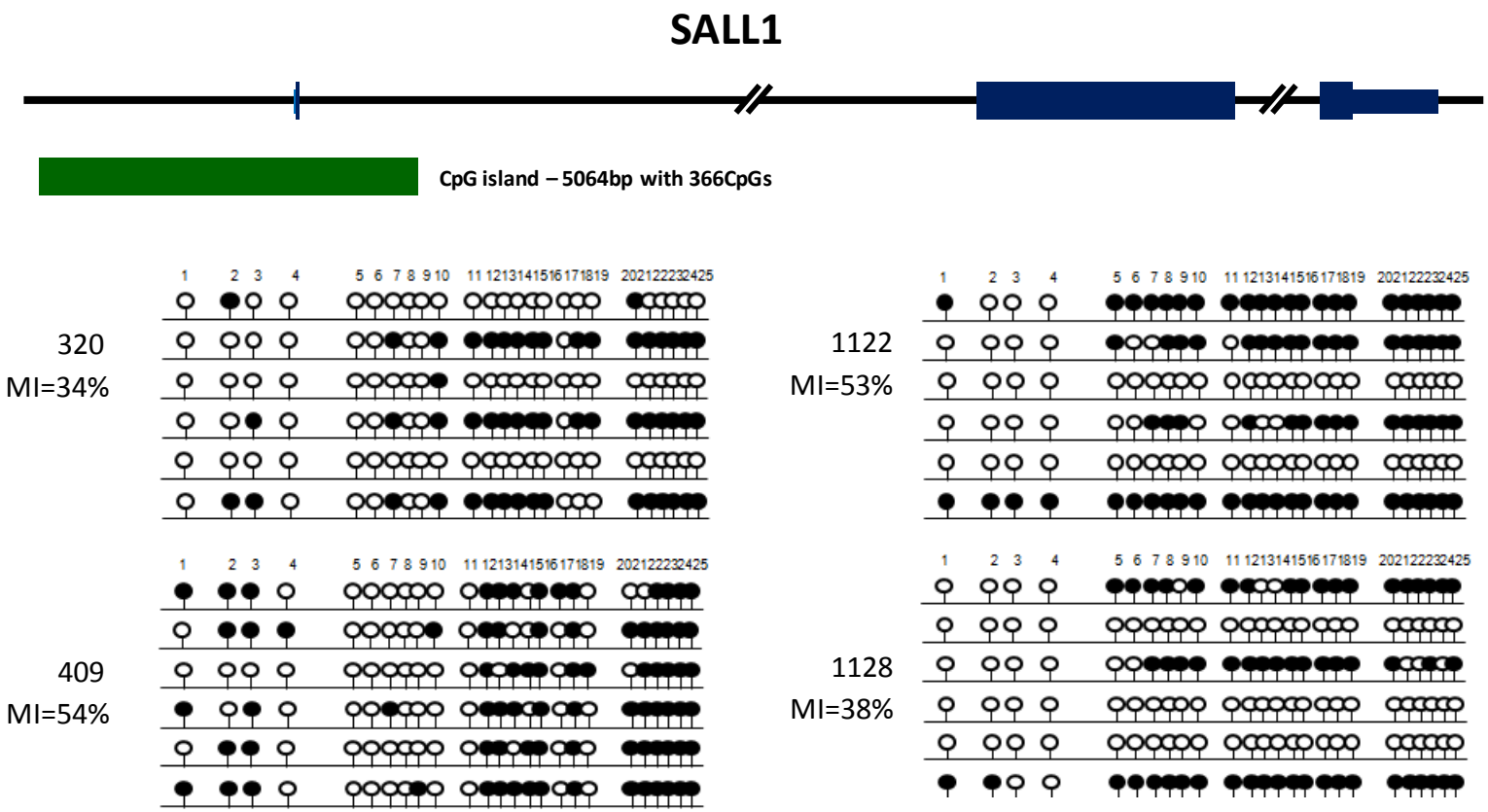


Figure 3.8_Sequencing results for SALL1

A gene schematic of *SALL1* is shown to demonstrate the region being analysed. Clone sequencing results are shown for four methylated samples, each line represents a single allele and black and white circles represent methylated and unmethylated CpGs respectively. Methylation indexes (MIs) are calculated as a percentage of methylated CpGs out of the total number of CpGs assessed.

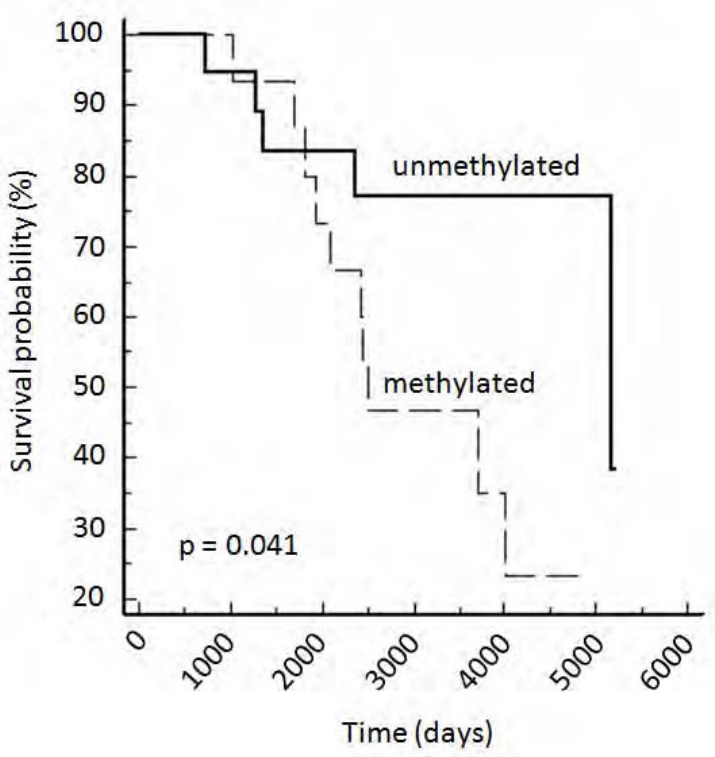


Figure 3.9 Kaplan-Meier analysis for *EMILIN2*
The Kaplan-Meier graph is shown for the association between *EMILIN2* methylation and poor disease free survival. This analysis was performed on results from 35 tumour samples. The minimum and median follow-up times were 247 days and 2501 days respectively. For non-relapsing tumours, the maximum period of follow-up was 5185 days.

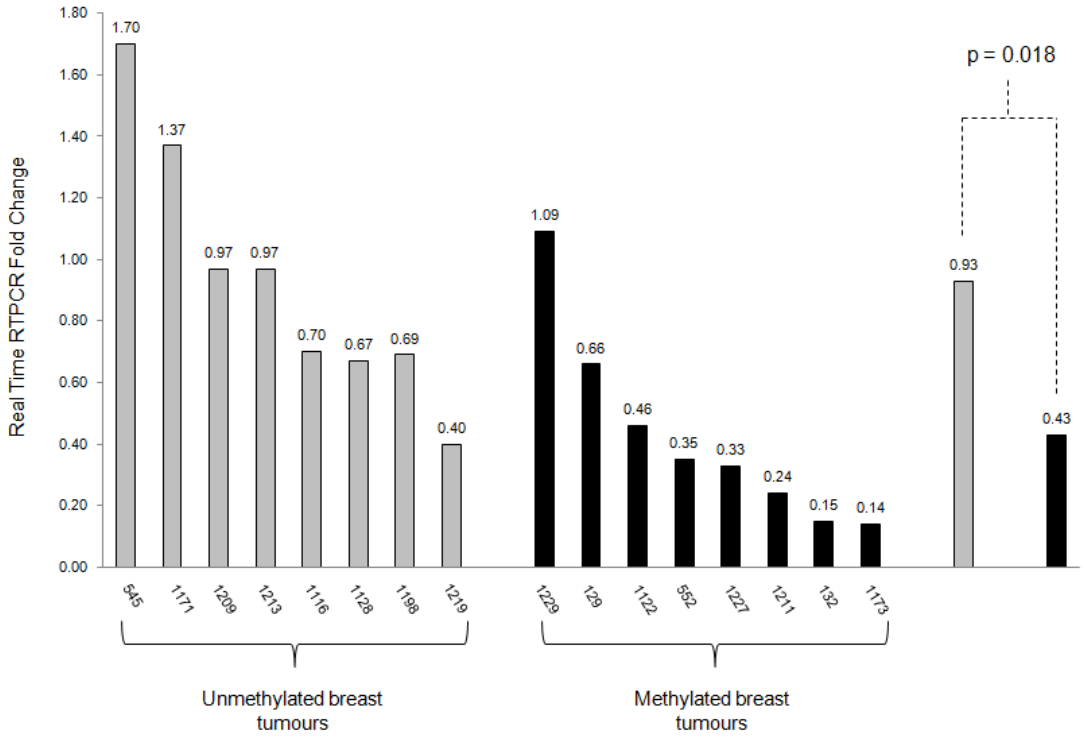


Figure 3.10 *EMILIN2* expression in primary tumour samples

This graph shows the expression levels of unmethylated and methylated tumours as determined by quantitative real-time RT-PCR. The mean expression levels of unmethylated tumours (grey bar) and methylated tumours (black bar) are shown on the far-right of the graph. The statistical significance of the difference in these means (using Student's T-test is) also shown. Data for these graphs were obtained by Dr. Sophie Vacher from Dr. Ivan Bieche's laboratory (Oncogenetic Laboratory, INSERM U 735, Centre René Huguenin, Saint Cloud, France).

methylation of *EMILIN2* is correlated with loss of *EMILIN2* expression in primary tumours and therefore likely to have a biological effect.

3.4.8 Analysis of *CIDE-A, DBC1, EMILIN2, FBLN2* and *SALL1* in other common solid tumour types

Hypermethylation of genes is often not restricted to one tumour type and so for the five good candidates methylation was assessed in other available common solid tumour types. Cancer cell lines were available for glioma, lung, colorectal and prostate. Frequent methylation was observed in all cancer types for all genes with the exception of *SALL1*, which was unmethylated in glioma cell lines (table 3.4). *DBC1* was not tested in lung cancer cell lines as methylation in lung cancer is known (Izumi *et al*, 2005).

The frequent methylation in cell lines led to analysing these genes in primary tumour samples. Due to the frequent nature of methylation in the multiple cell line types analysed ovarian primary tumour samples were also analysed despite ovarian cancer cell lines not being available. In-house, primary lung, glioma and ovarian tumour samples were available. Analysis of colorectal and prostate tumours was carried out by others in the laboratories of Dr. Luke Hesson (Prince of Wales Clinical School, University of New South Wales, Sydney, Australia) and Dr. Reinhard Dammann (Institute for Genetics, Justus Liebig University Giessen, Heinrich-Buff-Ring 58-62, D-35392 Giessen, Germany) respectively.

3.4.8.1 *EMILIN2*

The only frequent primary tumour methylation observed for *EMILIN2* was in colorectal tumours. CoBRA assay demonstrated frequent methylation in colorectal tumours (19/57; 33.3%) compared to normal colon tissue (1/82; 1.2%). Colorectal adenomas also showed

		EMILIN2	SALL1	FBLN2	CIDE-A	DBC1
Colorectal	Cell lines	4/5 (80%)	6/6 (100%)	4/4 (100%)	8/8 (100%)	8/8 (100%)
	Tumours	19/57 (33.3%)	44/53 (83.0%)	54/58 (93.1%)	7/8 (87.5%)	8/8 (100%)
	Adenomas	7/22 (31.8%)	17/19(89.5%)	23/23 (100%)	7/7 (100%)	8/8 (100%)
	Normal tissue	1/82 (1.2%)	30/79 (38.0%)	74/80 (92.5%)	16/16 (100%)	16/16 (100%)
Lung	Cell lines	4/10 (40%)	4/13 (30.8%)	7/9 (77.8%)	11/15 (73.3%)	-
	Tumours	2/9 (22.2%)	2/15 (13.3%)	0/10 (0%)	2/18 (11.1%)	-
	Normal tissue	1/7 (14%)	3/17 (17.6%)	0/10 (0%)	0/16 (0%)	-
Glioma	Cell lines	3/6 (50%)	0/7 (0%)	7/7 (100%)	7/7 (100%)	3/7 (42.9%)
	Tumours	2/11 (18.2%)	-	5/11 (45.5%)	3/19 (15.8%)	0/13 (0%)
	Normal tissue	0/1 (0%)	-	didn't work	0/4 (0%)	0/4 (0%)
Prostate	Cell lines	2/5 (40%)	3/4 (75%)	1/5 (20%)	3/4 (75%)	2/5 (40%)
	Tumours	0/14 (0%)	5/14 (35.7%)	5/14 (35.7%)	3/14 (21.4%	1/14 * (7.1%)
	Normal tissue	0/4 (0%)	1/4 (25.0%)	0/4 (0%)	1/4 (25.0%)	2/4 * (50%)
Ovarian	Cell lines	-	-	-	-	-
	Tumours	0/18 (0%)	3/19 (15.8%)	3/18 (16.7%)	1/9 (11.1%)	3/15 (20%)

Table 3.4 Results of other solid tumour analysis

This table shows the results of cell line and primary tumour analysis of *EMILIN2*, *SALL1*, *FBLN2*, *CIDE-A* and *DBC1* in colorectal, lung, gliomas, prostate and ovarian tumours. Primary tumour (and normal tissue) analysis of colorectal and prostate tissue DNA was carried out in the laboratories of Dr. Luke Hesson (Prince of Wales Clinical School, University of New South Wales, Sydney, Australia) and Dr. Reinhard Dammann (Institute for Genetics, Justus Liebig University Giessen, Heinrich-Buff-Ring 58-62, D-35392 Giessen, Germany) respectively.

frequent methylation (7/22; 31.8%), indicating that *EMILIN2* methylation may be an early event in colon carcinogenesis. All remaining tumours analysed for methylation of *EMILIN2* showed either infrequent or no methylation; NSCLC (2/9; 22.2%), gliomas (2/11; 18.2%), prostate (0/14; 0%) and ovarian (0/18 ,0%) (table 3.4).

3.4.8.2 *SALL1*

Colorectal and prostate tumours both showed frequent methylation of *SALL1*. Colorectal tumours demonstrated methylation in 44/53 (83%) of samples compared to 30/79 (38%) of normal colorectal tissue samples. As for *EMILIN2*, *SALL1* showed a high level of methylation in colorectal adenomas 17/19 (89.5%). Prostate tumours demonstrated methylation in 5/14 (35.7%) of tumours and 1/4 (25%) of normal prostate tissue samples. Both lung and ovarian tumours showed infrequent levels of methylation (2/15; 13.3%) and 3/19 (15.8%) respectively (table 3.4). Gliomas were not analysed for methylation since cell lines were unmethylated.

3.4.8.3 *FBLN2*

Frequent methylation of *FBLN2* was observed in gliomas and prostate tumours. Gliomas showed methylation in 5/11 (45.5%) tumour samples and 1/14 (7%) normal brain samples. Prostate tumours were methylated in 5/14 (35.7%) of samples and 0/4 (0%) normal prostate tissue. Infrequent methylation was observed in ovarian cancer samples (3/18; 16.7%) and no methylation was observed in lung tumour samples (0/10) (table 3.4). Methylation was highly prevalent in colorectal cancers, adenomas and normal colorectal tissue suggesting *FBLN2* may be normally methylated in this tissue type.

3.4.8.4 *CIDE-A*

Infrequent methylation of *CIDE-A* was observed for all tumour types analysed, methylation frequencies ranged from 11% in lung and ovarian cancers to 21.4% in prostate cancers (table 3.4). As for *FBLN2*, high methylation levels were observed for *CIDE-A* in cancer, adenoma and normal colorectal tissues, suggesting *CIDE-A* is normally methylated in colorectal tissue.

3.4.8.5 *DBC1*

DBC1 has previously been shown to be methylated in lung cancer (Izumi *et al*, 2005) and has therefore not been analysed in lung cancer in this study. Low level methylation of *DBC1* was observed in prostate cancer samples (1/14; 7%), but was present at a higher level in normal prostate tissue (2/4; 50%). The only cancer type to show more frequent methylation of *DBC1* was ovarian (3/15; 20%) (table 3.4). As for *FBLN2* and *CIDE-A*, high levels of methylation were observed in cancer, adenoma and normal colorectal tissue, suggesting *DBC1* is normally methylated in this tissue.

From this analysis, relevant levels of methylation have been observed in prostate (*SALL1* and *FBLN2*), colorectal (*EMILIN2*) and gliomas (*FBLN2*) (figure 3.11).

3.4.9 miRNA analysis

Analysis of probes considered to be associated with miRNAs identified 66 probes that showed a 5-fold enrichment in at least three out of five breast cancer cell lines compared to NHMECs. These 66 probes represented 25 miRNAs, of which three were located on the X chromosome and therefore discarded from any further analysis. Of the remaining 22, 7 were represented within the list by multiple probes and the remaining 15 were represented by single probes only. Of the 7 represented by multiple probes, 5 (hsa-mir-183, hsa-mir196a1, hsa-mir-

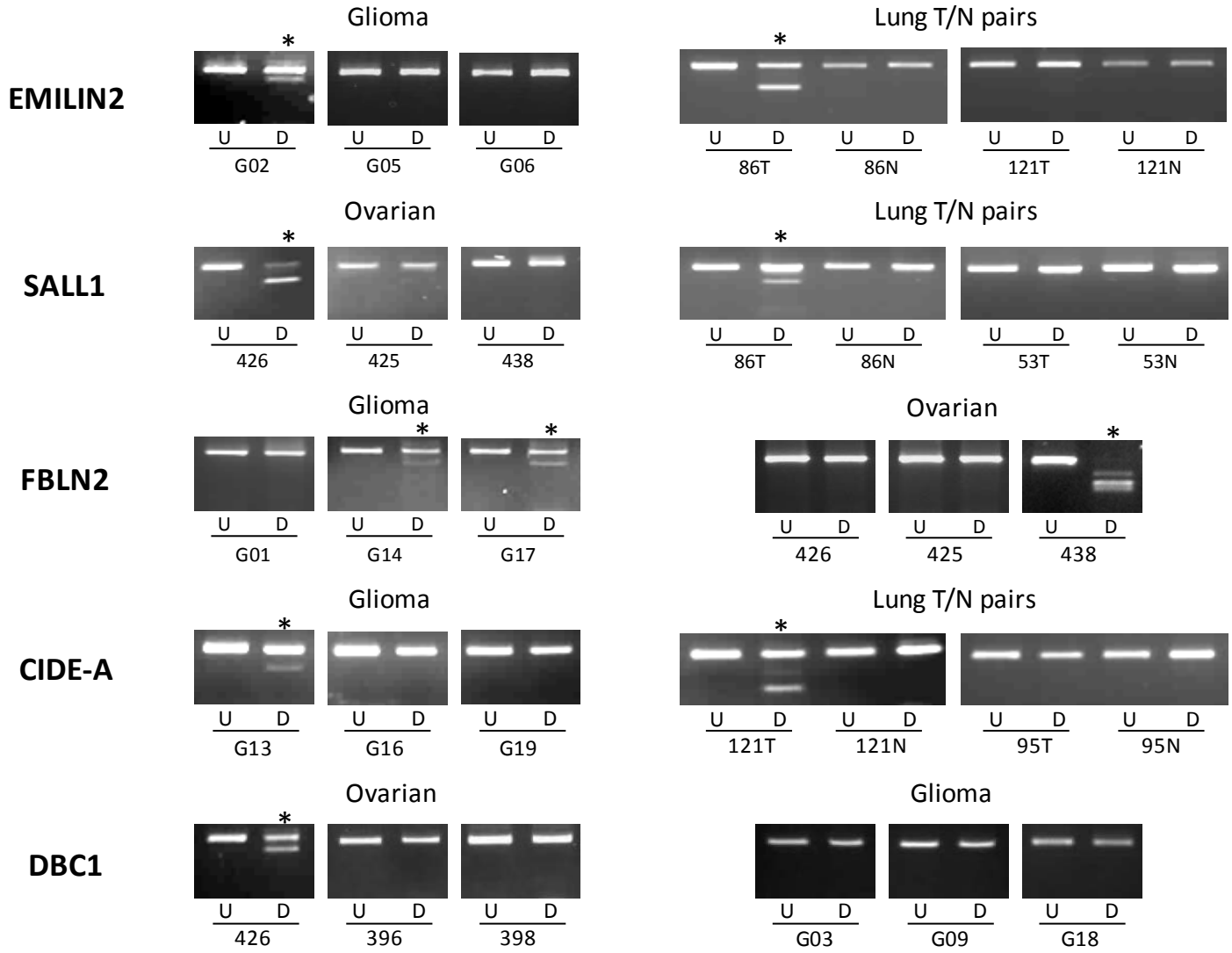

EMILIN2
Glioma
Lung T/N pairs
U D
G02
G05
G06
86T
86N
121T
121N
SALL1
Ovarian
Lung T/N pairs
426
425
438
86T
86N
53T
53N
FBLN2
Glioma
Ovarian
G01
G14
G17
426
425
438
CIDE-A
Glioma
Lung T/N pairs
G13
G16
G19
121T
121N
95T
95N
DBC1
Ovarian
Glioma
426
396
398
G03
G09
G18

Figure 3.11 CoBRA results for other solid tumour samples
CoBRA results are shown for two cancer types each for *EMILIN2*, *SALL1*, *FBLN2*, *CIDE-A* and *DBC1*. Undigested products (U) are shown next to digested products (D). For lung samples, results for associated tumour (T) and normal (N) tissue are shown for each patient sample. * indicates methylation

596, hsa-mir-9-3 and hsa-mir-181c) were within a promoter region rather than downstream. Promisingly, one of these 5 miRNAs detected, hsa-mir-196a-1, has previously been shown to have the associated CpG island methylated in breast cancer (Hoffman *et al*, 2009), one has been shown to have altered expression in lung cancer, hsa-mir-183 (Cho *et al*, 2009) and another, hsa-mir-181c, has two associated family members (hsa-mir-181a and hsa-mir-181b) that have been shown to have tumour suppressor effects in glioma cells (Shi *et al*, 2008). Although confirmed analysis of methylated miRNAs was not a planned part of this investigation, an initial analysis was performed upon the best candidate, hsa-mir-183 (has-mir196a-1 having already been previously demonstrated to be methylated in breast cancer), due to the number of positive probes and the possible biological significance. CoBRA assays of the breast cancer cell lines were performed to assess the region of the has-mir-183 CpG island with the greatest representation of probes and demonstrated infrequent methylation (1/9; 11%) in this region. Further investigation was not performed.

3.5 DISCUSSION

Breast cancer represents the most common cancer in women both worldwide and in the UK (Jemal *et al*, 2011; CRUK cancer statistics) and the second most common form of cancer related deaths in women in the UK (CRUK cancer statistics). Death rates from breast cancer have been falling over the past 10 years (CRUK cancer statistics). This has been through a combination of increased disease understanding leading to better, more targeted therapies and earlier disease detection and preventative measures. Great leaps forward have been made in more personalised/targeted therapies for breast cancer patients such as trastuzumab (Herceptin) treatment for *ERBB2* overexpressing tumours (Ménard *et al*, 2003; Hudis *et al*, 2007), tamoxifen treatment for ER positive tumours (Jordan, 1993) and the development of MammaPrint and Oncotype DX expression profile assays to identify patients that are most at risk of relapse and therefore benefit from more rigorous therapy (Glas *et al*, 2006; Cobleigh *et al*, 2005). Preventative measures, such as a reduction in the use of post-menopausal hormone therapies, and the development of improved detection methods, such as nation-wide mammogram screening programs, have also had a dramatic effect on overall disease outcome. Nonetheless, breast cancer still remains the second most common cancer related death in the UK and the identification of further deregulated molecular targets are still required to enhance breast cancer outcomes. Identifying cancer specific DNA hypermethylation events represents a useful strategy for identifying genes that may be good targets for the identification of deregulated processes in cancer, but also as potential events that could be developed into biomarkers for diagnosis/prognosis/treatment plans. This study represents the first example of using the methylated CpG island recovery assay (MIRA) coupled with CpG island arrays in invasive breast cancer and one of the first studies to use a methylated DNA pull-down approach to identify methylated genes in breast cancer with other examples including a

MeDIP study on BRCA mutated primary patient samples (Flanagan *et al*, 2010) and a MeDIP approach coupled with sequencing on breast cancer cell lines (Ruicke *et al*, 2010).

In addition to the present study, the MIRA assay in conjunction with CpG microarrays has now been used to assess the methylation status of lung cancers (Rauch *et al*, 2006), DCIS (Tommasi *et al*, 2009), childhood ALL (Dunwell *et al*, 2010) and astrocytomas (Wu *et al*, 2010). In each case, the assay has been useful in identifying novel methylated genes in the respective cancers indicating the assay represents an effective methodology for such studies.

Ductal carcinoma in situ (DCIS) is an early lesion of breast tissue that often acts as a precursor for invasive breast tumours. A MIRA study was carried out on 6 DCIS samples by the collaborators who performed the MIRA assay presented in this study using the same type of CpG island microarray. That study identified 81 target genes considered to be hypermethylated, and a further 27 target CpG islands that were unconnected to individual genes. Confirmation of some of these targets and subsequent analysis in invasive breast cancers found high levels of methylation to be both present and increased in invasive tumours (Tommasi *et al*, 2009). The study demonstrated one third of their target genes to be homeobox related family members and suggest methylation of multiple homeobox genes is a frequent and early event in breast cancer tumourigenesis (Tommasi *et al*, 2009). Comparison of the high stringency gene list from the work in this chapter against the 81 target genes from Tommasi *et al* identifies 27 genes that are within both lists, demonstrating a pleasing degree of overlap but also distinct differences. Of these, 13 (48%) are homeobox related family member genes. Of the 27 genes in both lists, two (*BARHL2* and *KLF11*), were analysed in this chapter, both of which showed infrequent methylation when analysed by CoBRA in breast cancer cell lines and therefore were not analysed further.

Use of the MIRA assay in this study has identified five good candidate genes that are frequently methylated genes in breast cancer cell lines and sporadic ductal breast cancer patient samples. In addition, this study has also demonstrated frequent methylation of some of these genes in other cancers. Each gene, through known functions or association with other cancers, shows some promise as a factor that could be involved in the pathogenesis of breast cancers.

3.5.1 *CIDE-A*

This study has identified hypermethylation in 53% of analysed breast tumours, tumour specific methylation and re-expression following de-methylation in cell lines. No relevant levels of methylation were observed in other cancer types analysed, suggesting hypermethylation in cancer is so far specific to breast cancer.

CIDE-A (cell death inducing DFFA-like effector a) is a 4 exon gene located at 18p11.2. It has a CpG island situated upstream of and across the transcription start point consisting of 942bp and 85 CpGs. Methylation of the CpG island has been shown to have a substantial effect on the expression of *CIDE-A*, providing a mechanism for tissue specific expression of the *CIDE-A* gene (Li *et al*, 2008). To date, expression has been observed in adipose, heart, lymph node and from the present study, breast, tissues. Of other analysed tissues, no expression has been observed in liver, stomach, spleen and kidney (Li *et al*, 2008). The authors identified the minimal promoter region of *CIDE-A* to be within the CpG island and contain Sp1/Sp3 transcription factor binding sites, which contain a CpG in the middle of their binding consensus sequence. They showed methylation of this region resulted in dramatically reduced expression of *CIDE-A* and that methylation was frequent in non-expressing tissues. They also demonstrated altered chromatin structure in methylated samples. The primers used in this

chapter amplify an overlapping region to that analysed in the study by Li *et al* (2008). Hypermethylation of *CIDE-A* has not previously been identified in cancer tissues respective to normal tissues, but the data from this study in breast cancer cell lines and Li *et al* (2008) suggests that the methylation detected is likely to reduce expression in methylated tumour samples. It would be nice to have data on the expression of *CIDE-A* in methylated and unmethylated primary breast tumours both at the mRNA and protein level. The work in this chapter also detected high levels of methylation in colorectal and prostate normal and cancerous tissues, suggesting expression of *CIDE-A* is unlikely in these tissues in their normal state.

CIDE-A protein is a member of the CIDE family of proteins which consists of three members, CIDE-A, CIDE-B and CIDE-C. It is a 24kDa protein that has been shown to be able to activate apoptosis in a way which is unaffected by caspase inhibitors but inhibited by DFF45 (DNA fragmentation factor 45kDa) (Inohara *et al*, 1998). CIDE-A also has a role in metabolism, in mice reduced *CIDE-A* levels produce a higher metabolic rate and lipolysis in brown adipose tissue accompanied by a resistance to diet-induced obesity and diabetes (Zhou *et al*, 2003). In cows, *CIDE-A* expression levels and that of an interacting partner, *UCP1*, are increased in mammary epithelial cells and lactating mammary glands when incubated with long chain fatty-acids and decreased when incubated with insulin. *CIDE-A* expression also appeared to be highest at periods of peak lactation (Yonezawa *et al*, 2009), suggesting a specific role for *CIDE-A* in normal mammary glands (in bovine tissue at least) and therefore conferring a possible role for *CIDE-A* deregulation in breast cells as a potential aggravating event for breast tumourigenesis. This functional information may also be relevant to the fact that *CIDE-A* methylation so far appears to be a breast cancer specific event.

The *CIDE-A* genes is located at 18p11.2, a region of LOH (loss of heterozygosity) in esophageal squamous cell carcinoma (Karkera *et al*, 2000) suggesting *CIDE-A* loss as a possible factor in esophageal squamous cell carcinoma.

Two mutations have been described in *CIDE-A* in the COSMIC database (COSMIC – http://www.sanger.ac.uk/genetics/CGP/cosmic/), one in an ovarian tumour and one in a glioma tumour. Both mutations are single base substitutions resulting in amino acid changes of which one was predicted to be probably damaging by PolyPhen2 (http://genetics.bwh.harvard.edu/pph2/) and the other intolerant by SIFT (http://sift.jcvi.org/www/SIFT_BLink_submit.html) but not vice versa (table 3.5). Taken together with the results from this study and the normal function of CIDE-A protein, *CIDE-A* appears to be a promising candidate for a role in tumourigenesis specifically in breast and possibly in esophageal squamous cell carcinoma. Further research on the potential role of *CIDE-A* in breast cancer is warranted to investigate this further.

3.5.2 *DBC1*

This study has identified frequent (26%), tumour specific methylation of *DBC1* in breast cancer. Methylation has also been demonstrated in ovarian cancer at a lower frequency (20%). *DBC1* expression has been shown to be affected by methylation status in breast cancer cell lines.

DBC1 (*deleted in bladder cancer 1*), NM_014618, not to be confused with *deleted in breast cancer 1* (*DBC1/KIAA1967*), is an eight exon, 202kb gene. It has a CpG island consisting of 1128bp containing 104 CpGs situated upstream of and including the transcription

GENE	Number of mutations in breast	Number of mutations in other cancer types	PolyPhen2 analysis (score)	SIFT analysis (p-value)
CIDE-A	0/11	p.P116T (Ovarian - 1/1)	Probably damaging (0.967)	Tolerant (0.49)
		p.E203K (Glioma - 1/21)	Benign (0.029)	Intolerant (0.00)
DBC1	0/48	p.G109C (Ovarian -1/2)	Possibly damaging (0.773)	Intolerant (0.01)
		p.G448S (Ovarian - 1/2)	Probably damaging (0.990)	Intolerant (0.00)
		p.V562I (Glioma - 1/446)	Possibly damaging (0.632)	Tolerant (0.31)
		p.G684D (Glioma - 1/446)	Benign (0.140)	Intolerant (0.01)
		p.P712T (Colorectal - 1/37)	Probably damaging (0.955)	Intolerant (0.00)
EMILIN2	0/11	p.T245K (Ovarian - 1/1)	Possibly damaging (0.326)	Tolerant (0.14)
FBLN2	0/11	p.A1101E (Ovarian - 1/1)	Possibly damaging (0.698)	Tolerant (0.20)
SALL1	0/11	p.T66P (Ovarian - 1/3)	Probably damaging (0.975)	Intolerant (0.01)
		p.P577H (Ovarian 1/3)	Possibly damaging (0.726)	Tolerant (0.37)
		p.E859D (Ovarian - 1/3)	Possibly damaging (0.447)	Tolerant (0.26)

Table 3.5_Mutational data of *CIDE-A*, *DBC1*, *EMILIN2*, *FBLN2* and *SALL1* from COSMIC

This table shows the mutational data from COSMIC for each of the five good candidate genes. The number of mutations identified for breast and other cancers are shown along with PolyPhen2 and SIFT predictions on whether these mutations are likely to be damaging/tolerant. COSMIC: http://www.sanger.ac.uk/genetics/CGP/cosmic/. PolyPhen2: http://genetics.bwh.harvard.edu/pph2. SIFT: http://sift.jcvi.org/www/SIFT_BLink_submit.html.

start point. The gene produces an 88 kDa protein. *DBC1* was initially identified as a potential tumour suppressor gene in bladder cancer due to its location at 9q32-q33, a common region for LOH in this cancer (Habuchi *et al*, 1998). The Habuchi study (1998) also showed it to be frequently hypermethylated in bladder cancer but not mutated. Subsequently, hypermethylation has also been identified in NSCLC (Izumi *et al*, 2005), ALL (San José-Enériz *et al*, 2006) and oral squamous cell carcinoma (Gao *et al*, 2004). In the COSMIC database, *DBC1* is listed as having been assessed for mutation in 533 samples, covering ovarian, breast, large intestine and glioma tumours. Five mutations have been identified, 2 in ovarian (out of 2 samples), 1 in large intestine (out of 37 samples) and 2 in gliomas (out of 446 samples). All mutations identified were base substitutions resulting in amino acid changes, 3 predicted to be damaging and intolerant by PolyPhen2 and SIFT while one was predicted to be probably damaging by PolyPhen2 and the other intolerant by SIFT but not vice versa (table 3.5).

The *DBC1* gene protein product is thought to function in cell cycle regulation, deregulation of which is highly amenable to tumourigenesis. Overexpression of *DBC1* decreases cell growth without appearing to affect apoptosis, cells appear to have a reduced ability to progress from the G1 to S phases of the cell cycle (Nishiyama *et al*, 2001). The data obtained in this study, the function of *DBC1* and involvement in other cancers suggests *DBC1* downregulation is a good candidate for contribution to the pathogenesis of breast cancer.

3.5.3 *EMILIN2*

This study has identified *EMILIN2* as being frequently methylated in breast and colorectal cancers (44% and 33.3% respectively). In both cases *EMILIN2* methylation was specific for the tumour and in colorectal cancers is potentially an early event. In addition, methylation has

been shown to be capable of repressing expression of *EMILIN2* in cell lines and downregulation of *EMILIN2* associates with methylation in primary tumour samples. *EMILIN2* methylation in breast cancer associated with ER status of tumours and, more importantly, associated with relapse and poor disease free survival.

EMILIN2 (*elastin microfibril interfacer 2*) is a 67kb 8 exon gene and has a CpG island consisting of 1610bp containing 146 CpGs situated upstream of and across the transcription start site. It produces a 115kDa protein that forms a homotrimer. *EMILIN2* was the second of three identified members of the EMILIN family. The family are a group of extracellular matrix glycoproteins characterised by a C-terminal gC1q domain and an N-terminal EMI domain (Colombatti *et al*, 2000). EMILIN2 is capable of inducing apoptosis by interacting with the TRAIL death receptor, DR4 in both *in vitro* and *in vivo* model systems (Mongiat *et al*, 2007). EMILIN2 has been shown to have contrasting effects on tumour cells. Confusingly, it behaves both as a pro-apoptotic agent and a pro-angiogenic agent (Mongiat *et al*, 2010). When introduced into both *in vitro* and *in vivo* models, EMILIN2 was able to reduce growth via a pro-apoptotic mechanism utilising caspase-8 via DR4 receptor binding, however, *in vivo*, the remaining tumour had greater vessel density than untreated controls suggesting a proangiogenic effect. In combination with an antiangiogenic drug, tumours could be completely reduced *in vivo*, suggesting a potentially exploitable mechanism for treatment utilising EMILIN2. If this ever comes to fruition, whether tumours are methylated for *EMILIN2* may help to identify patients who would be most susceptible to this treatment.

EMILIN2 is located at 18p11.3 which has been shown to be a common region of LOH in breast cancer (Kittiniyom *et al*, 2001). In the Kittiniyom study (2001) the authors detected a high frequency of LOH (63%) at 18p11.3 that was prevalent in a high proportion of early grade DCIS (ductal carcinoma in situ) lesions, suggesting that this chromosomal event occurs

at an early stage in breast cancer progression. LOH at this locus associated with microsatellite instability and LOH at other common sites in breast cancer (3p, 9p, 17p, 17q). LOH at this locus has also been described in CRC (Kawakami *et al*, 2002), the association of LOH at this locus with microsatellite instability in breast cancer may therefore be of relevance to the CRC data obtained in this chapter and it would be interesting to see if methylation of *EMILIN2* associates with microsatellite unstable CRC tumours.

EMILIN2 has not previously been shown to be methylated in cancer and the frequencies identified in this study suggest that methylation is most likely specific to a subset of cancers. The COSMIC database lists one mutation in *EMILIN2* out of 45 samples assessed (consisting of a range of ovarian, breast, large intestine and glioma tumours). The mutation was a heterozygous mutation found in ovarian cancer consisting of a C to A substitution (c.734C>A) resulting in a threonine to lysine substitution at the amino acid level (p.T245K) predicted only to be possibly damaging by PolyPhen2 (table 3.5). The low prevalence of mutations suggests this is unlikely to be a major mechanism for inactivation/altered activity in common cancers however, coupled with the location within a region of LOH in breast cancer, the frequent methylation levels in breast and CRC tumours, the potential early stage during which expression may be lost and its pro-apoptotic functions, *EMILIN2* is a promising candidate for further work in breast cancer. In addition, the association between *EMILIN2* methylation status with relapse and poor disease free survival suggests that the detection of *EMILIN2* methylation may not only be of use in terms of good candidate gene identification but also as a potential candidate for further analysis as a marker of disease outcome.

3.5.4 *FBLN2*

This study has identified frequent methylation (34%) in breast cancer that has been demonstrated to be tumour specific. Frequent methylation levels were also observed in prostate (35.7%) and glioma tumours (45.5%). In addition, breast cancer cell lines showed re-expression of *FBLN2* following treatment with the demethylating drug, 5azaDC suggesting methylation is capable of reducing *FBLN2* expression.

FBLN2 (*fibulin 2*) is located at 3p25.1, part of the larger 3p region commonly deleted in many cancers (Pandis *et al*, 1997; Kok *et al*, 1997). Currently three isoforms have been demonstrated at the locus, all of which are transcribed from one of two CpG island regions. This study has analysed the larger of the CpG islands identified by the MIRA probes, which consists of 594bp and contains 84 CpGs and is situated across the transcription start of two of the isoforms. Expression analysis covered both of these isoforms.

The FBLN2 protein is a 126kDa secreted protein that forms a homotrimer and resides in the extracellular matrix (ECM) where it is thought to act as a scaffold protein, binding various ECM proteins such as fibrillin, type IV collagen, fibronectin, laminins, agrecan and versican (Argraves *et al*, 2003). It is part of a family of six FBLN proteins that appear to have roles in development (Argraves *et al*, 2003).

Frequent hypermethylation of *FBLN2* has previously been described in childhood ALL at a frequency of 56% in B-ALL and 17% in T-ALL (Dunwell *et al*, 2009) and downregulation has been observed in both breast cancer cell lines and primary tumours (Yi *et al*, 2007). The study by Yi *et al* (2007) demonstrated loss of the *FBLN2* mRNA and protein in breast cancer cell lines and significant loss of the FBLN2 protein in 14/17(82.4%) primary breast tumours. In addition, the authors were able to demonstrate reduced cell migration and invasion when

re-introducing *FBLN2* into non-expressing cell lines. No differences were observed in growth rate or binding to other ECM proteins, suggesting *FBLN2* most likely has a role in breast cancer progression rather than initiation (Yi *et al*, 2007). The frequency of methylation observed in this chapter is much lower than the level of downregulation observed in the study by Yi *et al* (2007), suggesting other mechanisms play a role in downregulating this gene in breast cancers. The COSMIC database lists one mutation in *FBLN2*, detected in ovarian cancer (out of the only ovarian cancer sample analysed), no mutations were identified in the 11 breast cancer samples analysed nor any of the other cancer types, bringing the total number of samples analysed to 45 (table 3.5). This suggests other mechanisms are likely to be responsible for downregulation of *FBLN2* and as it is situated on the commonly deleted chromosome 3p region, tumour specific somatic deletion could cause the loss of at least one allele. Hypermethylation and downregulation has also been observed in other members of the FBLN family, *FBLN3* (*EFEMP1*) has shown reduced expression at the mRNA level and methylation in breast cancer cell lines (Sadr-Nabavi *et al*, 2009). Frequent methylation was also observed in primary breast tumour samples coupled with reduced expression on the protein level (Sadr-Nabavi *et al*, 2009). *FBLN1*, *FBLN4* and *FBLN5* have been shown to be downregulated in prostate cancer (Wazlinski *et al*, 2007).

These results add to the notion that *FBLN2* hypermethylation is a common event in many cancers and that it often results in downregulation. Coupled with other results that have demonstrated the effects of loss of *FBLN2* on tumourigenic properties in breast cancer cells (Yi *et al*, 2007), *FBLN2* is an excellent candidate gene for further analysis, as are the other members of the *FBLN* gene family.

3.5.5 *SALL1*

Methylation was identified in 63% of breast tumours analysed. This study has also shown the methylation in breast tumours to be tumour specific and capable of reducing expression in breast cancer cell lines. Frequent methylation has also been demonstrated in prostate tumours (35.7%) and colorectal tumours (83%), although the methylation in their respective normal tissues (25% and 38% respectively) was notably higher than the small amount observed in normal breast tissues (5%).

SALL1 (*sal-like 1 (drosophila)*) is a three exon gene located at 16q12.1. It has a CpG island consisting of 5064bp with 366 CpGs. It encodes for a 140 kDa protein that functions as a zinc finger transcriptional repressor. Mutations within *SALL1* cause Townes-Brock syndrome, a disorder characterised by imperforate anus, abnormally shaped ears and hand malformations, among other developmental malformations, suggesting *SALL1* transcriptional activity is required at particular stages of development.

SALL1 involvement in cancer is also likely to be due to transcriptional activity, it has recently been demonstrated that *SALL1* is able to promote angiogenesis via activation of *VEGF-A* (Yamamoto *et al*, 2010). Hypermethylation has previously been shown in two types of leukaemia, ALL (Kuang *et al*, 2008) and CLL (Tong *et al*, 2010). *SALL1* also resides within a region of LOH in multiple cancers, including breast, prostate, ovarian and retinoblastoma (Argos *et al*, 2008; Chin *et al*, 2007). The COSMIC database lists three different mutations in three different ovarian cancer samples, each a single base substitution resulting in an amino acid change, all of which predicted to be damaging by PolyPhen2 and 2 of which predicted to be intolerant by SIFT. Other cancers, including breast, have been analysed but no additional mutations identified (table 3.5).

The results from this study suggest that SALL1 is frequently methylated in breast and prostate cancer and due to its known functions, LOH and implications in other cancers, it's downregulation as a result of hypermthylation is likely to be biologically relevant in the progression of breast cancers.

3.5.6 Conclusions

The aims of this study were to identify frequently methylated genes in breast cancer cell lines by use of the MIRA pull down assay coupled to a human CpG island array and use this data to identify frequently methylated genes in sporadic ductal breast cancers. The study was successful in identifying frequently methylated genes in breast cancer cell lines, identifying 16 genes that showed frequent methylation in cell lines and no methylation in a normal breast sample, nine of which also showed frequent methylation in primary tumour samples. However, 32 genes were analysed from the array initially, meaning only a 50% success rate of genes that showed methylation when analysed by CoBRA in cell lines i.e. a 50% false positive rate. There are multiple reasons why this may be the case. The MIRA assay itself is likely to have some degree of intrinsic false positive rate from non-specific pull-down. This may be altered and improved by optimisation of the pull-down reaction. However the screening methodology and subsequent laboratory confirmation are the influences most heavily affecting this high false positive rate. The regions analysed by CoBRA did not necessarily cover the region in which the positive probes were situated, instead, because the aim was to identify functionally relevant genes, CoBRA primers were designed upstream of and around the transcription start site. Therefore, it may be that the microarray probes that detected tumour specific methylation at a specific region were not within the region assessed by CoBRA. The regions that were chosen to be assessed by CoBRA may have either not contained any probes or may have been negative for enrichment. Similarly, in the cases where

methylation was present in the selected region in both the tumour and the normal sample, the positive enriched probes may again sit outside of the probed area. Nevertheless, the regions selected for study normally represent the most relevant region in respect to identifying epigenetic gene silencing, which was the object of the study and was successful. However, important regions may have not been enriched for in the MIRA assay or overlooked in the array analyses, providing false negatives. Additionally, the CoBRA digest screen used will only provide a positive result if the methylated CpGs reside within a BstUI restriction site, it is therefore possible that methylation could be present, resulting in enrichment during the MIRA pull-down but be missed by the CoBRA assay. This could be corrected for by analysing each candidate region by bisulphite sequencing opposed to CoBRA. In this study CoBRA was chosen for two reasons; firstly it is an accepted, fast and efficient method for accessing methylation in a region, while bisulphite sequencing is a costly and lengthy process, and secondly to detect an expansive level of general tumour specific methylation rather than small amounts of possibly non-specific methylation. CoBRA is an effective methodology for both of these aims.

For this type of study, where cell lines were used as the starting material, it should be taken into account that cell lines are known to demonstrate increased methylation both in terms of the regions demonstrating methylation and the depth of methylation at given regions (Smiraglia *et al*, 2001). The lack of methylation in primary samples compared to breast cancer cell lines for some genes is therefore likely to be due to the cell lines having acquired additional methylation.

One advantage of using a CpG island microarray is that many CpG islands are being assessed, including those that relate to miRNAs and other chromosomal locations, giving a global assessment of methylation. In contrast, one downside to the CpG island microarray is that

probes are designed to represent general areas and so a larger CpG island has more probes. Therefore, if, as in this study, selection criteria take into account the number of positive probes for a given feature, which is a sensible and logical approach to whittling down the number of candidate genes, there is an intrinsic bias to genes with large CpG islands. Nevertheless, the success of this study, and others, demonstrates the validity of this methodology but it should be noted that other methodologies may be equally as successful or may identify alterative candidates missed by this screen or not so strongly represented by this breast cancer tumour cohort.

Additionally, the CpG island microarray identified a subset of miRNAs that were potentially methylated in breast cancer cell lines, including some good candidates that have previously been shown to be involved in breast and/or other cancers. Since this was neither an aspect which we normally study nor one of the main aims of this chapter it was not extensively investigated, but does present as something that would be interesting to do in the future.

Overall, this study was successful in identifying five genes that were methylated in breast tumours, and in some cases other tumours also. The genes also represented good candidates to be downregulated in cancer. Only primary mRNA levels for *EMILIN2* were assessed in this study, however, it would be nice to see the same data for the other genes and also to see downregulation at the protein level. It would also be advantageous to demonstrate tumour suppressor properties for those of the five genes that have not previously been shown in breast cancer cells. All of the findings of this study would also benefit from repeat confirmation in other, larger cohorts of the relevant cancer type, which would hopefully highlight the strongest candidates and remove any cohort specific positive results. In conclusion, this study has identified 5 good candidates for involvement in sporadic breast cancer, of which the methylation status of one, *EMILIN2*, also associates with disease outcome.

This work has now been published under Hill *et al*, 2010. *Molecular Cancer* **9**(51). A copy is provided at the back of the thesis.

CHAPTER 4

ILLUMINA INFINIUM METHYLATION ANALYSIS OF SPORADIC BREAST TUMOURS

4.1 ABSTRACT

Use of the Illumina Infinium HumanMethylation27 array on a panel of clinically well characterised sporadic ductal breast tumours and non-cancerous breast tissue has identified 263 candidate tumour specific hypermethylated genes (291 CpG loci probes) within sporadic tumours.

Hierarchical clustering based on methylation levels splits the samples into three groups characterised by high, medium or low level methylation of the 291 probes ($p<0.0001$). Certain clinical features, most notably relapse status and oestrogen and progesterone receptor (ER and PR) status, clustered within these groups. ANOVA analysis also confirmed overall methylation levels differ significantly between relapse ($p=0.035$), oestrogen and progesterone positive receptor status ($p=0.001$) and lymph node positive tumours ($p=0.042$).

To identify individual genes associated with particular clinical features, Fisher's exact tests (with FDR correction) were performed, identifying genes where methylation status correlated with relapse (n=9), lymph node (n=2) and ER and PR receptor status (n=19). Among these were *RECK* and *ACADL*, both of which are novel methylated genes in breast cancer and have shown reduced expression in methylated primary tumour samples. In both cases, methylation status also associated with disease free survival ($p=0.0009$).

4.2 INTRODUCTION

Work in chapter 3 focused on identifying novel genes hypermethylated in breast cancer by using a genome wide approach involving the initial assessment of breast cancer cell lines. The use of primary tumours as opposed to cell lines should be more representative of epigenetic aberrations within the endogenous tumour, however, the increased heterogeneity of primary tumours requires the use of a larger sample number when looking for relatively less common events. As technology has improved it has become possible to use a lower amount of starting material and analyse more samples at once for a relatively low cost. This has given rise to a plethora of large genome wide methylation profiling studies (Christensen *et al*, 2011; Noushmehr *et al*, 2010; O'Riain *et al*, 2009; Martín-Subero *et al*, 2009a). This is particularly useful when it is desirable to look at DNA methylation in relation to clinical characteristics. There are many factors that can affect clinical and prognostic outcomes in breast cancer, for example, patient age, menopause status, lymph node status, parity, dietary factors, hormone exposure and receptor status of the oestrogen, progesterone and ERBB2 receptors. Survival data from patients can also be used in these studies to identify factors that relate back to either overall survival (OS) or disease free survival (DFS).

In an attempt to profile genome wide methylation changes in sporadic breast tumours in relation to clinical features and to identify additional novel hypermethylated genes, this study has utilised the Illumina Infinium HumanMethylation27 array on a cohort of clinically well characterized sporadic ductal breast tumours.

4.3 AIMS

1. To use the Illumina Infinium assay to create a list of hypermethylated genes within this cohort of breast tumours and cluster potential hypermethylated candidates to identify groups with similar methylation profiles
2. To compare overall methylation levels between tumours with different clinical characteristics
3. To identify individual genes where hypermethylation associates with specific clinical features
4. To confirm selected hypermethylated genes associated with either clinical features or highly methylated/multiple positive probes

4.4 RESULTS

4.4.1 Use of Illumina Infinium HumanMethylation27 BeadChip to measure genome wide DNA methylation changes

This study used the Illumina Infinium HumanMethylation27 Bead Chip on a panel of 39 clinically well characterized sporadic breast tumours and 4 further sporadic tumours along with matched normal breast tissue DNA from these 4 patients, see table 4.1 for clinical details of the 39 sporadic samples.

The Infinium array assesses methylation levels of 27,578 single CpG dinucleotides within 14,475 consensus coding sequencing (CCDS) genes. This results in an average of two probed CpG dinucleotides being measured for each gene, unless genes are known to be cancer related or imprinted, in which case the number of probed CpG dinucleotides raises to between 3 and 20 per gene. In addition, 110 miRNAs are assessed using 254 probed CpG dinucleotides. Methylation levels are given as a quantitative numerical value (β value) ranging between 0 and 1. This value represents the ratio of methylated signal to unmethylated signal, where 0 indicates no methylation and 1 indicates complete methylation. The basis of Infinium CpG assay design was discussed in chapter one, see section 1.3.7.3 for more information.

Scatter plots of all β values for a single sample against another demonstrates the amount of correlation between the two samples. Repetition of the Illumina assay on the same DNA sample shows a very high level of correlation, $R^2 = 0.9955$, showing that the assay is highly reproducible (figure 4.1). It should be noted that the repetition sample is an RCC tumour done at the same time and on the same chip (each chip accommodating 12 samples) as some of the breast samples used in this study and not DNA of one of the tumours analysed in this study. This scatter plot has been provided courtesy of Dr. C.J. Ricketts. The four normal breast DNA

Table 4.1_Clinical features of the 39 sporadic breast tumours used in the array.
This table shows clinical features for the 39 tumours used in the array. Parameters cover; age at diagnosis (years), tumour size (mm), grade (SBR), lymph node status (positive or negative), menopause status (positive or negative),ER status (positive or negative), PR status (positive or negative), ERBB2 overexpression status (overexpressed or not overexpressed), relapse status (positive or negative) and disease free survival (days since diagnosis until relapse event). 1=positive 0=negative.

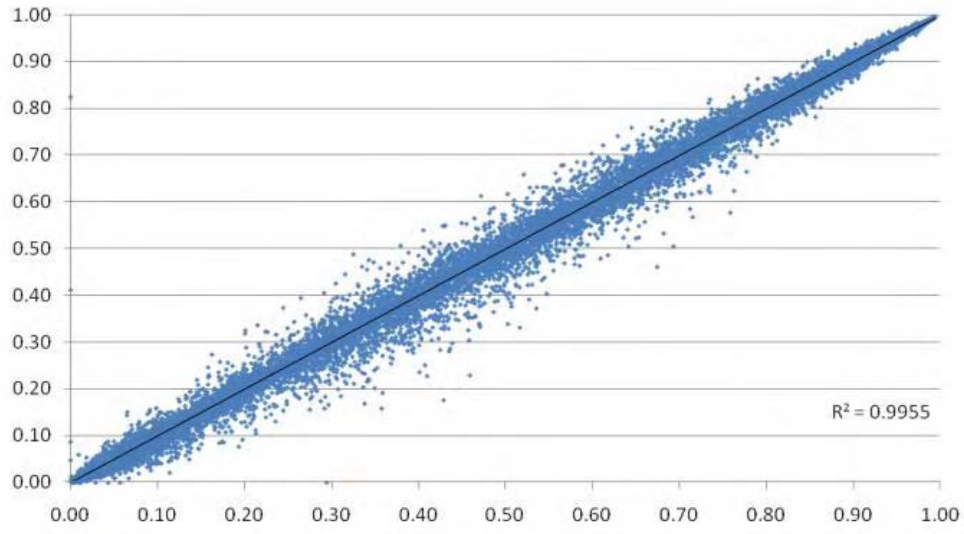


Figure 4.1_Correlation of a replicated sample
This graph shows the correlation between the array results for one sample run twice on the Infinium array. This graph represents an RCC tumour duplicated on the same run as some of the breast cancer samples. The graph has been provided by Dr. Chris Ricketts

samples showed high levels of correlation between them, ranging from R^2=0.9302 (normal 1 versus normal 4) to R^2=0.974 (normal 2 versus normal 4) (figure 4.2), suggesting all normal samples have a similar spread of DNA methylation and little variation, making suitable controls to be used for further analysis. As expected, the correlation between tumour normal pairs showed a greater level of variation, particularly in pairs 1 and 2 which showed R^2 values of 0.7243 and 0.8337 respectively, whilst pairs 3 and 4 showed less variation (R^2=0.9381 and R^2= 0.909 respectively) (figure 4.3). These results show that pairs 1 and 2 have much higher levels of DNA methylation alterations than pairs 3 and 4 which may be indicative of DNA methylation deregulation playing a larger role in the onset or progression of tumours 1 and 2. This may be due to DNA methylation changes acting as a driving factor within these two tumours as oppose to pairs 3 and 4 which may have alternative overriding factors, such as genetic events like mutation. Unfortunately extensive clinical details of these tumour normal paired samples is unavailable and therefore cannot be analysed for any obvious differences that may associate with these differential levels of DNA methylation changes.

4.4.2 Identification of candidate hypermethylated loci

The earliest Infinium published work and Illumina literature based on said work has shown a β-value below 0.25 as being representative of unmethylated DNA (Martín-Subero *et al*, 2009a). For this reason, all probes showing $\beta > 0.25$ in any of the four normal samples were removed (n=11,211). The remaining probes (n=16,367) were analysed in two different ways; probes were removed if less than 20% exhibited either (a) a β-value above a certain cut off (cut-offs ranged from 0.4 to 0.6), or, (b) a difference in tumour and mean normal β-values above a certain cut-off (cut-offs ranged from a difference of 0.3 to 0.5). The number of probes detected as being hypermethylated by these methods ranged from 69 probes representing 64 genes to 606 probes representing 524 genes, selection criteria being a β-difference cut-off of

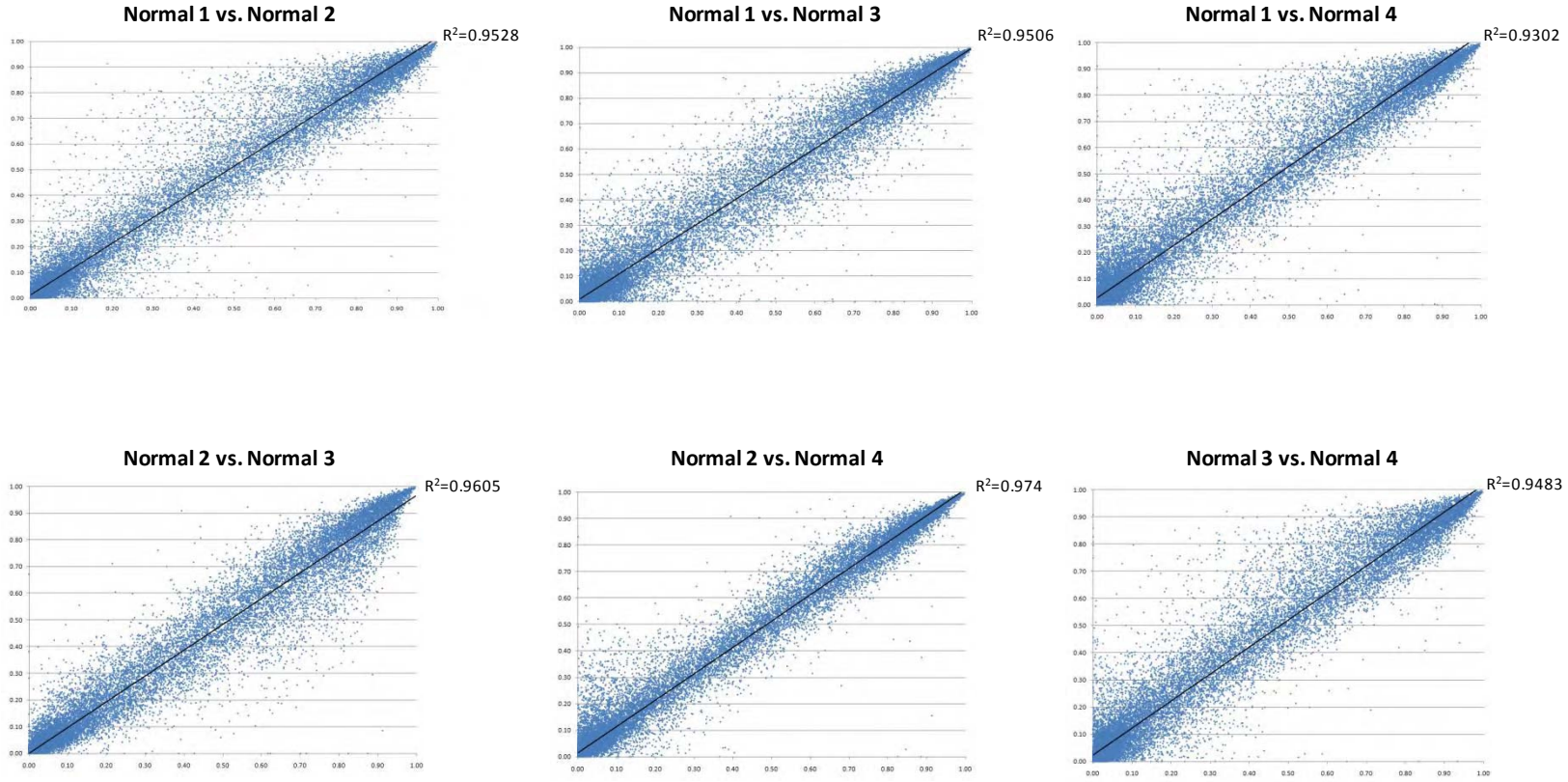


Figure 4.2_β value correlation between normals.
Scatter plots are shown for comparisons of normal samples.
R^2 = correlation coefficient.

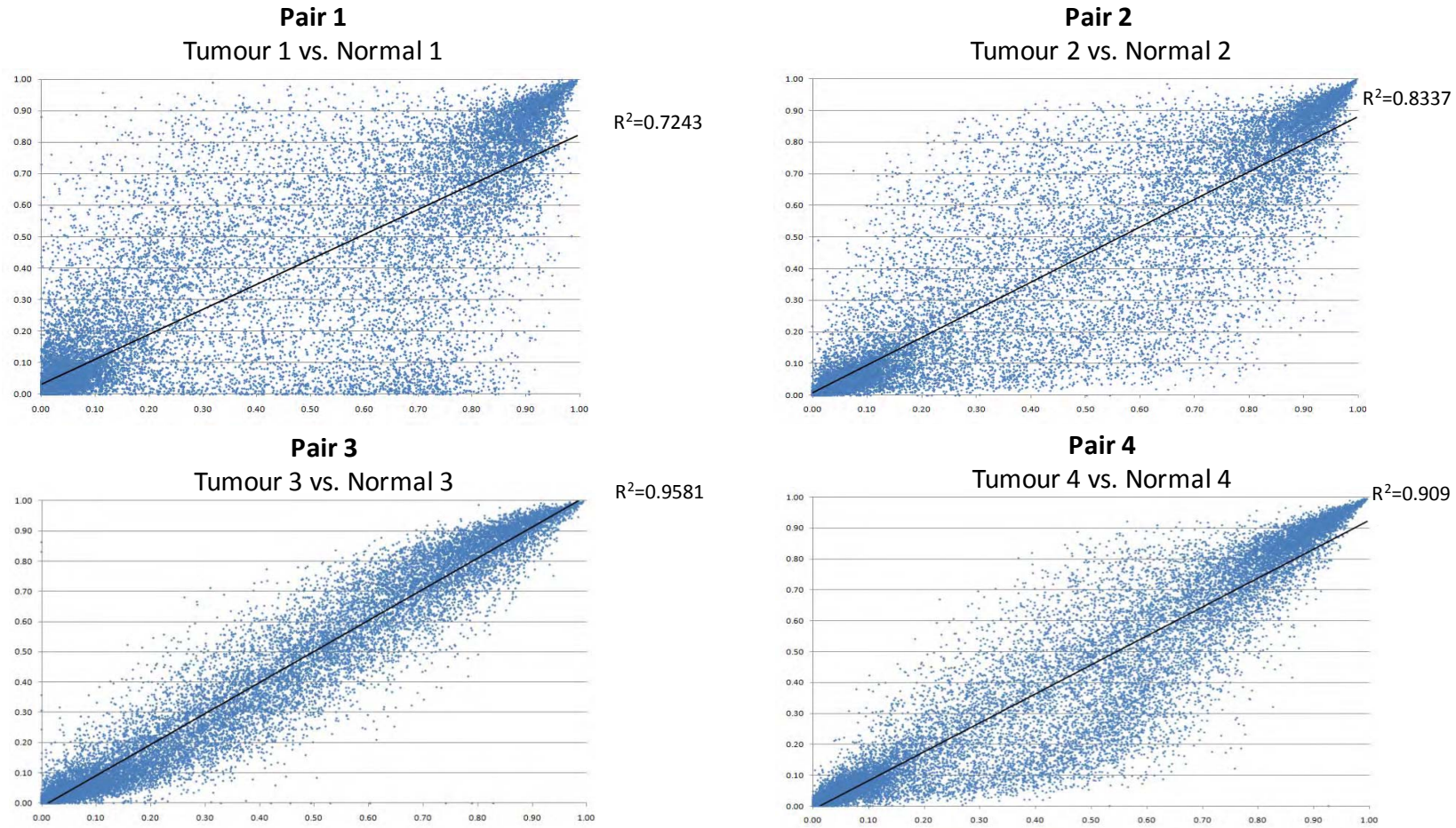


Figure 4.3_β value correlation between tumour normal paired samples.
Scatter plots are shown for each of the four tumour normal pairs used in the array. R^2 = correlation coefficient

0.6 and a β-value cut off of 0.4 respectively (figure 4.4). To select a workable number of genes and also take into account the unfairness of a direct β-value cut-off on samples very close to the cut-off value, the lists for β-value >0.5 and β-difference >0.4 were combined, giving 153 genes that were in both lists, 81 only in β-value >0.5 and 30 only in β-difference >0.4 (figure 4.4) and an either/or list of 291 probes representing 264 genes, all genes are shown in appendix B.III. All genes on the X chromosome or known to be imprinted were removed from the analysis. The either/or list of 291 probes/264 genes was used for further analysis.

4.4.3 Array validation

The list contained a number of genes already known to be methylated in breast cancer, for example, *APC, CCND2, GSTP1* and *PTPRO* (Jin *et al*, 2001; Evron *et al*. 2001; Esteller *et al*, 1998; Ramaswamy *et al*, 2009) suggesting the list produced was suitable for further analysis, however, to confirm this further, the list was tested in two ways; (i) individual β-values were assessed, and (ii) the function of the genes within the list were considered.

4.4.3.1 Assessment of individual β-values

Three genes were chosen from the list previously shown to have an association with cancer, *SFRP5, CLDN6,* and *SIPA1* (Veeck *et al*, 2008a, Tsuonda *et al*, 2009 and Gaudet *et al*, 2009), to be used for β-value validation. A region encompassing the significant CpG being measured by the Infinium probe was clone sequenced for DNA samples showing β-values considered to be methylated and unmethylated.

For *SFRP5*, samples were sequenced with β-values of 0.54 (sample 304), 0.01 (sample 1116) and 0.04 (sample 43), all of which showed methylation at the probe CpG to be of a similar

β-value cut off

	Probes	Genes	X-chromosome	Imprinted
0.4	606	524	4	12
0.5	261	241	1	6
0.6	69	66	0	2
0.7	6	6	0	0

β-value difference cut off

	Probes	Genes	X-chromosome	Imprinted
0.3	587	501	3	14
0.4	207	189	1	4
0.5	41	40	0	1

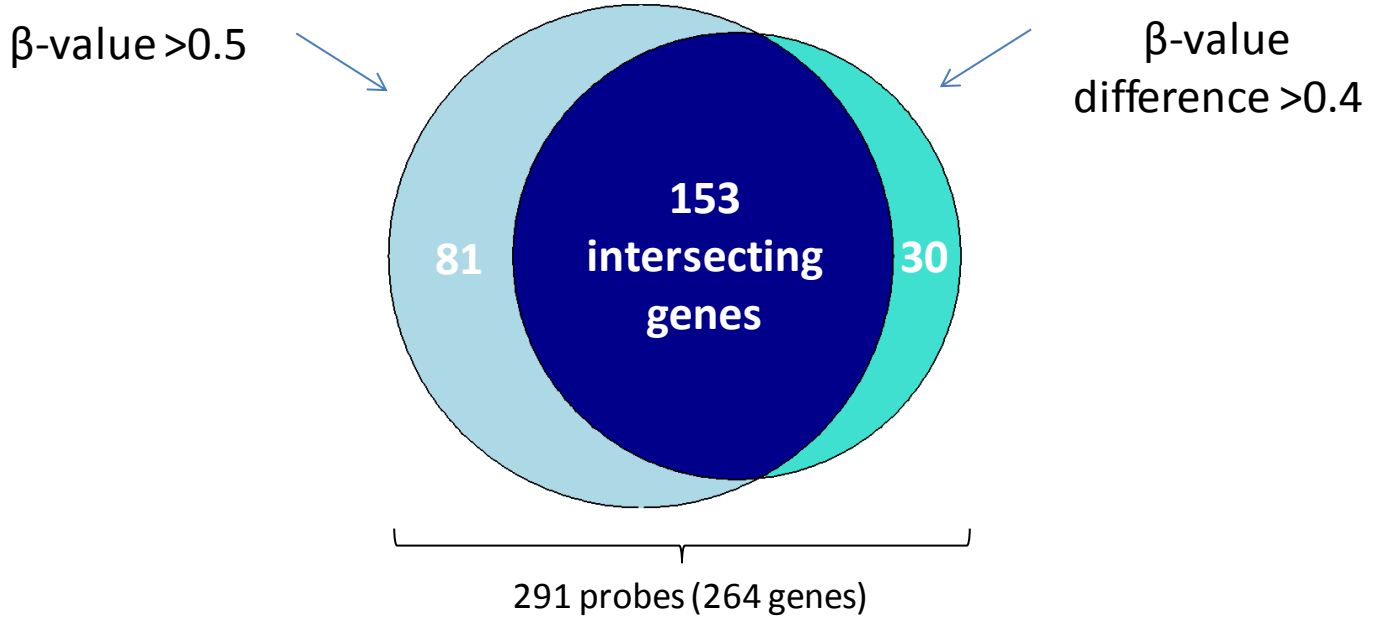


Figure 4.4_ Identification of hypermethylated loci.
The numbers of probes and genes for each selection criteria are shown, along with the numbers of X-chromosome genes and imprinted genes. A Venn diagram representative of combining β-value >0.5 and β-difference >0.4 is also shown.

level to the β-value, 0.67, 0.14 and 0.00 respectively, and comparable methylation indexes (MIs) (76.0%, 1.93% and 2.4% respectively) (figure 4.5). The same trend was observed for *SIPA1*, where high β-values (0.53 (sample 1265) and 0.70 (sample 207)) showed both high probe values and MIs (figure 4.6) whilst the remaining low β-value sample (sample 43, β-value = 0.13) showed a low probe value (0.00) and higher MI (20.0%). *CLDN6* results show a similar trend of high β-values corresponding to high probe values and MIs (samples 375 and 198, figure 4.7) with the only slight exception of sample 1178, showing a low β-value and a low probe value with a higher MI (33.9%). This is due to relatively frequent methylation 5' (upstream) of the CpG being assessed by the Infinium assay (figure 4.7).

Results for *SFRP5*, *SIPA1* and *CLDN6* demonstrate β >0.5 and β <0.25 represent methylation and no methylation at the CpG respectively but also that β-values appear to be indicative of methylation status of the immediate surrounding region. These results show the list criteria to be a reliable indicator of methylation.

4.4.3.2 Functional assessment of gene list

A good indication of whether the list produced is acceptable to do further work on is to assess the types of genes within the list.

Funcassociate (http://llama.mshri.on.ca/funcassociate/) is a freely available program that identifies gene ontology (GO) terms within a set list of genes and determines whether any particular GO terms are over represented. When performed on this dataset, 21 GO terms are identified as being enriched, many of these GO terms are involved with development, for example, 84 genes, 16 genes and 33 genes are associated with the GO terms 'developmental process', 'pattern specification process' and 'multicellular organismal development' respectively.

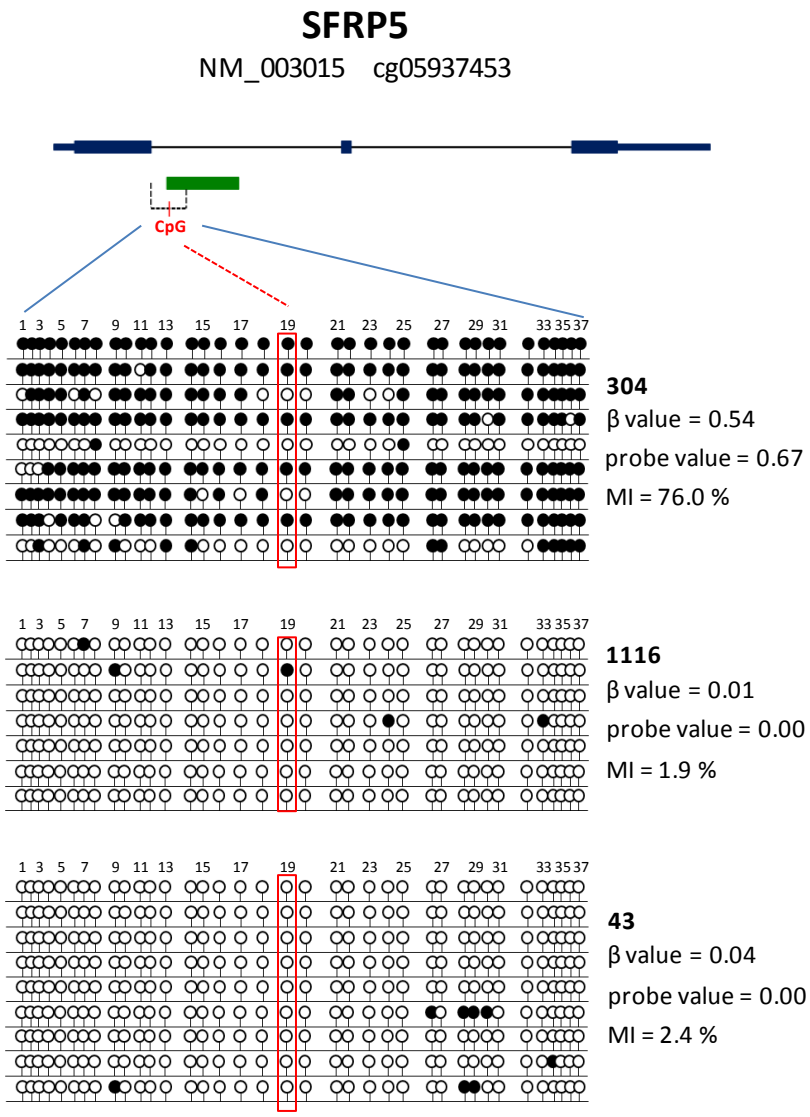


Figure 4.5_β-value validation for *SFRP5*.
Clone sequencing results are shown for *SFRP5* (NM_00315) for the region surrounding Illumina ID (ILMNID) cg05937453. A gene schematic is shown with exons in blue, introns as straight black lines and CpG islands in green. The schematic illustrates the location of the region being sequenced (black dotted line) and the CpG being assessed by the Infinium assay (highlighted in red). Sequencing results are shown for three samples (304, 1116 and 43) with β-values, probe values (value between 0 and 1 as determined by clone sequencing for CpG being assessed in Infinium assay) and methylation index (MI) (given as a percentage of methylated CpGs out of total CpGs analysed). For sequencing results, each black line represents a single allele, black and white circles represent methylated and unmethylated CpGs respectively.

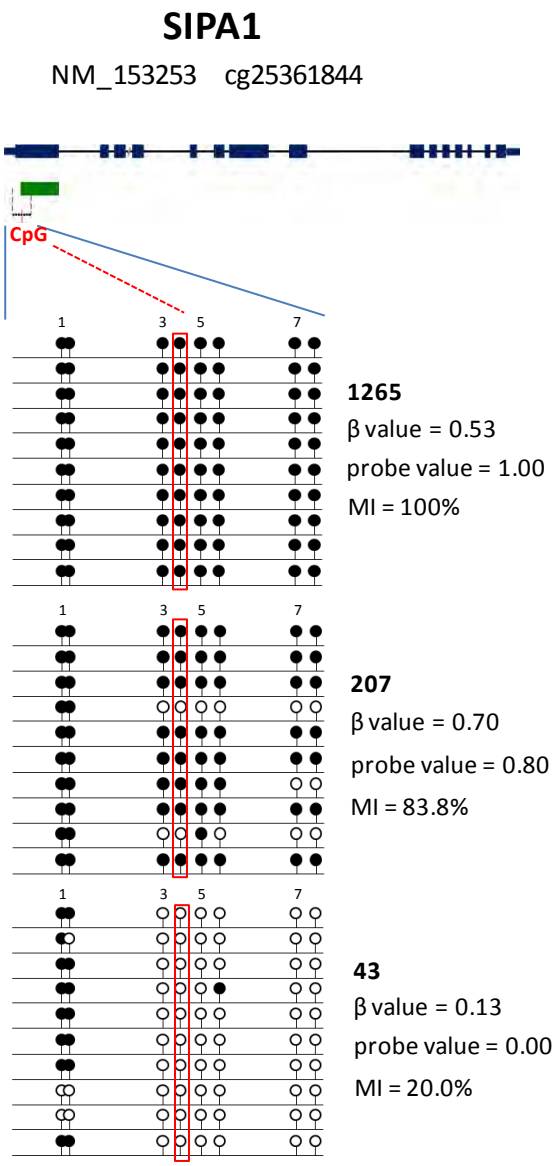


Figure 4.6_β-value validation for *SIPA1*.
Clone sequencing results are shown for *SIPA1* (NM_153253) for the region surrounding Illumina ID (ILMNID) cg25361844. A gene schematic is shown with exons in blue, introns as straight black lines and CpG islands in green. The schematic illustrates the location of the region being sequenced (black dotted line) and the CpG being assessed by the Infinium assay (highlighted in red). Sequencing results are shown for three samples (1265, 207 and 43) with β-values, probe values (value between 0 and 1 as determined by clone sequencing for CpG being assessed in Infinium assay) and methylation index (MI) (given as a percentage of methylated CpGs out of total CpGs analysed). For sequencing results, each black line represents a single allele, black and white circles represent methylated and unmethylated CpGs respectively.

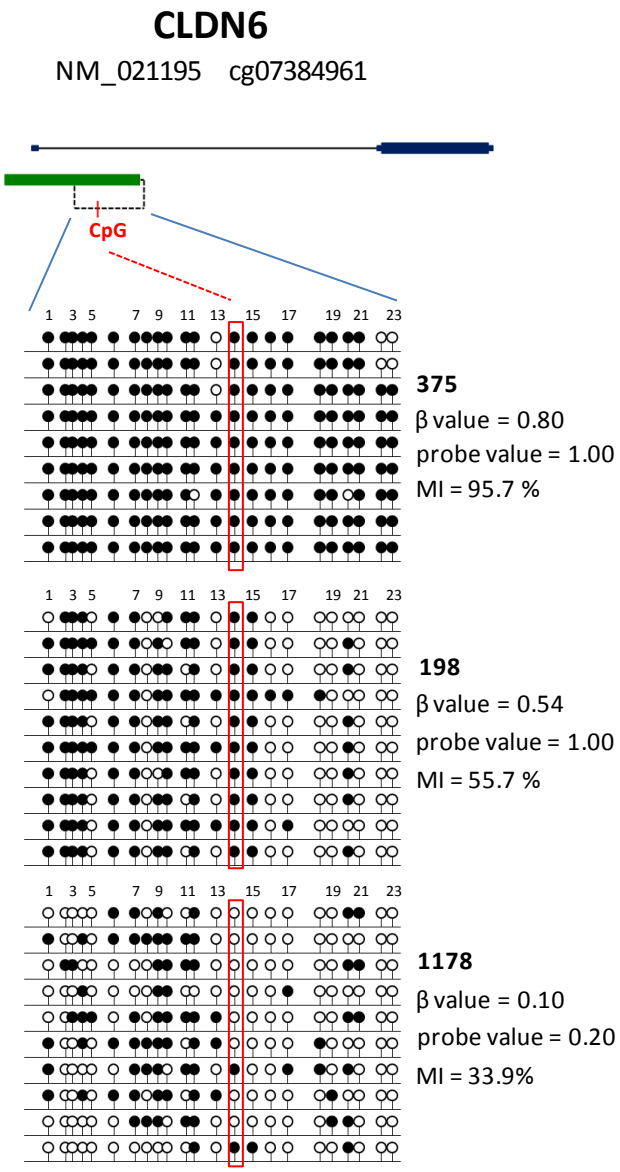


Figure 4.7_β-value validation for *CLDN6*.
Clone sequencing results are shown for *CLDN6* (NM_021195) for the region surrounding Illumina ID (ILMNID) cg07384961. A gene schematic is shown with exons in blue, introns as a straight black line and CpG islands in green. The schematic illustrates the location of the region being sequenced (black dotted line) and the CpG being assessed by the Infinium assay (highlighted in red). Sequencing results are shown for three samples (375, 198 AND 1178) with β-values, probe values (value between 0 and 1 as determined by clone sequencing for CpG being assessed in Infinium assay (highlighted in red) and methylation index (MI) given as a percentage of methylated CpGs out of total CpGs analysed. For sequencing results, each black line represents a single allele, black and white circles represent methylated and unmethylated CpGs respectively.

The Database for Annotation, Visualization and Integrated Discovery (DAVID) (http://david.abcc.ncifcrf.gov/) was used to assess functional properties of genes within the list. The functional annotation chart tool identified the presence of many genes related to areas of carcinogenesis: cell adhesion (24 genes), regulation of cell proliferation (21 genes), negative regulation of cell death (15 genes), cell migration (12 genes), regulation of cell cycle (10 genes) and tumour suppressors (6 genes) (figure 4.8). Relevant pathways were also identified through the DAVID pathway analysis tool; pathways in cancer (12 genes), cell adhesion molecules (7 genes) and the Wnt signalling pathway (6 genes) (figure 4.8).

Taken together, results from the β-value validation and assessment of genes within the list suggest the initial analysis of the β-values produces a suitable selection for further analysis. Further work on this list involved (i) looking at good candidate genes where breast cancer hypermethylation is a novel event and (ii) analysis of methylation with clinical features, both globally and individually.

4.4.4 Analysis of candidate genes

To analyse candidate genes from within the list prior to assessing clinical features, three genes were chosen each from those that were most highly methylated, and those that showed more than one probe within the list of 291.

4.4.4.1 Highly methylated genes

The use of two selection criteria means that a single sample can be considered hypermethylated in one of two ways. To simplify this for further analysis, samples were given a value of either 1 or 0 representing methylated and unmethylated respectively by either criteria. Probes could then be ranked in order of highest to lowest frequency of methylation.

DAVID functional chart analysis

Regulation of cell death; 26 genes (10.7%)					
APC	IHH	CCK	GSTP1	STAT5A	DDAH2
BDNF	POU4F1	CDKN2A	NEUROD1	KIT	EDNRB
CARD11	ACVR1	DCC	PAX7	WRN	FOXL2
GRIP2	RUNX3	RYR2	SFRP1	SST	SOC2
TNFRSF10D	BRAF				
Cell adhesion; 24 genes (9.9%)					
CD9	AEBP1	CDH9	IGFBP7	PDPN	MCAM
APC	CD34	CHL1	ICAM2	PCDHAC1	NCAM2
AJAP1	CNTNAP2	CLDN11	KIRREL2	PCDHGB7	RELN
AOC3	FLRT2	CLDN6	LAMA2	PCDHGA12	SCARF2
Regulation of cell proliferation; 21 genes (8.7%)					
IHH	BTG4	APC	CDKN2A	RUNX3	PDGFRB
KLF11	BTG3	BDNF	EDNRB	STAT5A	
CDX2	CLEC11A	CARD11	IGFBP7	KIT	
CCND2	CD9	PRKCQ	PLAU	SST	
Negative regulation of cell death; 15 genes (6.2%)					
APC	IHH	DDAH2	PAX7	SOCS2	GSTP1
BDNF	ACVR1	EDNRB	SFRP1	TNFRSF10D	NEUROD1
KIT			STAT5A	BRAF	
Cell migration; 12 genes (5%)					
POU4F1	CD34	APC	EDNRB	RELN	PDGFRB
DCC	ISL1	CCK	PLAU	KIT	
Cell-cell adhesion; 11 genes (4.5%)					
CLDN11	CD34	CLDN6	PPDN	PCDHGA12	PCDHGB7
NCAM2	CDH9	ICAM2	PCDHAC1	SCARF2	
Regulation of cell cycle; 10 genes (4.1%)					
	BTG4	CCND2	CDKN2A	STAT5A	
	BTG3	PKIA	RUNX3		
	APC	PRKCQ	SIPA1		
Tumour supressor; 6 genes (2.5%)					
RASSF2	APC	CDKN2A	DCC	KLK10	RECK

DAVID KEGG pathways

Pathways in cancer; 12 genes (5%)					
APC	DCC	GSTP1	RALBP1	STAT5A	KIT
CDKN2A	FZD10	HHIP	LAMA2	PDGFRB	BRAF
Cell adhesion molecules; 6 genes (2.5%)					
CD34	CLDN11	CLDN6	CNTNAP2	ICAM2	NCAM2
Wnt signalling pathway; 7 genes (2.7%)					
CCND2	APC	FXD10	SFRP2	SFRP5	SFRP4
SFRP1					

Figure 4.8._DAVID analysis of hypermethylated loci

This figure shows the genes involved in functions and pathways relevant to carcinogenesis as identified through DAVID chart analysis and KEGG pathway analysis respectively. The numbers of genes in each functional group/pathway is shown along with the percentage this accounts for out of total number of genes.

The top ten methylated probes represent eight genes, *INPP5B* (*inositol polyphosphate-5-phosphatase*), *ABCA3* (*ATP-binding cassette, sub-family A (ABC1), member 3*), *COX7A1* (*cytochrome c oxidase subunit VIIa polypeptide 1 (muscle)*), *C12orf34* (*chromosome 12 open reading frame*), *SST* (*somatostatin*), *ZNF154* (*zinc finger protein 154*), *C1orf114* (*chromosome 1 open reading frame 114*) and *GPC2* (*glypican 2*). Of these eight genes, all but *INPP5B* and *C12orf34* are assayed by the Infinium array within a 'true' CpG island (occasionally CpG probes lie outside of well-defined CpG islands, those that are within a well-defined CpG island are designated as being within a 'true' CpG island by Illumina) (figure 4.9). The top three of these genes, *ABCA3* (69.2%), *COX7A1* (69.2%) and *SST* (59.0%) were chosen to be analysed further.

Firstly, to determine if methylation was likely to have a functional impact, gene expression in breast cancer cell lines was determined pre and post 5-azaDC treatment. Re-expression was observed in *COX7A1* and *SST* but not *ABCA3*. Cell line methylation status was subsequently determined for *COX7A1* and *SST*, both of which showed methylation by CoBRA assay in re-expressing lines (figure 4.10). *COX7A1* and *SST* CoBRA assays were further used on tumour normal paired samples to assess the wider presence of methylation and tumour specificity in a separate cohort (figure 4.11). High overall frequencies of methylation were observed for both *COX7A1* and *SST*, 100% and 68.4% respectively. *COX7A1* showed increased methylation in 70.6% (12/17) of tumours compared with low levels of methylation present in corresponding normal tissue and a further 5 tumours showed equal amounts of low level methylation in both tumour and corresponding normal tissue. *SST* showed tumour-specific methylation in 31.6% (6/9) of pairs and increased methylation in a further 31.6% (6/19) of tumours compared with corresponding normal tissue. One tumour showed an equal amount of low level methylation in the tumour and normal. No methylation was observed in the remaining tumours.

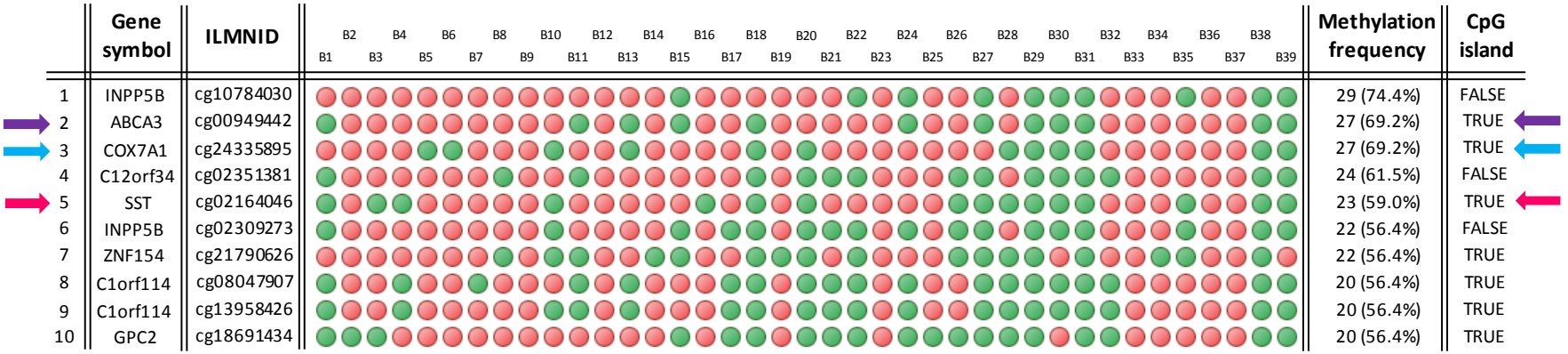


Figure 4.9_Top ten highly methylated probes.

The figure shows the ten most methylated probes according to selection criteria ranked in order of methylation frequency. Methylation status is indicated for each of the 39 sporadic tumours (B1-B39) with either a red circle (methylated) or green circle (unmethylated). The frequency of methylated samples is also shown and whether the CpG being assessed resides in a true CpG island (TRUE) or not (FALSE). Purple, blue and pink arrows are aligned to the three top genes with true CpG islands.

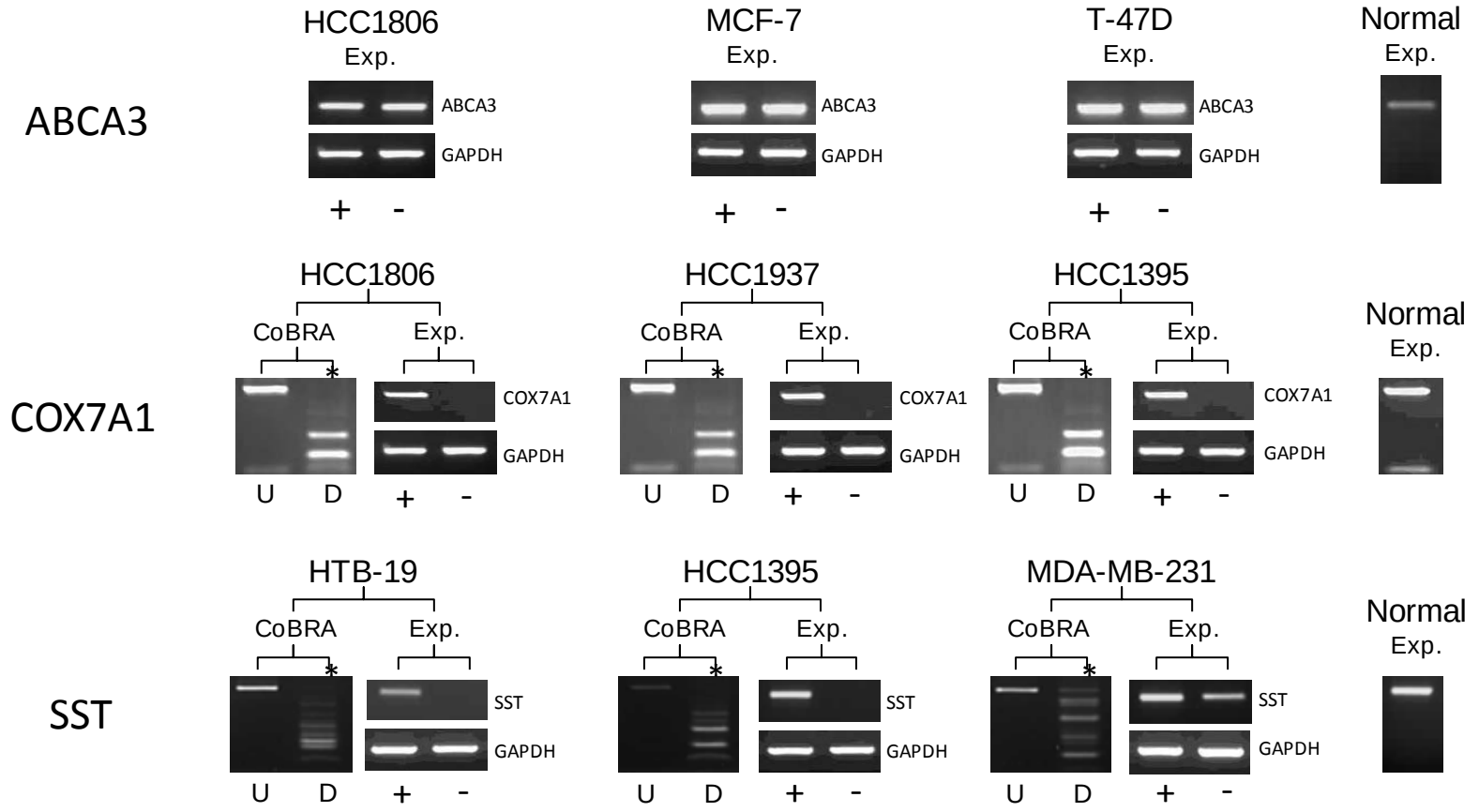


Figure 4.10_Cell line CoBRA and expression analysis for highly methylated genes.
Expression results (Exp.) are shown for *ABCA3*, *COX7A1*, *SST* and *GAPDH*. Three cell lines are shown in each case, all with (+) and without (-) treatment with 5-azaDC. Expression for normal breast tissue RNA is also shown. Where re-expression was seen (*COX7A1* and *SST*) CoBRA results are also shown for the corresponding cell line. Undigested products (U) are run next to digested products (D). * indicates methylated samples

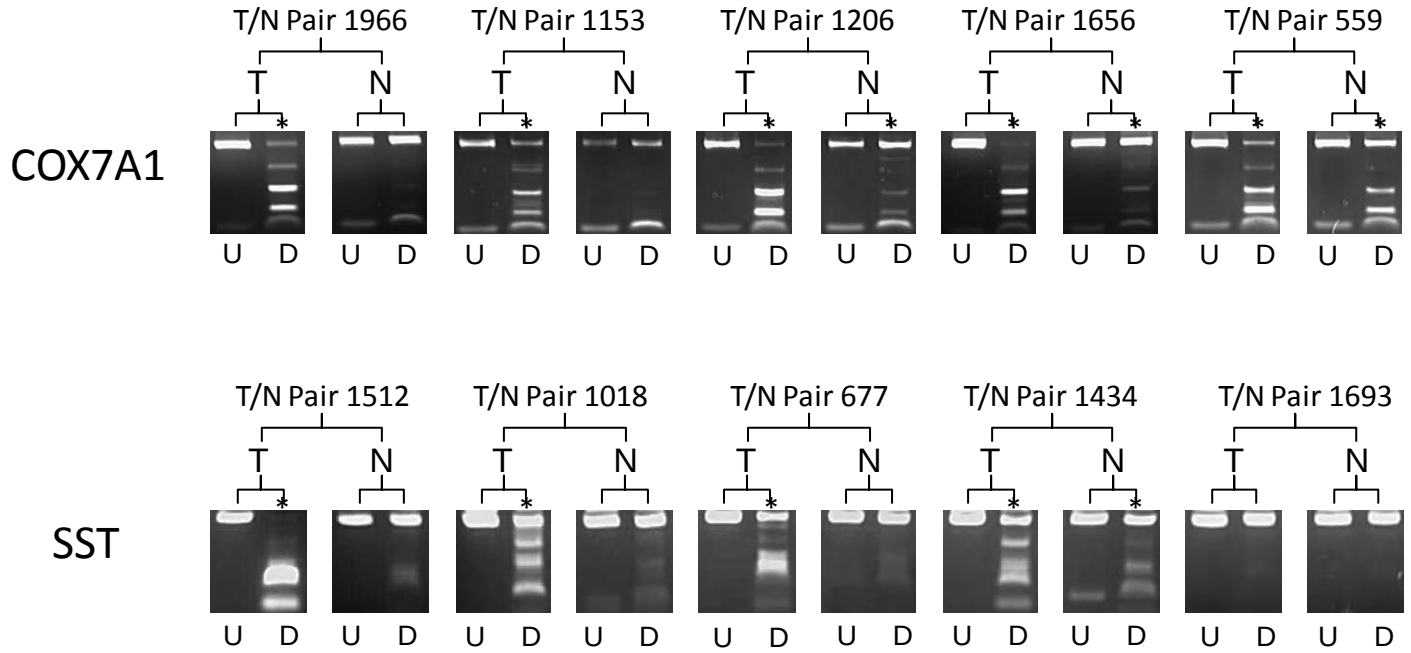


Figure 4.11_CoBRA tumour/normal pair results for *COX7A1* and *SST*.
CoBRA results are shown for five tumour (T) and normal (N) paired samples with undigested product (U) run adjacent to digested product (D). * indicates methylated samples

4.4.4.2 Genes with multiple probes

Twenty four genes were represented by more than one probe within the list of 291, four of which have previously been shown to be methylated in breast cancer, *APC, CCND2, GSTP1*, and *PTPRO* (Virami et al, 2001; Evron et al, 2001; Esteller et al, 1998; Ramaswamy et al, 2009). Of the remaining twenty genes, three, *LAMA2*, *CDKL2* and *ZNF154*, were selected for further analysis because of multiple probes (figure 4.12). A further four genes present within this list were analysed later in this chapter due to an association with clinical features (see 4.4.5.3).

Both *CDKL2* and *ZNF154* showed re-expression following treatment with 5-azaDC whilst *LAMA2* showed uniform expression, subsequent CoBRA analysis of *CDKL2* and *ZNF154* in breast cancer cell lines showed methylation in re-expressing lines (figure 4.13). Tumour normal pairs demonstrated high tumour methylation levels for *ZNF154* (87.5%), of which the majority (56.2%; 9/16) showed tumour specific methylation. Of the remaining samples, 25% (4/16) showed increased methylation in the tumour compared to corresponding normal, 6.3% (1/16) showed equal methylation in the tumour compared to normal and 12.5% (2/16) were unmethylated. High levels of methylation were also observed for *CDKL2* (56.3%; 9/16). Tumour specific methylation was observed in 25% (4/16) samples, whilst 12.5% (2/16) of samples showed increased methylation in the tumour compared to normal, 18.8% (3/16) showed equal methylation in the tumour compared to normal and 43.8% (7/16) were unmethylated (figure 4.14).

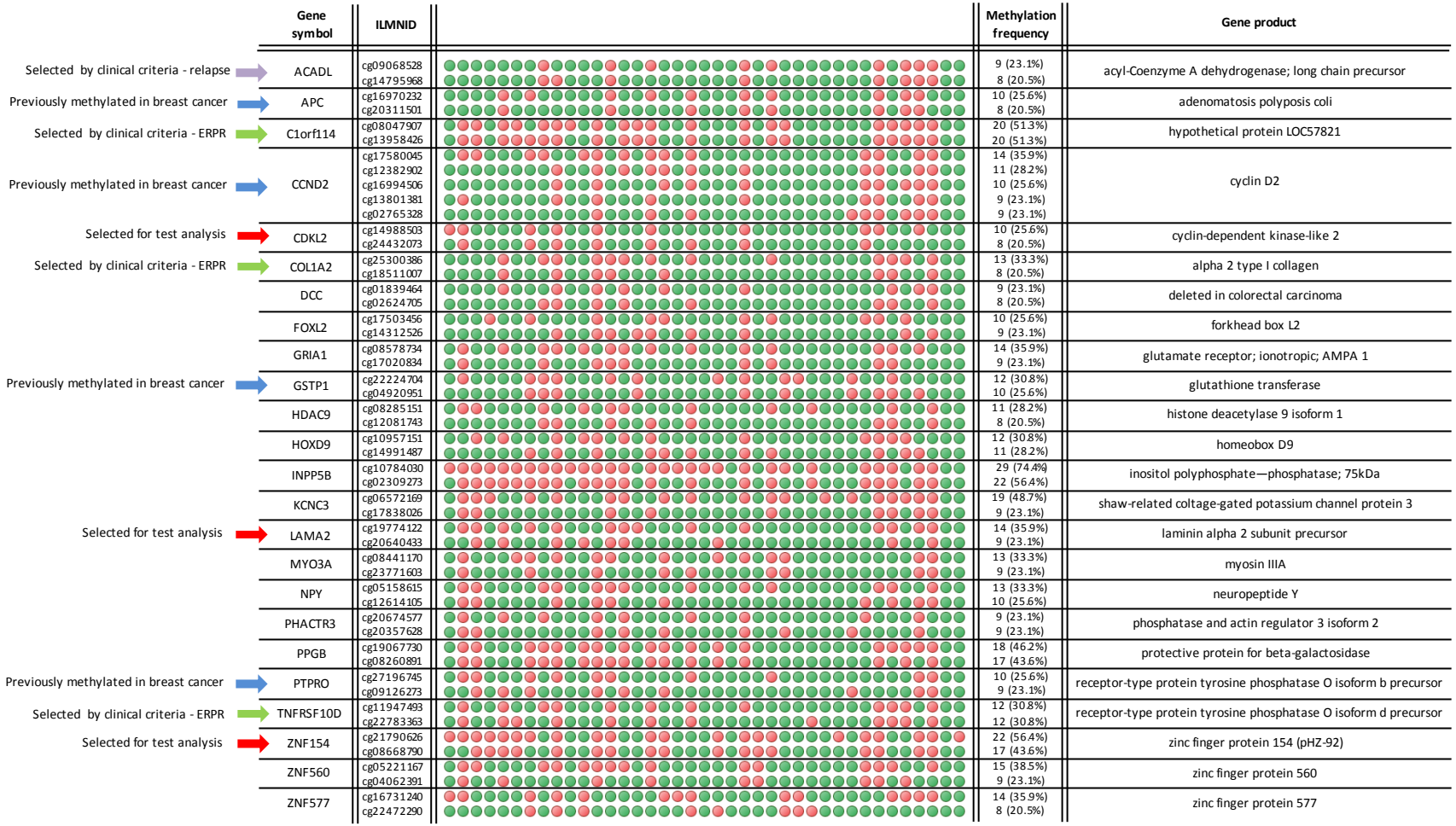


Figure 4.12_Genes with multiple probes.

The methylation status for all genes with multiple probes is represented by red and green circles for methylation and no methylation respectively. Methylation frequencies and gene product names are also shown. Genes chosen to be analysed because of multiple probes are highlighted with red arrows, genes previously shown to be methylated in breast caner are highlighted with blue arrows and purple and green arrows highlight genes analysed because of an association with clinical characteristics (see 4.4.5.3).

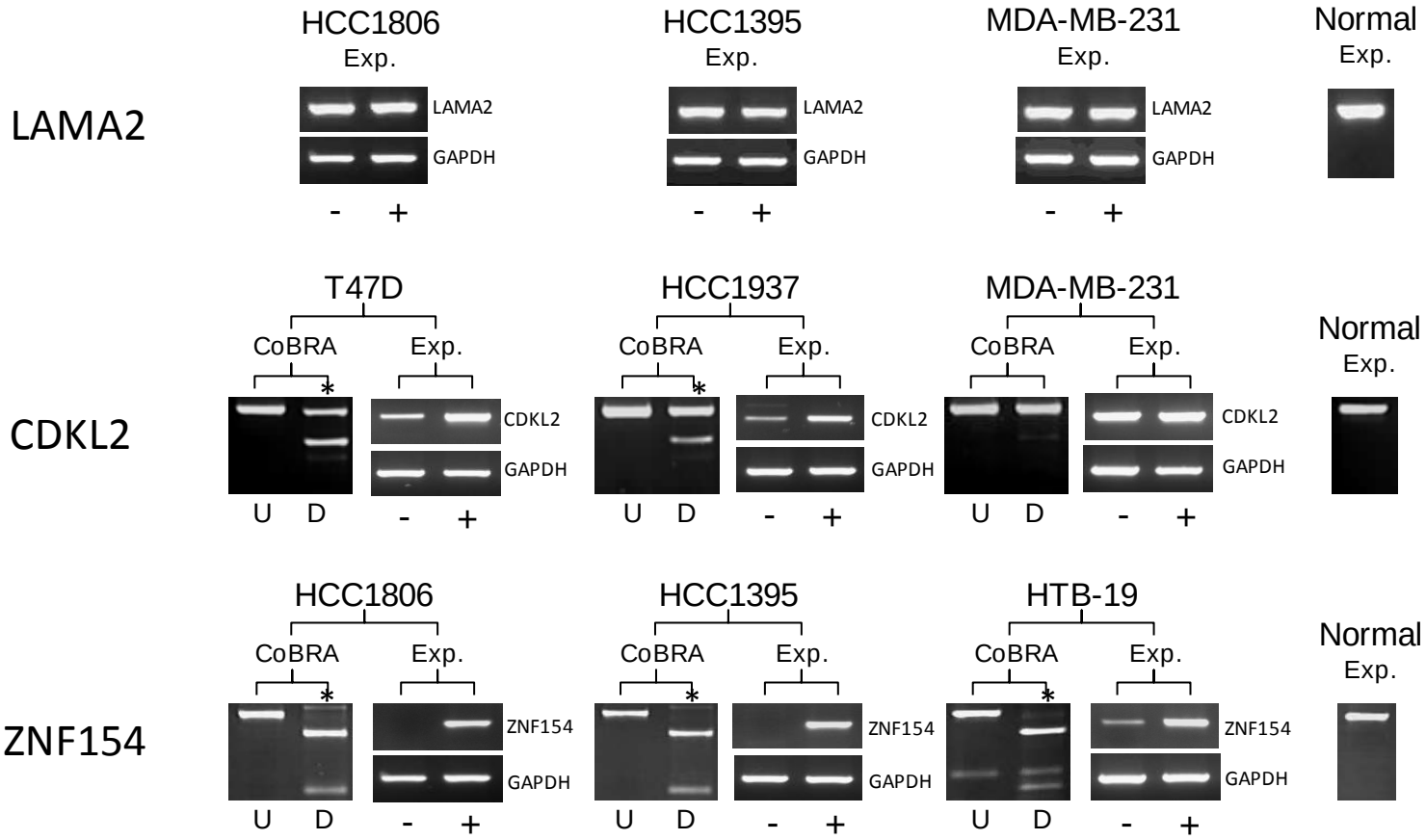


Figure 4.13_Cell line CoBRA and expression analysis for highly methylated genes.
Expression results (Exp.) are shown for *LAMA2*, *CDKL2*, and *ZNF154* and *GAPDH*. Three cell lines are shown in each case, all with (+) and without (-) treatment with 5-azaDC. Expression for normal breast tissue RNA is also shown. Where re-expression was seen (*CDKL2* and *ZNF154*) CoBRA results are also shown for the corresponding cell lines. Undigested products (U) are run next to digested products (D). * indicates methylated samples

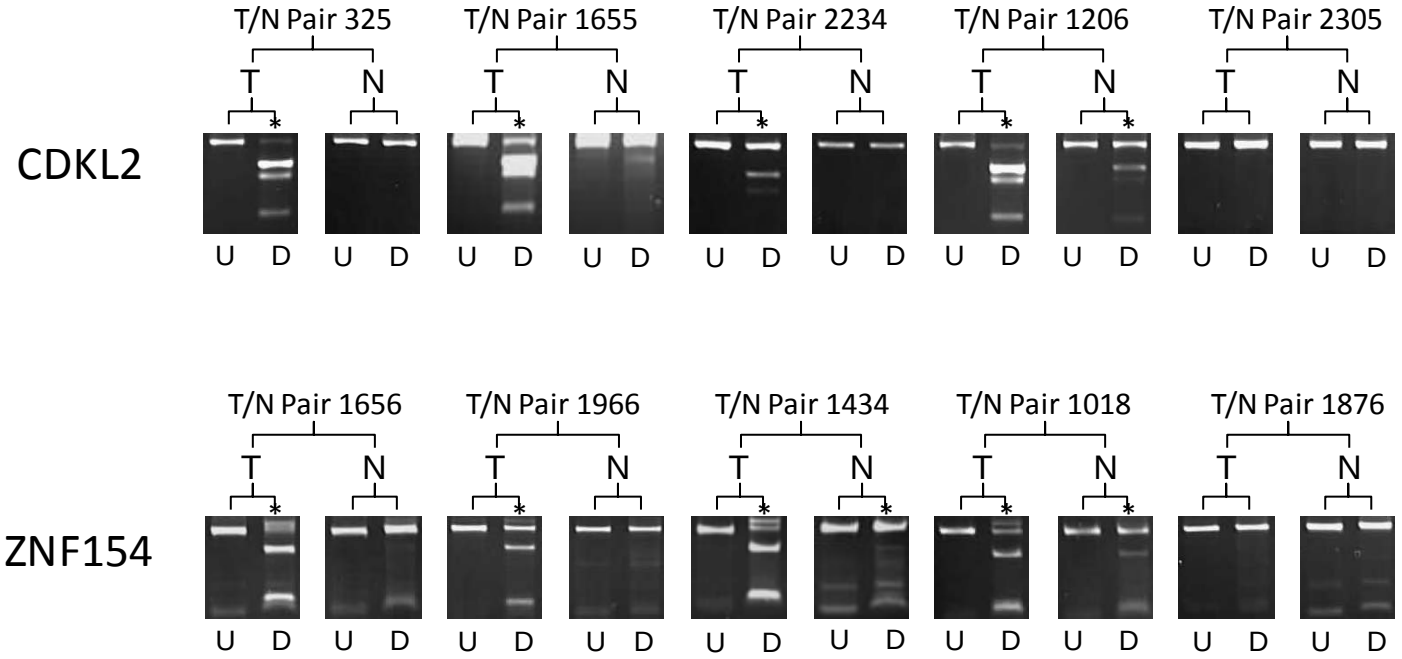


Figure 4.14_CoBRA tumour/normal pair results for *CDKL2* and *ZNF154*.
CoBRA results are shown for five tumour (T) and normal (N) paired samples with undigested product (U) run adjacent to digested product (D). * indicates methylated samples

4.4.5 Analysis of clinical features

Overall tumour hypermethylation levels were assessed using Euclidean hierarchical clustering and analysis of variance.

4.4.5.1 Hierarchical clustering

Hierarchical clustering was performed upon the selected hypermethylated probes to cluster the 39 tumours into groups showing the greatest similarity in methylation profiles. Three distinct clusters were produced with two outliers (figure 4.15). Cluster 1 consisted of 18 highly methylated tumours (mean β value = 0.433), cluster 3 (n=12) showed low overall methylation levels (mean β value = 0.176) and cluster 2 (n=7) showed an intermediary level of methylation (mean β value = 0.317) ($p<0.0001$; ANOVA) (figure 4.16a). Samples in cluster 1 were enriched in samples that relapsed (66.7% of samples compared to 14.3% and 25% for clusters 2 and 3 respectively) and samples that had both oestrogen and progesterone receptors (ER and PR) (77.8% of samples compared to 14.3% and 16.7% for clusters 2 and 3 respectively). Clusters 2 and 3 both contained high levels of samples with no ER and no PR expression (71.4% and 58.3% respectively). The most notable difference in clinical feature between clusters 2 and 3 was the high level of triple negative tumours in cluster 3 (58.3% compared to 28.6% in cluster 2). These results show different levels of methylation changes within the tumours are associated with clinical features. This association between global methylation and clinical feature is worthy of further analysis both on a global and individual gene level.

4.4.5.2 Analysis of Variance (ANOVA) analysis

Clustering results suggested an association between overall methylation levels and certain clinical features such as relapse and receptor status. To assess clinical features and overall

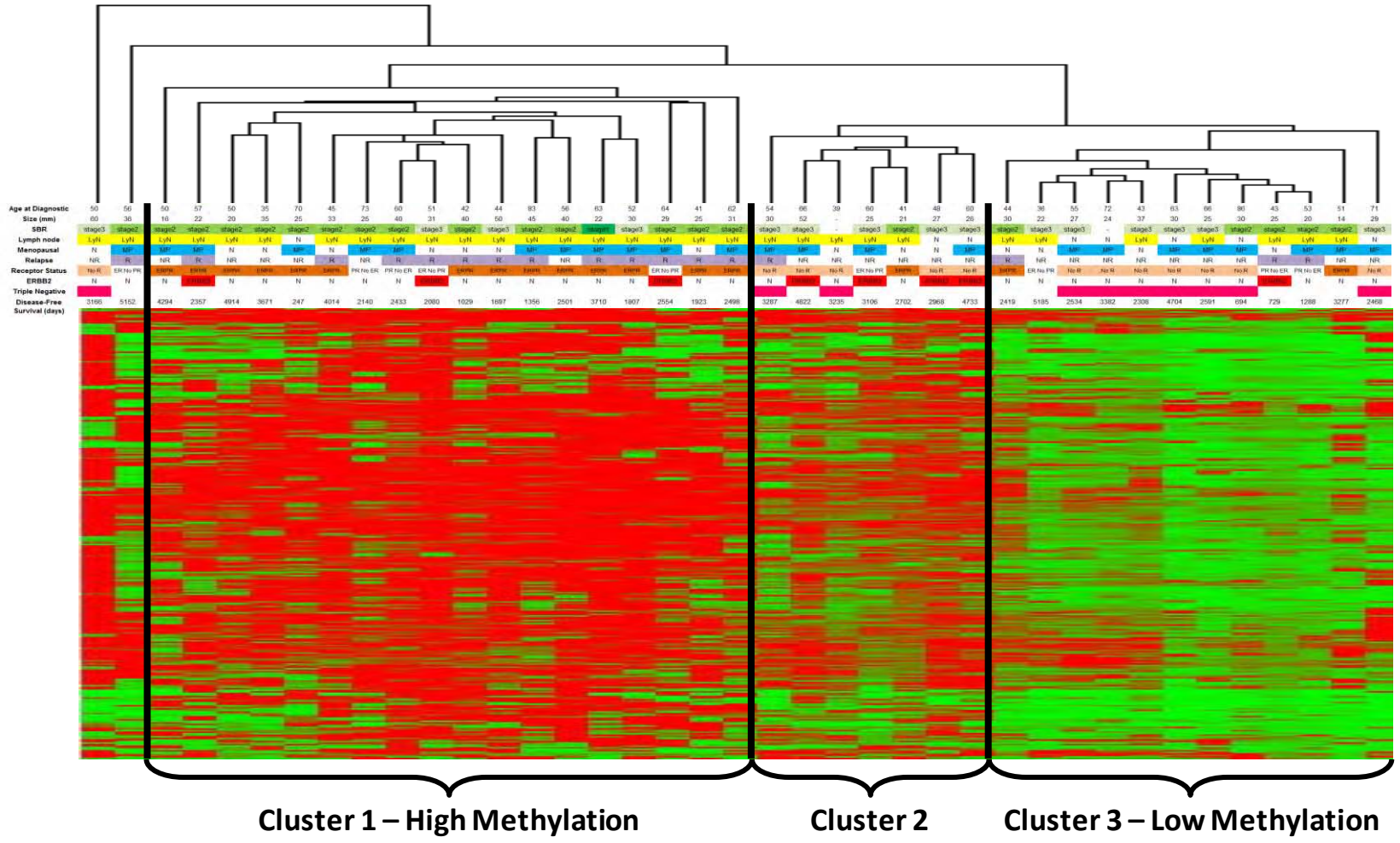

Age at Diagnostic
Size (mm)
SBR
Lymph node
Menopausal
Relapse
Receptor Status
ERBB2
Triple Negative
Disease-Free Survival (days)
Cluster 1 – High Methylation
Cluster 2
Cluster 3 – Low Methylation

Figure 4.15_Hierarchical clustering splits tumour samples into three main groups
The dendogram can be seen aligned with sample information showing clinical status of: age at diagnosis (years), tumour size (mm), SBR stage (1,2 or 3), lymph node status (LyN = lymph node positive; N = lymph node negative), menopause status (MP = menopausal; N = not menopausal), relapse status (R= relapse; NR = no relapse), estrogen and progesterone receptor status (ERPR = estrogen and progesterone positive; ERnoPR = estrogen positive only; PRnoER = progesterone positive only; noR = estrogen and progesterone negative), ERBB2 status (ERBB2 = ERBB2 overexpression, N=normal expression), triple negative (pink highlighting = triple negative) and survival (disease free survival in days). The three cluster groups, showing high, medium and low methylation are separated with thick black lines and labelled underneath. Two outliers can be seen on the extreme left of the clustering image. Red and green refer to β-values >0.5 and <0.5 respectively.

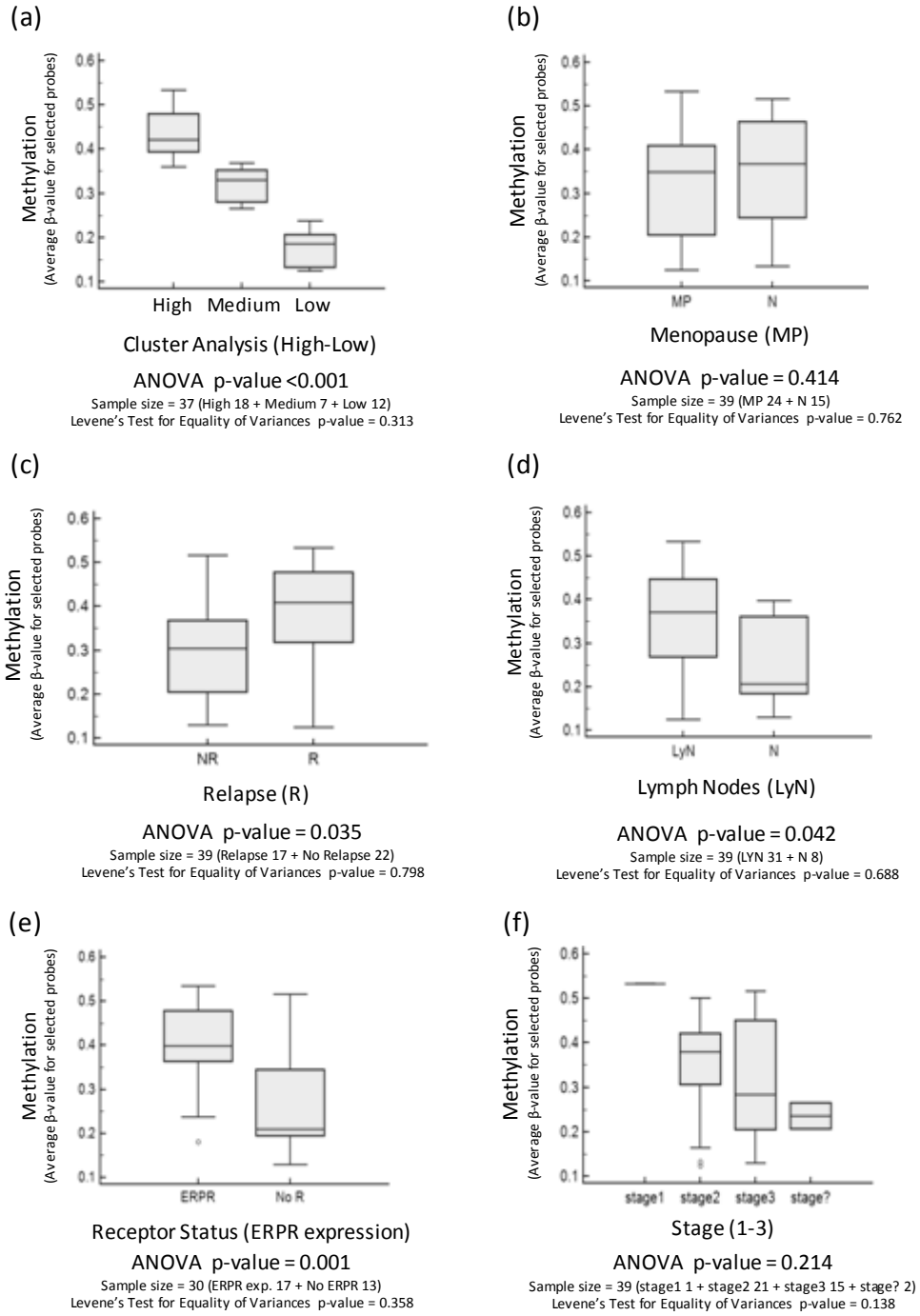


Figure 4.16_ANOVA analysis of methylation and clinical features.
Results are shown for ANOVA analysis of overall methylation levels against cluster groups (a), menopause status (b), relapse status (c), lymph node status (d), ER and PR receptor status (e) and tumour stage (f). Central boxes within the box-plot figures represent the interquartile range, lines within the central box represent the median value and upper and lower horizontal lines represent the highest and lowest values (circles above or below represent any outliers).

methylation levels further, analysis of variance (ANOVA) testing was used. Overall methylation levels show significant differences in relapsing, lymph node positive and ER and PR receptor positive tumours (figures 4.16c, d and e respectively), in each case, presence of the particular clinical feature associated with a greater level of overall hypermethylation. Patient menopausal status and tumour stage did not show any significant association with methylation (figures 4.16b and f). For this cohort of samples, disease free survival (DFS) did not significantly associate with any clinical feature or methylation cluster groups. However, DFS and relapse was just below significance (0.052) (figure 4.17).

4.4.5.3 Identification of individual genes with association to clinical features

In order to identify individual gene loci where methylation status correlated with clinical features, Fisher's exact tests were carried out on all 291 probes. After the use of false discovery rate (FDR) correction, 9 probes (9 genes; *RECK*, *SFRP2*, *ITR*, *UGT3A1*, *SYDE1*, *UGT3A2*, *ACADL*, *UAP1L1* and *PKRCB1*) were identified where methylation status correlated significantly with relapse, 19 probes (18 genes; *B3GAT1*, *TNFRSF10D*, *GRIA4*, *C1orf114*, *DNAJC6*, *C6orf34*, *CORIN*, *LAMA2*, *ITR*, *MCAM*, *C12orf34*, *NPY*, *COL1A2*, *RSNL2*, *ZNF660*, *ADAMTSL1*, *DPP4*, *MAT1A*) where methylation status correlated with ER/PR receptor expression status and 2 probes (2 genes; *LDHD* and *ZNF660*) where methylation status significantly associated with lymph node positive tumours (table 4.2). Genes associating with relapse and ER/PR receptor status were chosen to be analysed further.

Relapse

The methylation status for the 9 relapse associated genes are shown in figure 4.18. Since probes specifically methylated in relapsing tumours should show strong correlations with DFS, Kaplan-Meier analysis was carried out on all nine probes/genes. Using Bonferroni

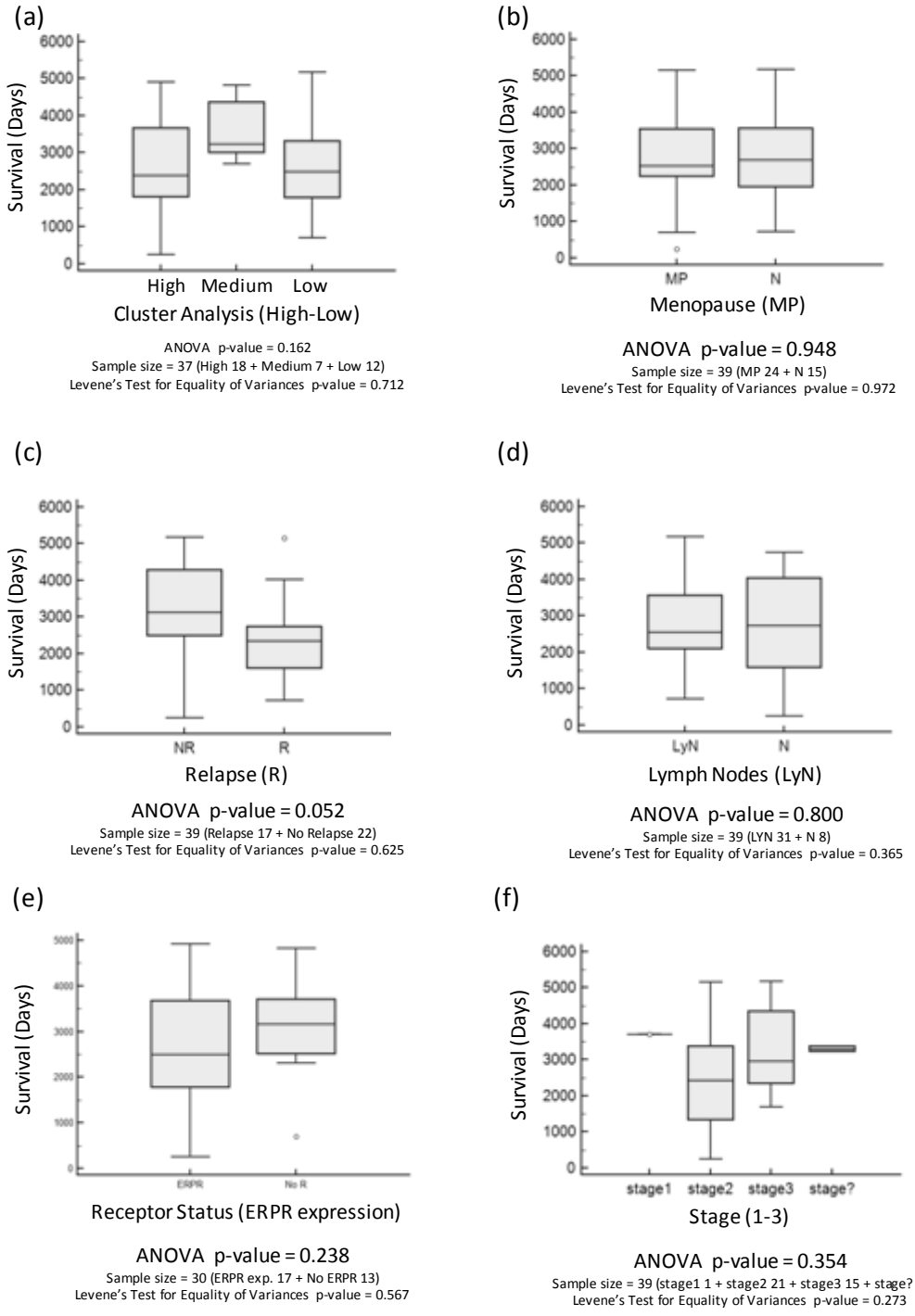


Figure 4.17_ANOVA analysis of disease free survival and clinical features.
Results are shown for ANOVA analysis of disease free survival levels against cluster groups (a), menopause status (b), relapse status (c), lymph node status (d), ER and PR receptor status (e) and tumour stage (f). Central boxes within the box-plot figures represent the interquartile range, lines within the central box represent the median value and upper and lower horizontal lines represent the highest and lowest values (circles above or below represent any outliers).

Clinical feature	Number of probes	Number of genes	Gene names
Relapse	9	9	RECK, SFRP2, ITR, UGT3A1, SYDE1, UGT3A2, ACADL, UAP1L1, PKRCB1
ER and PR	19	18	B3GAT1, TNFRSF10D, GRIA4, C1orf114, DNAJC6, C6orf145, CORIN, LAMA2, ITR, MCAM, C12orf34, NPY, COL1A2, RSNL2, ZNF660, ADAMTSL1, DPP4, MAT1A
Lymph node positive	2	2	LDHD, ZNF660

Table 4.2_Individual gene methylation associating with clinical features.
For relapse, ER and PR positive, triple negative and lymph node positive tumours, the number of probes and genes are shown with gene names.

Gene symbol	ILMNID	RELPASE	NO RELPASE	Number methylated relapse (total = 17)	Number methylated no relapse (total = 22)	FDR corrected P-value
RECK	cg12717594			10	1	0.00026
SFRP2	cg23207990			11	2	0.00086
ITR	cg09582042			11	2	0.00086
UGT3A1	cg23317501			10	2	0.00554
SYDE1	cg04981492			11	3	0.00972
UGT3A2	cg07084163			11	3	0.00972
ACADL	cg09068528			8	1	0.03490
UAP1L1	cg04582938			10	3	0.04377
PRKCB1	cg04582938			10	3	0.04377

Figure 4.18_Probes that show significant association between methylation and relapse.
The figure shows the nine significant probes associating with relapse status ranked in order of FDR corrected p-value. Methylation status is indicated for each of the 39 tumours with either a red circle (methylated) or green circle (unmethylated). The frequency of methylation is shown for relapsing and non-relapsing tumours along with FDR-corrected p-values.

correction, 6 genes (*ACADL*, *RECK*, *SFRP2*, *UAP1L1*, *UGT3A1* and *ITR*) were identified that showed significant association between methylation status and DFS (figure 4.19).

For the three most significant genes, *ACADL*, *RECK*, and *SFRP2*, it was determined whether the methylation detected at the single CpG dinucleotide in the Infinium array was indicative of a wider level of methylation and whether this could be capable of affecting expression, indicating possible functional importance in addition to clinical relevance. Cell line re-expression after treatment with 5-azaDC and corresponding methylation in breast cancer cell lines was observed for all three genes (figure 4.20). CoBRA results of tumour normal paired samples showed frequent methylation of *RECK* (22.2%; 4/18), *ACADL* (30%; 6/20) and *SFRP2* (57.1%; 8/14). All methylation of *RECK* was tumour specific whilst some methylation was observed at a lower level in corresponding normal tissue in a subset of methylated samples for *ACADL* (33.3%; 2/6) and *SFRP2* (50%; 4/8). Clone sequencing of *ACADL* and *RECK* confirmed the CoBRA results, showing frequent, wide-spread, tumour-specific methylation (figures 4.21 and 4.22). In addition, quantitative real-time PCR data obtained by Dr. Sophie Vacher in Ivan Bieche's laboratory (Oncogenetic Laboratory, INSERM U 735, Centre René Huguenin, Saint Cloud, France) showed significant downregulation of methylated tumour samples compared to a panel of normal controls for both *RECK* ($p<0.0001$) and *ACADL* ($p = 0\ 043$) (figure 4.23a and b), but not *SFRP2*, confirming a possible functional consequence of *RECK* and *ACADL* methylation.

Oestrogen and progesterone receptor status

Of the 19 probes associated with ER and PR receptor status, three of them also had multiple probes within the final list of 291, *TNFRSF10D, C1orf114* and *COL1A2* (figures 4.24 and 4.12) and so were chosen for further analysis.

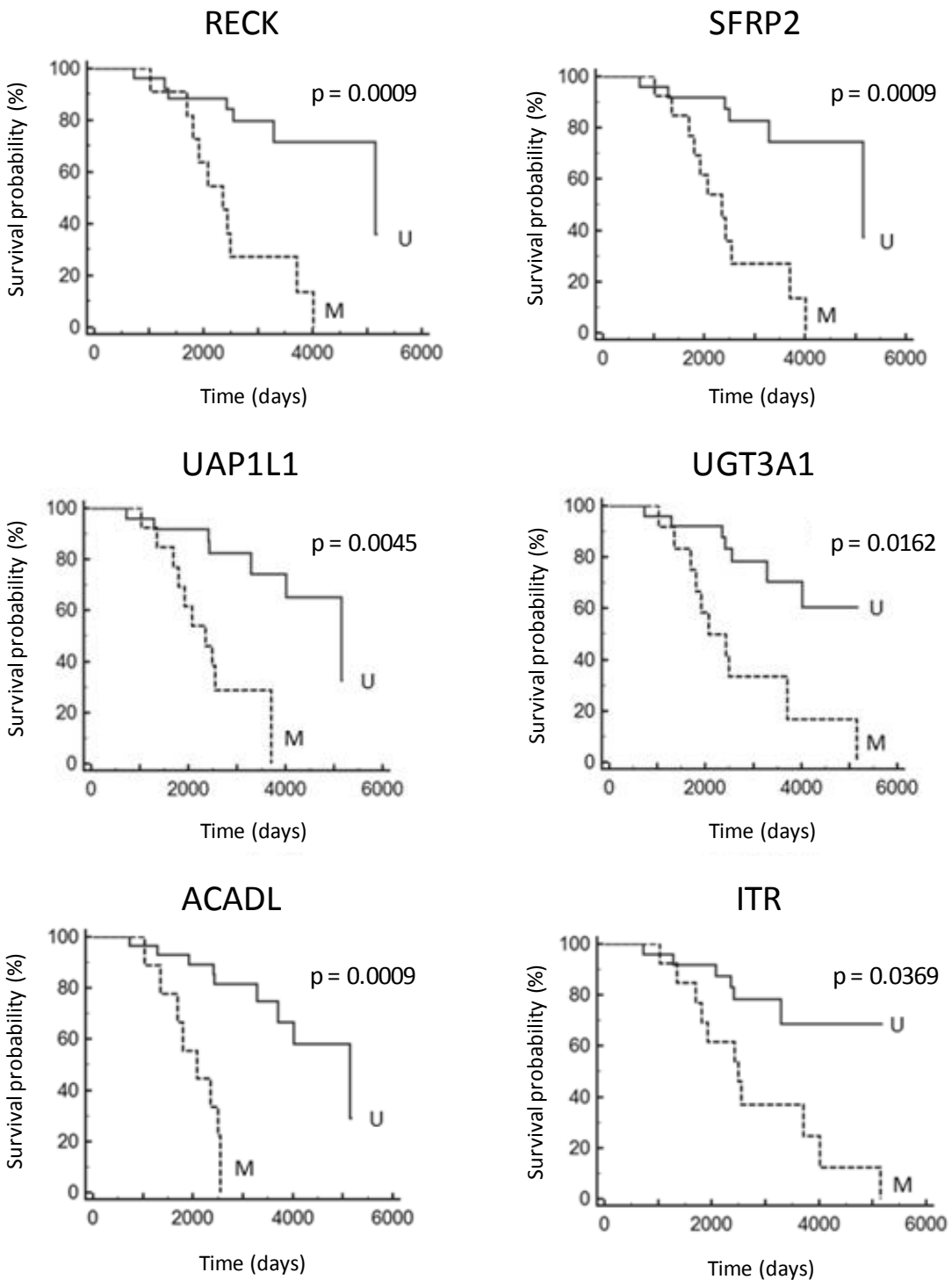


Figure 4.19_Kaplan-Meier analysis.
Kaplan-Meier survival graphs are shown for the six genes demonstrating a significant association between methylation and disease free survival (DFS). This analysis was performed on results from 39 tumour samples. The minimum and median survival times were 257 days and 2554 days respectively. For non-relapsing tumours, the maximum period of follow-up was 5185 days.

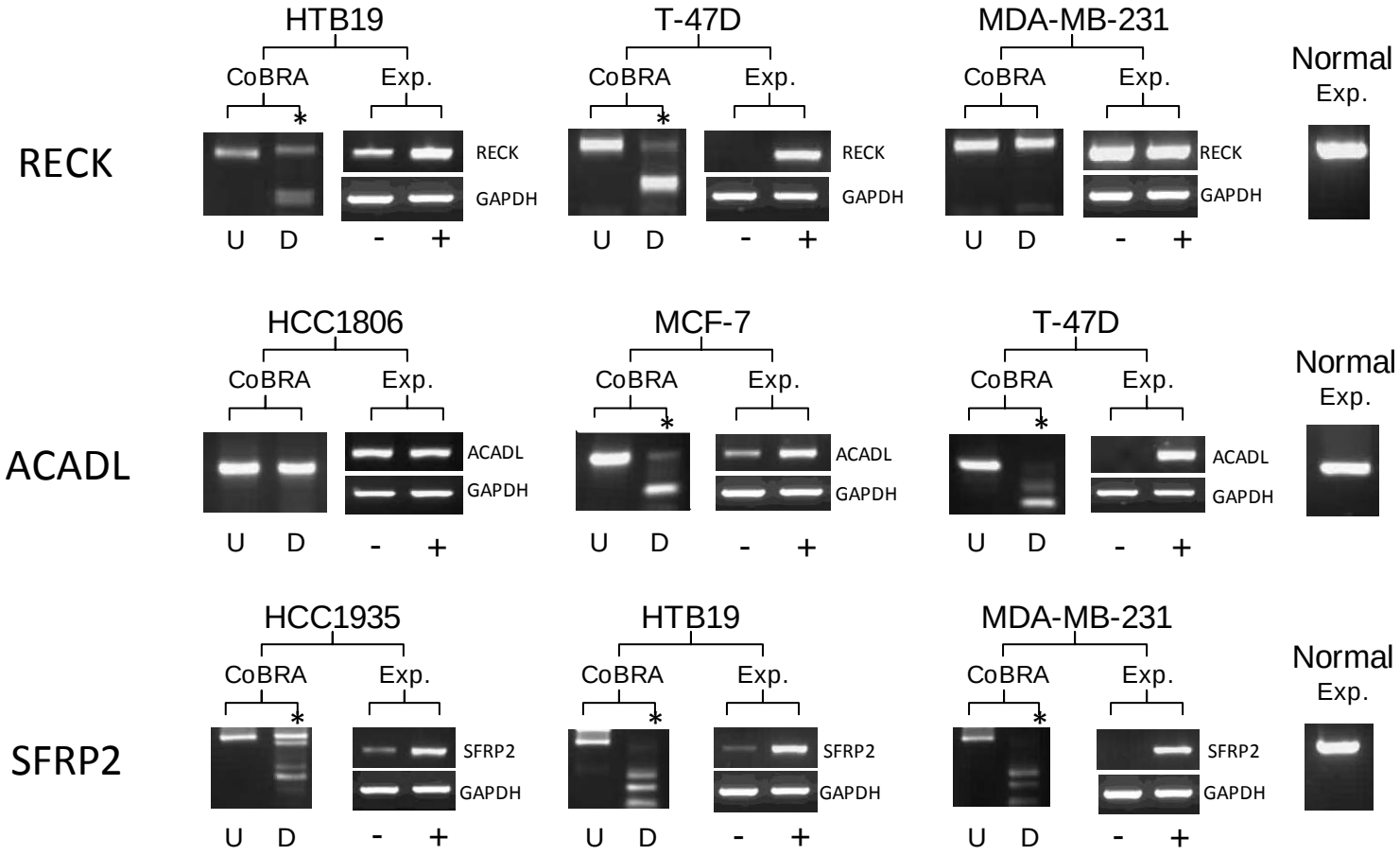


Figure 4.20_Cell line CoBRA and expression analysis of *RECK, ACADL* and *SFRP2*.
Expression results (Exp.) are shown for *RECK*, *ACADL*, *SFRP2* and *GAPDH*. Three cell lines are shown in each case, all pre (-) and post (+) treatment with 5-azaDC. Expression for normal breast is also shown. CoBRA results are also shown for the corresponding cell line. Undigested products (U) are run next to digested products (D). * indicates methylation samples

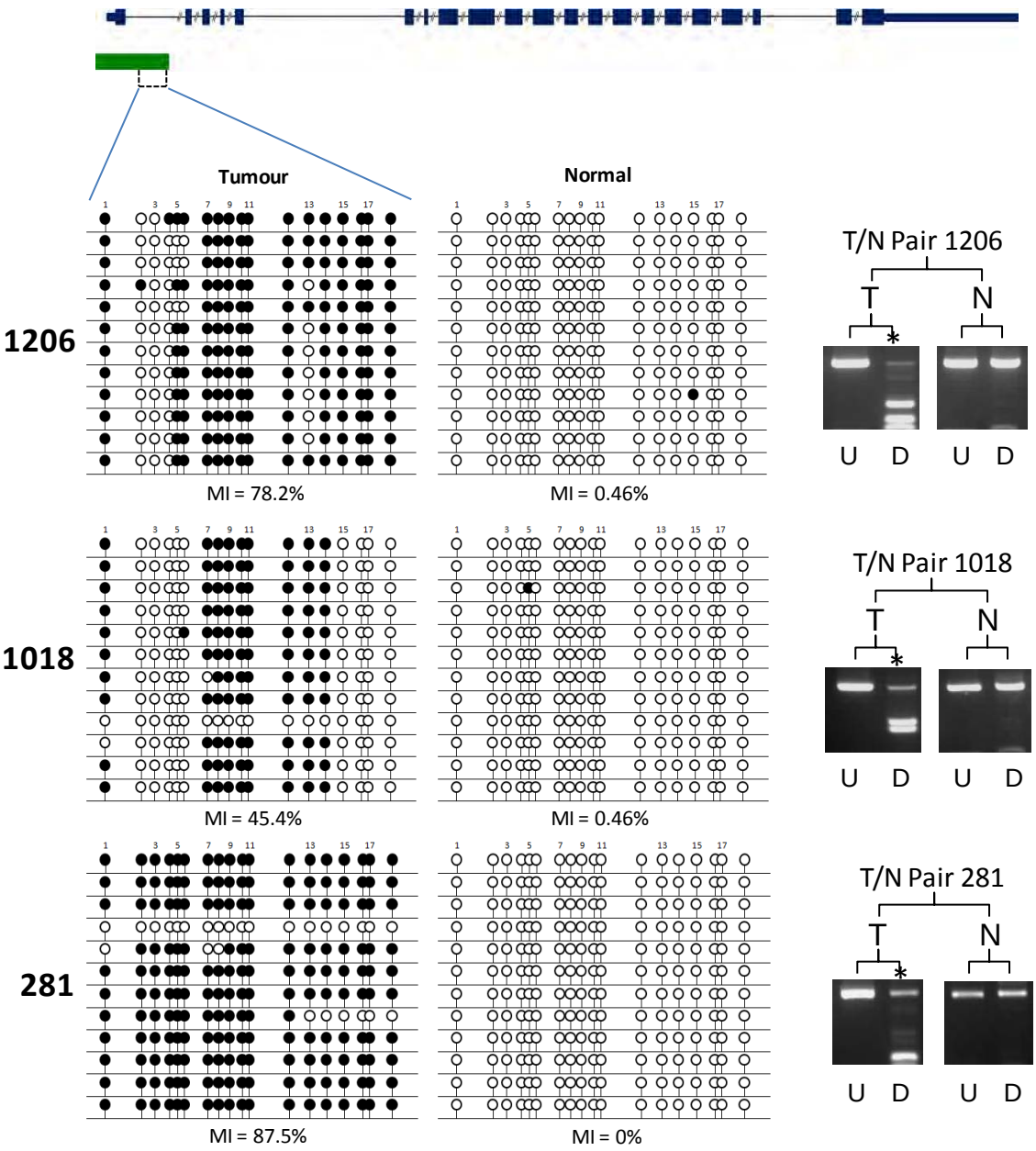


Figure 4.21_Clone sequencing and CoBRA results for *RECK* tumour normal pairs.
A gene schematic for *RECK* is shown illustrating the location the of the CpG island (green box) and the region being analysed (black dotted lines). Exons and introns are represented by blues boxes and black lines respectively. Clone sequencing and corresponding CoBRA results are shown for three pairs. For each clone sequencing result black lines represent single alleles and black and white circles represent methylated and unmethylated CpGs respectively. MIs are shown for each sample as a percentage of methylated CpGs out of the total number of CpGs analysed. CoBRA results are shown with undigested product (U) run next to digested product (D). * indicates methylation

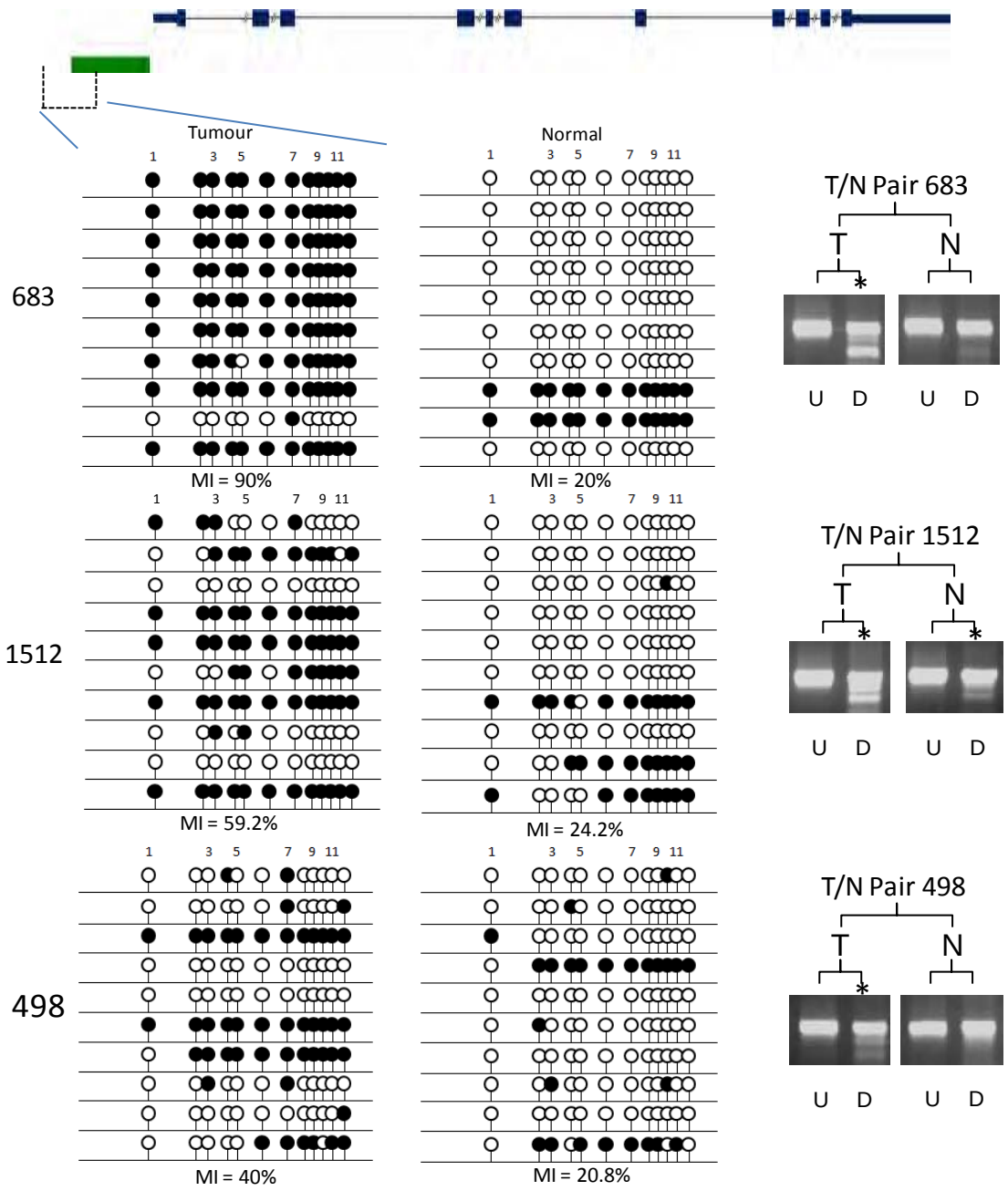


Figure 4.22_Clone sequencing and CoBRA results for *ACADL* tumour normal pairs.
A gene schematic for *ACADL* is shown illustrating the location the of the CpG island (green box) and the region being analysed (black dotted lines). Exons and introns are represented by blues boxes and black lines respectively. Clone sequencing and corresponding CoBRA results are shown for three pairs. For each clone sequencing result black lines represent single alleles and black and white circles represent methylated and unmethylated CpGs respectively. MIs are shown for each sample as a percentage of methylated CpGs out of the total number of CpGs analysed. CoBRA results are shown with undigested product (U) run next to digested product (D). * indicates methylated samples

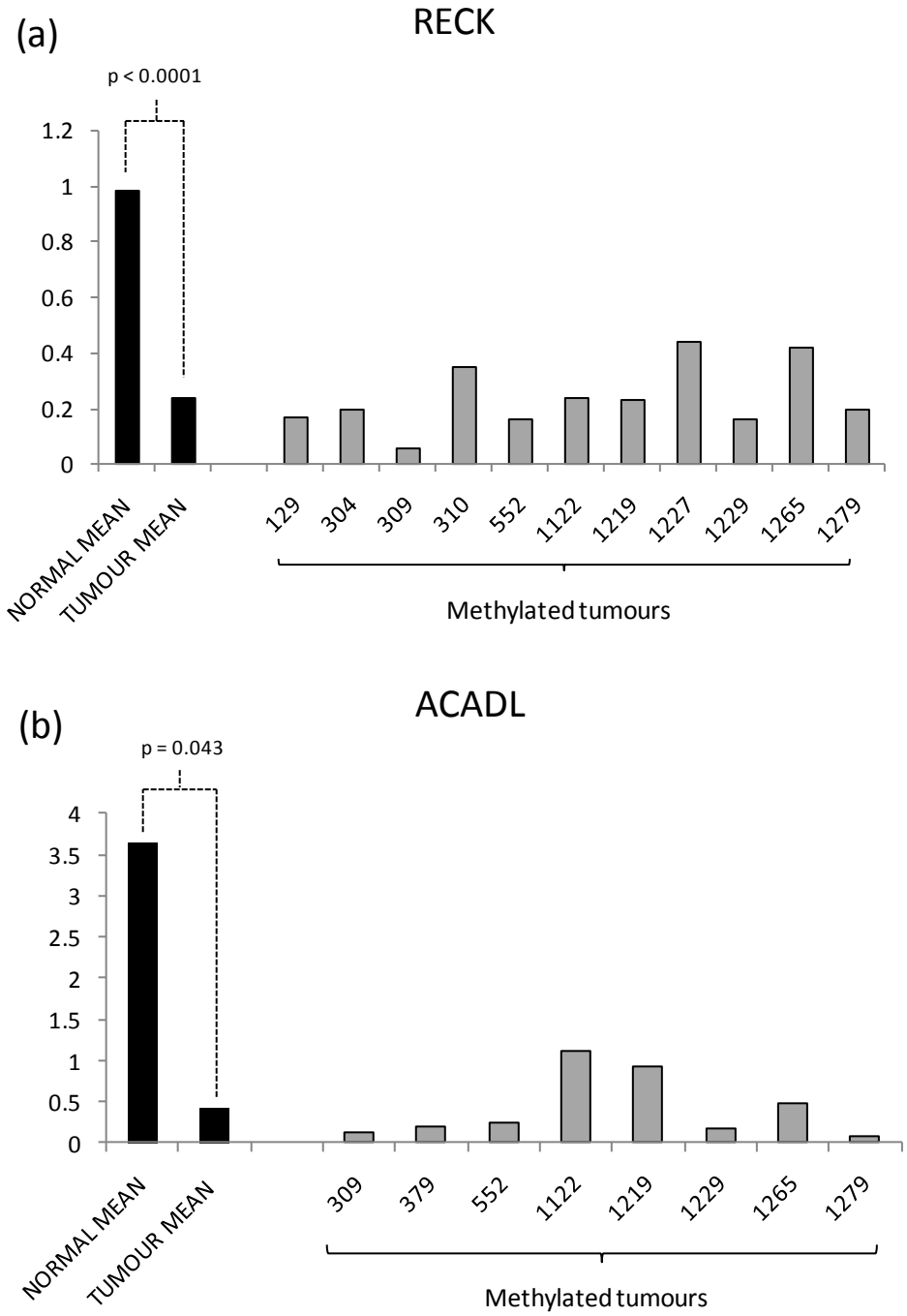


Figure 4.23_Quantitative real-time expression results for *RECK* and *ACADL*. Graphs are shown for real-time results for both *RECK* (a) and *ACADL* (b). Mean results are shown for normal and tumours in both cases alongside results for all available methylated tumours. P-values are given for the significance of downregulation in methylated tumour samples compared to normal samples (Student's t-test). Data for these graphs were obtained by Dr. Sophie Vacher from Dr. Ivan Bieche's laboratory (Oncogenetic Laboratory, INSERM U 735, Centre René Huguenin, Saint Cloud, France).

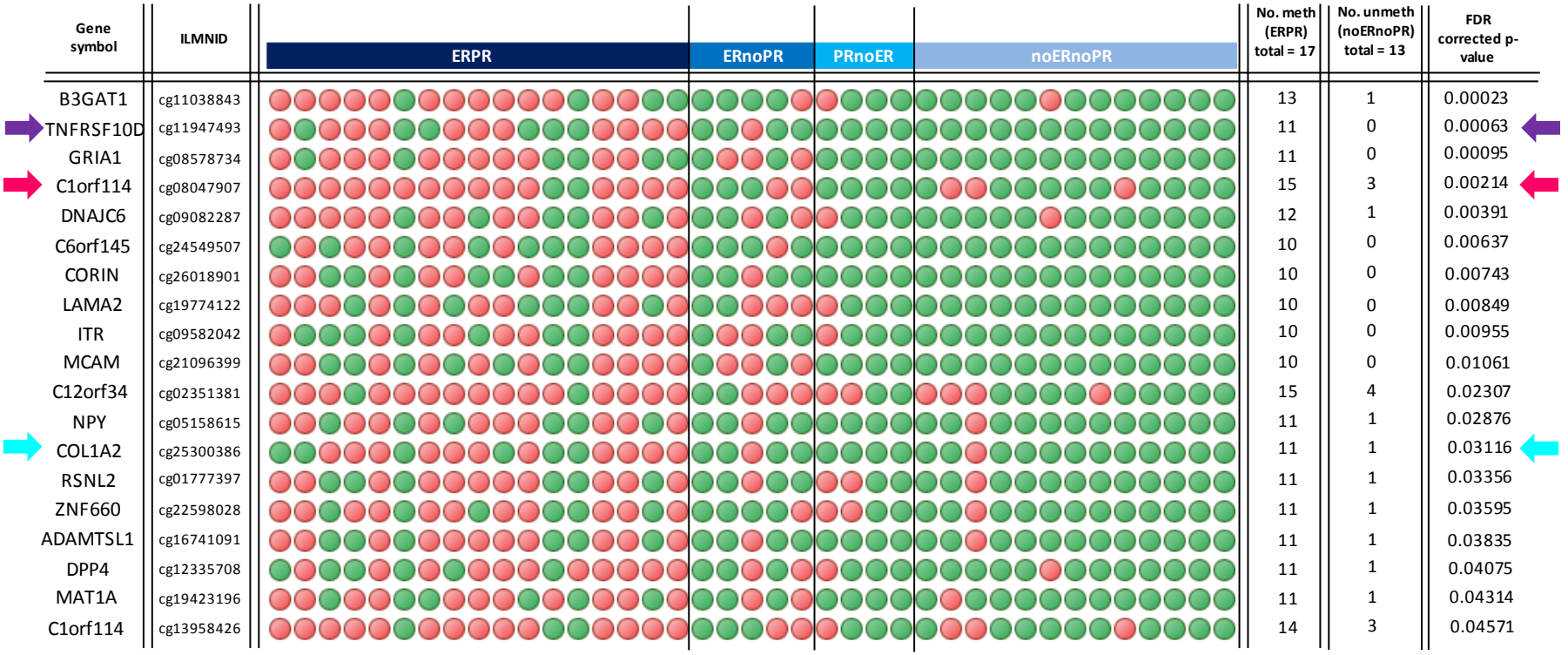


Figure 4.24_Probes showing a significant association between methylation and ER and PR expressing tumours
The 19 probes whose methylation status significantly associates with ER and PR positive tumours are shown, ranked in order of p-values. Methylated and unmethylated samples are represented by red and green circles respectively. Methylation frequency for each probe in ERPR positive tumours and ERPR negative tumours is also shown alongside FDR corrected p-values. Purple, pink and blue arrows are aligned to genes that have been analysed further.

All three genes demonstrated both increased expression in 5-azaDC treated cell lines and methylation in corresponding lines (figure 4.25). CoBRA analysis demonstrated frequent methylation in tumour normal paired samples for *COL1A2*, *TNFRSF10D*, and *C1orf114*. *COL1A2* showed tumour specific methylation in 76.5% (13/17) of samples, the remainder being unmethylated. *TNFRSF10D* showed tumour specific methylation in 50% (8/16) of samples, increased methylation in the tumour compared with normal tissue in 18.8% (3/16) of samples, equal amounts of methylation in the tumour and normal tissue of 12.5% (2/16) of samples, whilst the remaining 3 samples (18.8%) were unmethylated. *C1orf114* showed tumour specific methylation in 35.5% (6/17) of samples and increased methylation in tumour compared to corresponding normal tissue in 42.2% (7/17) of samples, the remaining 4 samples (23.5%) were unmethylated (figure 4.26). Bisulphite sequencing of *C1orf114* demonstrated tumour specific methylation throughout the selected area of the CpG island (figure 4.27). Quantitative real-time RT-PCR by Sophie Vacher in Dr. Ivan Bieche's laboratory (Oncogenetic Laboratory, INSERM U 735, Centre René Huguenin, Saint Cloud, France) demonstrated significant downregulation of *TNFRSF10D* (p=0.016) in methylated primary sporadic tumours compared to a panel of normal breast controls (figure 4.28).

4.4.6 Analysis of MIRA genes

The cohort of 39 sporadic breast tumours used in the Infinium array were the same samples used for confirmation of genes analysed in the MIRA array (chapter 3) In total, sixteen genes (*EMIIN2*, *DBC1*, *FBLN2*, *COMP*, *SESN3*, *EPSTI1*, *CD44*, *TP53INP1*, *NRXN2*, *SIM2*, *PAX9*, *POU4F1*, *FOXF2*, *CIDE-A*, *GSC*, *SALL1*) were assessed by CoBRA in chapter 3 and were also present on the Infinium array. Results for both assays are shown in table 4.3. Of the 16 MIRA genes, 10 (*EMILIN2*, *DBC1*, *FBLN2*, *COMP*, *SESN3*, *EPSTI1*, *SIM2*, *PAX9*, *CIDE-A* and *SALL1)* were methylated in >20% primary tumours, as assessed by CoBRA. The

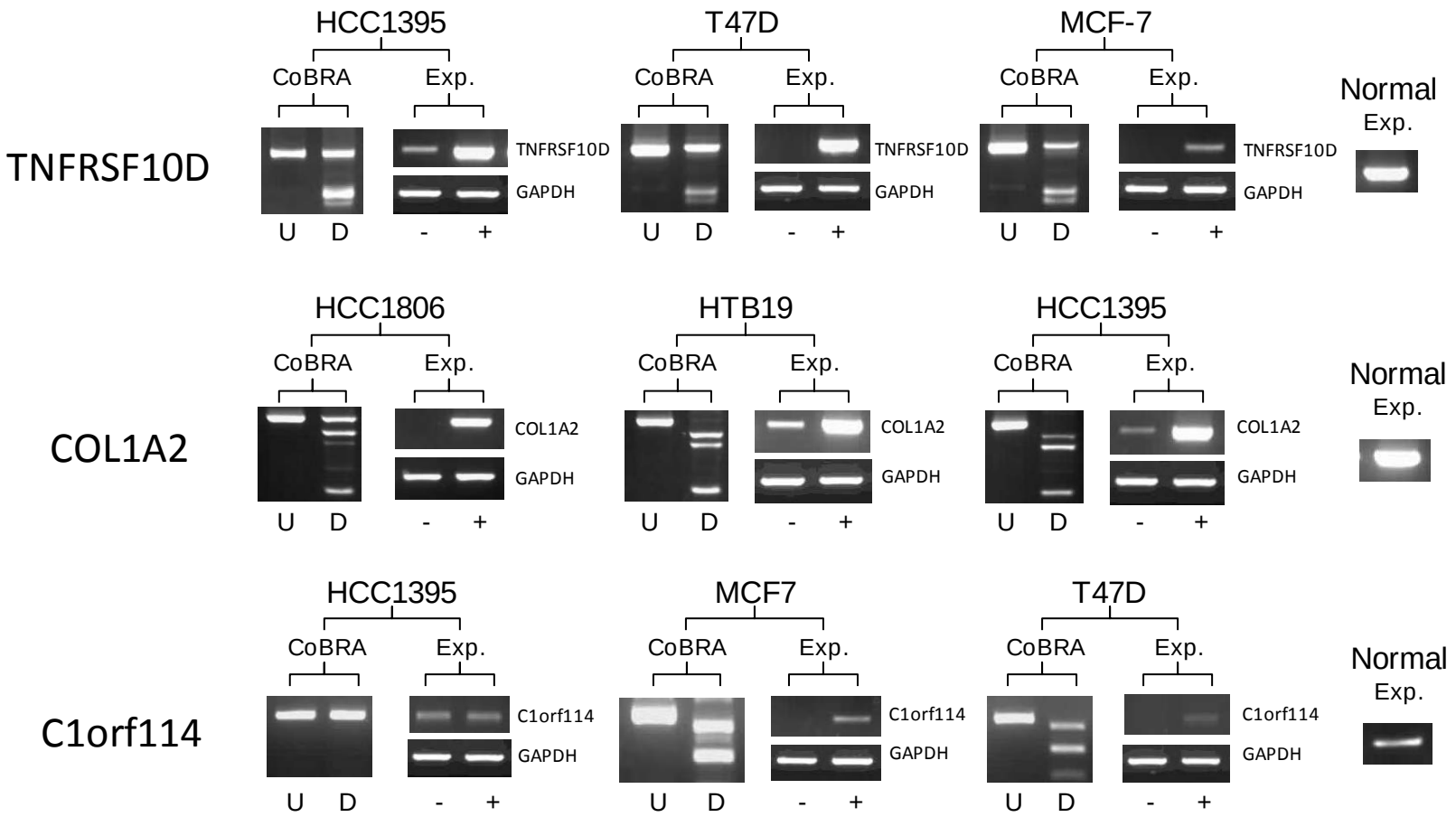


Figure 4.25_Cell line CoBRA and expression analysis for ER and PR associated genes.
Expression results (Exp.) are shown for *TNFRSF10D*, *COL1A2*, *C1orf114* and *GAPDH*. Three cell lines are shown in each case, all with (+) and without (-) treatment with 5-azaDC. Expression for normal breast tissue RNA is also shown. CoBRA results are also shown for the corresponding cell lines. Undigested products (U) are run next to digested products (D). * indicates methylated samples

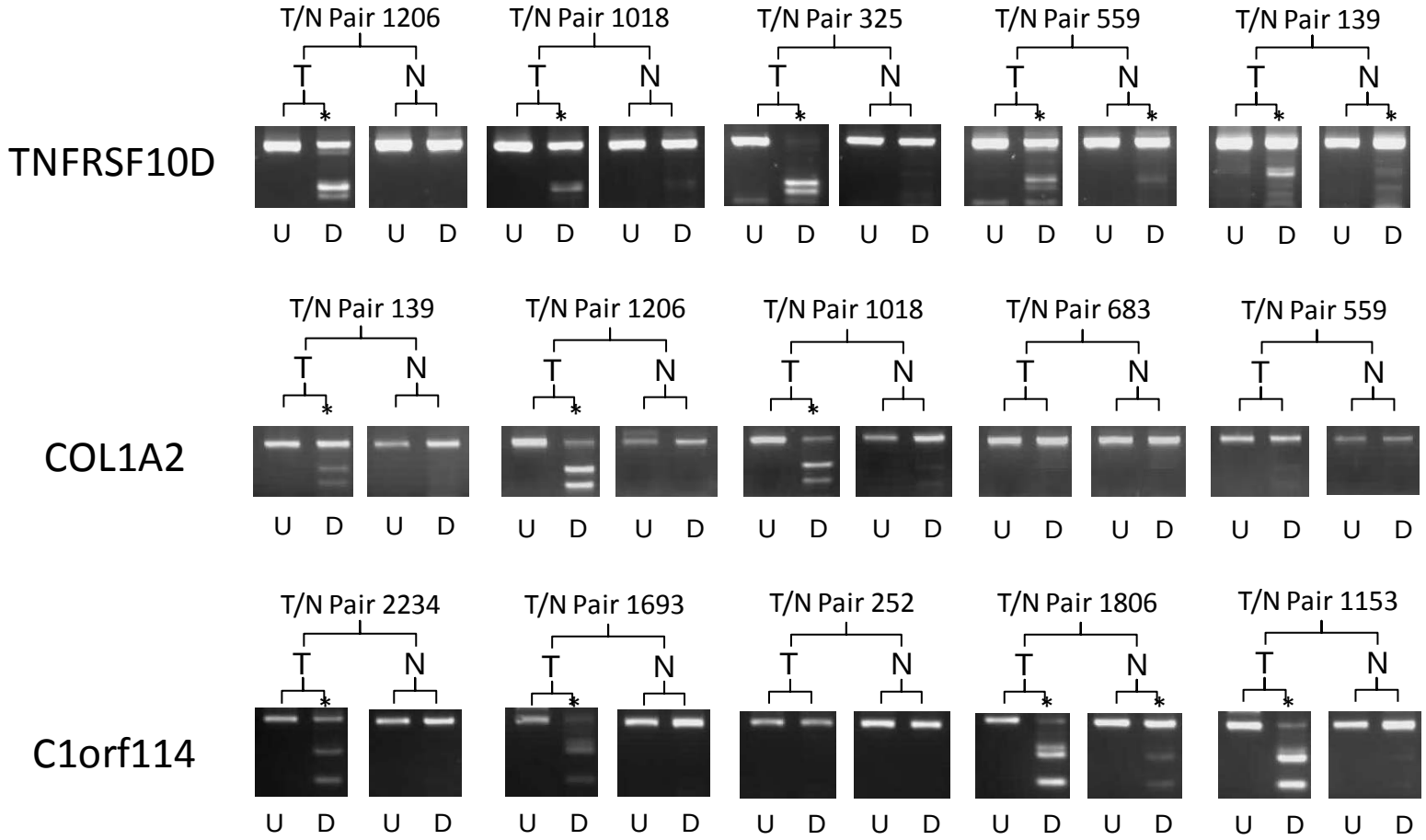


Figure 4.26_CoBRA tumour/normal pair results for *TNFRSF10D, COL1A2* and *C1orf114*.
CoBRA results are shown for five tumour (T) and normal (N) paired samples with undigested product (U) run adjacent to digested product (D). * indicates methylated samples

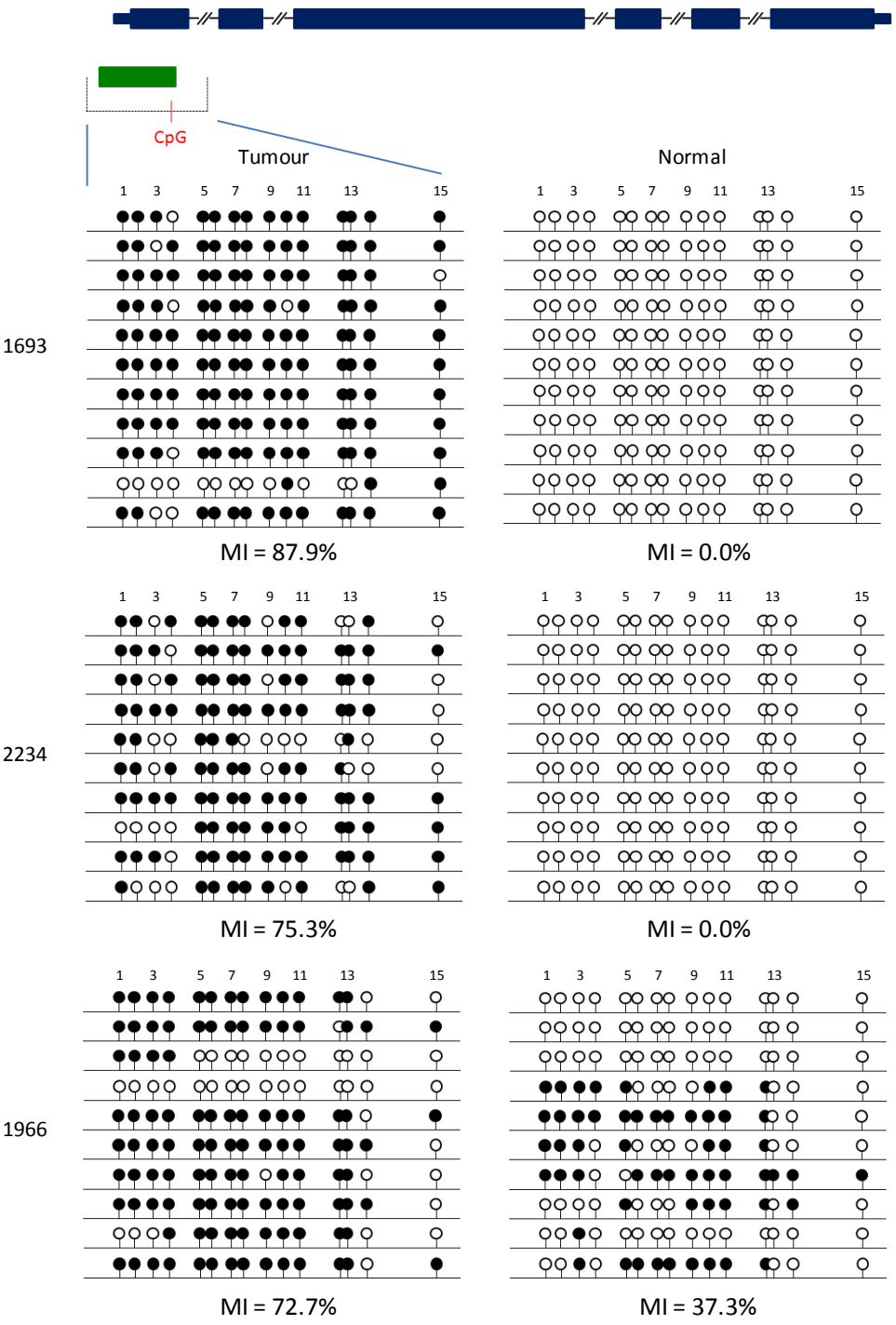


Figure 4.27_Clone sequencing and CoBRA results for *C1orf114* tumour normal pairs.
A gene schematic for *C1orf114* is shown illustrating the location the of the CpG island (green box) and the region being analysed (black dotted lines). Exons and introns are represented by blues boxes and black lines respectively. For each clone sequencing result black lines represent single alleles and black and white circles represent methylated and unmethylated CpGs respectively. MIs are shown for each sample as a percentage of methylated CpGs out of the total number of CpGs analysed.

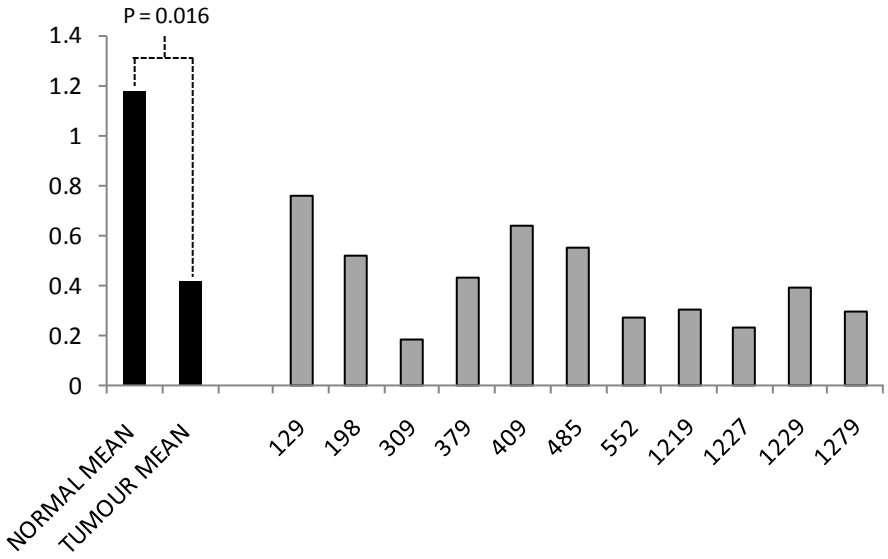


Figure 4.28_Qunatitative real-time expression results for *TNFRSF10D* .
A graphs is shown for real-time results for *TNFRSF10D*. Mean results are shown for normal and tumours alongside results for all available methylated tumours. A P-value is given for the significance of downregulation in methylated tumour samples compared to normal samples (Student's t-test). Data for these graphs were obatined by Dr. Sophie Vacher in Dr. Ivan Bieche's laboratory (Oncogenetic Laboratory, INSERM U 735, Centre René Huguenin, Saint Cloud, France).

Gene	CoBRA methylation frequency	Infinium methylation frequency	Infinium probe	Present in Infinium list	Reason for not being in list	Overlapping region being analysed
EMILIN2	16/36 (44.4%)	10/39 (25.6%)	cg09009111	no	>0.25 in at least 1 normal	no
		1/39 (2.6%)	cg14133708	no	>0.25 in at least 1 normal and tumour frequency <20%	yes
DBC1	8/31 (25.8%)	15/39 (38.5%)	cg25935911	no	>0.25 in at least 1 normal	no
FBLN2	13/38 (33.3%)	6/39 (15.4%)	cg00201234	no	>0.25 in at least 1 normal and tumour frequency <20%	no
		13/39 (33.3%)	cg16604516	no	>0.25 in at least 1 normal	yes
COMP	20/33 (60.6%)	1/39 (2.6%)	cg15784332	no	>0.25 in at least 1 normal and tumour frequency <20%	no
		21/39 (53.8%)	cg09949775	no	>0.25 in at least 1 normal	no
SESN3	7/10 (70%)	2/39 (5.1%)	cg11665588	no	tumour frequency <20%	no
		0/39 (0.0%)	cg13037275	no	tumour frequency <20%	no
EPSTI1	13/36 (36.1%)	2/39 (5.1%)	cg14539231	no	tumour frequency <20%	no
		0/39 (0.0%)	cg22905097	no	tumour frequency <20%	no
CD44	1/24 (4.2%)	0/39 (0.0%)	cg17640322	no	tumour frequency <20%	yes
		0/39 (0.0%)	cg18652941	no	tumour frequency <20%	no
		0/39 (0.0%)	cg01879488	no	tumour frequency <20%	yes
		0/39 (0.0%)	cg04125208	no	tumour frequency <20%	yes
		0/39 (0.0%)	cg08530414	no	tumour frequency <20%	no
		0/39 (0.0%)	cg08606356	no	tumour frequency <20%	yes
		0/39 (0.0%)	cg20143092	no	tumour frequency <20%	no
TP531NP1	0/18 (0.0%)	2/39 (5.1%)	cg18059933	no	tumour frequency <20%	no
		0/39 (0.0%)	cg24434118	no	tumour frequency <20%	yes
NRXN2	0/12 (0.0%)	0/39 (0.0%)	cg16718678	no	tumour frequency <20%	no
SIM2	24/29 (82.7%)	4/39 (10.3%)	cg02672220	no	tumour frequency <20%	no
		0/39 (0.0%)	cg13694867	no	tumour frequency <20%	no
PAX9	8/39 (20.5%)	27/39 (69.2%)	cg26620157	no	>0.25 in at least 1 normal	?
		27/39 (69.3%)	cg00509670	no	>0.25 in at least 1 normal	?
POU4F1	4/27 (14.8%)	13/39 (33.3%)	cg08097882	yes	-	yes
		0/39 (0.0%)	cg15604467	no	tumour frequency <20%	yes
FOXF2	3/35 (8.6%)	0/39 (0.0%)	cg03848675	no	tumour frequency <20%	no
		4/39 (10.3%)	cg08045570	no	>0.25 in at least 1 normal and tumour frequency <20%	no
CIDE-A	21/40 (52.5%)	2/39 (5.1%)	cg19883905	no	tumour frequency <20%	no
		4/39 (10.3%)	cg20950011	no	>0.25 in at least 1 normal and tumour frequency <20%	no
GSC	1/40 (2.5%)	0/39 (0.0%)	cg19224837	no	tumour frequency <20%	yes
		0/39 (0.0%)	cg20804555	no	tumour frequency <20%	no
SALL1	25/40 (62.5%)	0/39 (0.0%)	cg22674717	no	tumour frequency <20%	no

Table 4.3_Infinium results for genes identified through MIRA.
The sixteen genes analysed by CoBRA in chapter 3 and the infinium array are shown with methylation frequencies for both assays. The single probe identified in both assays is highlighted with a red arrow. The reasons for other probes not being identified in the infinium list are shown.

remaining 6 genes (*CD44, TP53INP1, NRXN2, FOXF2 GSC*, and *POU4F1*) were either unmethylated or methylated at a very low frequency by CoBRA assay. These results are also true for the Infinium assay, showing general agreement between the two methods for negligible methylation (table 4.3). The only slight discrepancy is *POU4F1*, there are two probes are present on the Infinium array for *POU4F1*, one of which shows both a higher level of methylation (33.3%) than the other (0%) and of the CoBRA result from chapter 3 (14.8%). Of the 10 positive genes, only 4 (*EMILIN2, DBC1, FBLN2, COMP*), showed similar methylation frequencies by CoBRA and Infinium assays (table 4.3). The remaining 6 (*SESN3, EPSTI1, SIM2, PAX9, CIDE-A* and *SALL1*) showed larger discrepancies between CoBRA and Infinium results. In most cases methylation was detected at a higher level by CoBRA, for example, 82.7% samples were methylated by CoBRA compared to 10.2% by Infinium for *SIM2*. The 16 genes analysed were represented by 34 probes, only one of which was picked up through the criteria used in the Infinium array and therefore present in the list of 291, *POU4F1* (Infinium probe number: cg08097882). The remaining probes where CoBRA results were in general agreement were not within the list of 291 because of either (a) the frequency being <20%, (b) at least one normal with a β-value>0.25, or (c) a combination of the two (table 4.3). These results do not show a large amount of concordance between the two methods, however, it should be noted that in the majority of cases, the individual CpG being assessed by Infinium does not sit within the region assessed by CoBRA (table 4.3). Considering the more sensitive nature of the Inifinium assay, it is surprising that in general higher methylation frequencies were observed by CoBRA, however this is most likely due to the region being analysed being much larger. Despite this, it is reassuring that the lowly methylated samples are not picked up at high frequencies in either assay.

Analysis of both MIRA and Infinium candidate selections (irrelevant of any further laboratory analyses) identifies 34 genes which are present in both of the initial selection lists (table 4.4).

Gene symbol	Gene product
C1orf114	chromosome 1 open reading frame 114
C6orf206	chromosome 6 open reading frame 206
CCK	cholecystokinin
CDKL2	cyclin-dependent kinase-like 2 (CDC2-related kinase)
CLDN6	claudin 6
FZD10	frizzled homolog 10 (drosophila)
GHSR	growth hormone secretagogue receptor
HAND1	heart and neural crest derivatives expressed 1
HIF3A	hypoxia inducible factor 3, alpha subunit
HIST1H3J	histone cluster 1, H3j
HOXB13	homeobox B13
KLF11	Kruppel-like factor 11
LYPD5	LY6/PLAUR domain containing 5
NEUROD1	neurogenic differentiation 1
PAQR9	progestin and adipoQ receptor family member IX
PCDHAG12	protocadherin gamma subfamily A, 12
POU4F1	POU class 4 homeobox 1
RAB37	RAB37, member RAS oncogene family
RBP1	retinol binding protein 1, cellular
RUNX3	runt-related transcription factor 3
SFRP2	secreted frizzled-related protein 2
SIX6	SIX homeobox 6
SLC2A14	solute carrier family 2 (facilitated glucose transporter), member 14
TBX20	T-box 20
TF	transferrin
VIPR2	vasoactive intestinal peptide receptor 2
VSX1	visual system homeobox 1
ZNF132	zinc finger protein 132
ZNF177	zinc finger protein 177
ZNF560	zinc finger protein 560

Table 4.4_Genes identified through both MIRA and Infnium analyses.
The table shows the gene symbol and product name of the 34 genes identified in MIRA and Infinium lists irrelevant of any further laboratory analyses.

4.5 DISCUSSION

Genome wide methylation studies have been successful for sometime in identifying methylated genes in cancer that may have tumour suppressor properties. Due to technology improvements it is now possible to use a smaller amount of starting material and cheaper to do many samples. This has made it feasible to assess a larger number of primary tumours, giving rise to many large scale methylation profiling studies (Christensen *et al*, 2011; Noushmehr *et al*, 2010; O'Riain *et al*, 2009; Martín-Subero *et al*, 2009a). On a smaller scale, this work has used the Illumina Infinium array to produce a profile of methylation at 14,475 CpG loci in 39 sporadic ductal breast tumours and four tumour/normal paired samples.

Results have shown that overall methylation profiles unsurprisingly differ between tumours and normals and that methylation changes are widely heterogeneous, in agreement with other similar studies (Van der Auwera *et* al, 2010; Christensen *et al*, 2010). This suggests that some tumours are less susceptible to changes in methylation, possibly as a result of powerful genetic events meaning tumour formation can occur and progress without a large prevalence of alternative epigenetic events. It may also be indicative of a lack of general deregulation of methylation in some tumours. A previous breast cancer Infinium study compared levels of DNMT3b, DNMT1 and DNMT3a between cluster groups with high and low mean β-values and found significantly increased DNMT3b expression and higher levels of DNMT1 approaching significance in the high β-value clusters. Although, these associations were not so strong when taking into account the expression of a cell proliferation marker and therefore the different rates of growth (Van der Auwera *et al*, 2010).

A list of 291 probes representing 164 genes considered to be frequently hypermethylated in tumours was produced containing numerous genes already known to be methylated in breast

cancer. In addition, genes within the list were associated with functions or pathways involved in tumourigenicity indicating the list had produced a relatively probable and relevant set of methylated genes in breast cancer. Although this study has not confirmed wider methylation frequencies for all hypermethylated genes, the validation work and further analysis of candidate genes and genes associated with clinical features suggests that the β-value from one CpG will be generally indicative of the wider level of methylation surrounding it. Although this may be considered to be slightly biased, due to the selection for sequencing analysis being based upon genes previously shown to be hypermethylated in other cancers. Nonetheless, wider methylation was also been observed for novel genes that had not been previously analysed, for example, *C1orf114* was found to be both extensively methylated and highly specific to tumours.

Clustering and ANOVA analyses identified increased aberrant methylation associated with lymph node, hormone receptor (ER/PR) and relapse positive tumours, with a marked lack of hypermethylation in triple-negative tumours. This lack of altered methylation in triple negative (-ER, -PR, -ERBB2 amplification) tumours agrees with previous published work in basal-like breast cancers which are also triple negative (Holm *et al*¸ 2010)

4.5.1 Oestrogen (ER) and progesterone (PR) positive tumours

Recent breast cancer methylation profiling studies using either Illumina Infinium or GoldenGate platforms have shown a similar trend for ER and PR positive tumours (Li *et al*, 2010 and Holm *et al*, 2010), suggesting that ER/PR positive tumours may be more susceptible to hypermethylation events. The presence of these hormone receptors are very import to breast cancer cells as they allow the cancer cells to react to outside hormones and bring about cellular changes through their transcription factor activity. Both receptors work as

transcription factors that are activated by their ligand binding. It is generally thought that at least some of their transcription targets are useful in tumourigenesis and this is evidenced by the negative response ER+ tumours can have to drugs that block oestrogen's effects, such as tamoxifen (Jordan, 1993). It would make a reasonable amount of sense that not all the activated genes would be advantageous to tumourigenesis and these could well be counteracted by *de novo* methylation.

This study identified 19 CpG loci (18 genes) where positive hypermethylation status correlated with ER and PR expression. Further analysis of three of these genes (*TNFRSF10D, COL1A2* and *C1orf114*) showed extensive methylation levels around the CpG probe loci and that methylation of these genes can affect their expression, suggesting this may also be the case for the 15 remaining genes.

4.5.1.1 TNFRSF10D

TNFRSF10D (*tumour necrosis factor receptor superfamily, member 10d, decoy with truncating death domain*) is an eight exon gene located at 8p21 with a 532bp CpG island containing 53 CpGs. It is a member of the TRAIL (tumour necrosis factor-related apoptosis-inducing ligand) family, a group of 5 receptors involved in apoptosis initiation. Binding of TRAIL to either of the death receptors *TNFRSF10a* (also known as DR4) or *TNFRSF10b* (also known as DR5) induces apoptosis via caspase-8/-10 however, binding of TRAIL to any of the three decoy receptors, such as *TNFRSF10D* does not induce apoptosis, in fact overexpression of the decoy receptors enhances protection against apoptosis (reviewed in Mahalingam *et al*, 2009), suggesting that downregulation of *TNFRSF10D* would not be advantageous to the tumour. Despite this, *TNFRSF10D* has previously been shown to be methylated in melanomas and colorectal cancers (Liu *et al*, 2008; Prabhu *et al*, 2009). After

carrying out this work, it was also noted that it had previously been shown to be hypermethylated in breast tumours and malignant mesothellomas (Shivapurkar *et al*, 2004), but reported under an uncommonly used synonym, DrC2. A GoldenGate study published after this work was done also showed that lower methylation was associated with basal-like tumours (predominantly triple-negative) compared to *ERBB2* overexpressing tumours (Bediaga *et al*, 2010). Interestingly, this study did not find any association between lower methylation levels and triple negative tumours, but rather it was found that greater *TNFRSF10D* methylation levels were observed in ER/PR expressing tumours. This does not prove or disprove either of these associations, as they may just be differences due to the use of different cohorts of breast cancer tumours and further investigation in further cohorts will be needed to provide a more definitive answer.

4.5.1.2 COL1A2

COL1A2 (*collagen, type 1, alpha 2*) is a large gene situated at 7q22.1 consisting of 52 exons that produce a 129 kDa protein. Type I collagen alpha 2 chain is an integral part of the type I collagen triple helix. Mutations within *COL1A2* can result in various forms of both Ehlers-Danlos syndrome and Osteogeneis Imperfecta. Patients with these diseases do not have a predisposition to cancer however *COL1A2* hypermethylation has been observed in many cancer types including medulloblastoma, bladder cancer and melanomas (Anderton *et al*, 2008, Mori *et al*, 2009 and Koga *et al*, 2009). In agreement with our results, the three studies mentioned all showed lower expression in methylated cell lines compared with 5-azaDC treated lines. In addition, both Anderton *et al* (2008) and Mori *et al* (2009) showed decreased expression in primary tumour samples by quantitative real-time PCR. In medulloblastomas, reduced expression was also demonstrated at the protein level in cell lines (Anderton *et al*, 2008). Mori *et al* were also able to demonstrate tumour suppressor effects of *COL1A2* on cell

growth and proliferation when introduced back into a methylated, non-expressing line. These results suggest *COL1A2* hypermethylation is a frequent event in cancers and can result in loss of expression and subsequent loss of tumour suppressing properties. The results from this study have demonstrated a possible role of *COL1A2* involvement in breast cancer.

Following this work, an Illumina GoldenGate study on 162 breast tumours showed differential methylation of *COL1A2* associated with lymph node positive tumours and tumour size (Christensen *et al*, 2010). The work done within this thesis demonstrated neither a significant difference nor a trend towards this association with lymph node status. There are many factors that could contribute to this, for example, the sample size used in this thesis is much smaller than that used by Christensen *et al* and the identification of individual genes was carried out using a different approach. The cohort used in this thesis also had a very large number of lymph node positive tumours, making significant associations less likely. In this thesis, probes were identified as being hypermethylated by stringent criteria and then analysed for clinical association based on a methylated or unmethylated classification. Christensen *et al* compared average β-values using a linear regression model able to take into account the mean distribution of β-values. When data for this study is analysed on raw β-values using student's t-test, one of the *COL1A2* probes associated with lymph node positive tumours at just below significance (FDR corrected p-value=0.065).

4.5.1.3 C1orf114

C1orf114 (*chromosome 1 open reading frame 114*) is a 6 exon gene located at 1q24. It has a CpG island of 248bp containing 20 CpGs situated across the transcription start site. Although thought to produce a 509 amino acid protein, no information about any possible function is known. However, results from this study suggest hypermethylation is frequent and specific in

breast tumours. In addition, 5-azaDC studies suggest methylation is capable of altering expression of this gene and that mRNA is present in normal breast tissue. *C1orf114* was also detected in the high stringency gene list from the MIRA assay on breast cancer cell lines (chapter 3) and in a MIRA assay on DCIS samples (Tomassi *et al*, 2009), although in both cases no further analysis was carried out. Taken together, these results suggest that it may be worth looking into the possible role of *C1orf114* in breast cancer, particularly as the methylation in DCIS suggests it may be an early event in breast tumourigenesis. Due to the high frequency and specificity of methylation and novel value it may also be worth screening other cancers for promoter hypermethylation. Additionally, functional investigation of the protein product would be useful and aid in assessing the importance of loss of expression.

Since the remaining 15 genes were not analysed any further they will not be discussed in detail, except to say that none of them have previously been shown to be hypermethylated in breast cancer nor have they been identified in any other Infinium/GoldenGate study at the time of writing.

4.5.2 Relapse and DFS

Although this is a small study, nearly all relapsing tumours were within the highly methylated group, suggesting an overall increase in hypermethylation events are more common in relapsing tumours. However, as factors were only considered on a univariate level, it would be necessary to do this on both a multivariate level and on different, larger cohorts to be certain of this association.

4.5.2.1 SFRP2

SFRP2 (*secreted frizzled-related protein 2*) is a three exon gene that generates a 295 amino acid protein. It has a CpG island situated across the transcription start point that is 1315bp

long with 112 CpGs. *SFRP2* is a member of five Wnt signalling inhibitors and as such functions to sequester signalling of this pathway, known to be over-active in many cancers (Prosperi *et al*, 2010). Promoter hypermethylation of *SFRP2* has been demonstrated in many cancers, including breast (Veeck *et al*, 2008b; Nojima *et al*, 2007; Urakami *et al*, 2006; Müller *et al*, 2004). The authors describe highly specific frequent methylation in primary breast tumours (83% of samples) and high levels of loss of *SFRP2* expression in tumour tissue. No associations with clinical features were observed with either methylation or loss of expression. The results obtained in this thesis identified a lower overall frequency of methylation (34%) and found methylation to be associated with relapse status and DFS. The discrepancy between the two studies is most likely due to the assay being used, although the Inifnium array does seem to be representative of the wider region it is still only the methylation status of the actual CpG being assayed that has shown to have the association with relapse/DFS, therefore, if the wider region were to be analysed, the larger amount of methylation may no longer associate with relapse/DFS. However, this finding needs to be confirmed on a larger cohort of samples using the same assay to determine if there is a true association between methylation status and DFS in ductal breast tumours. Either way, if methylation at this specific probe does associate strongly with relapse, even if the total CpG island methylation does not, it could be used as a potential biomarker.

4.5.2.2 RECK

RECK (*reversion-inducing-cysteine-rich protein with kazal motifs*) is a 21 exon gene situated on chromosome 9p. *RECK* was initially identified through a cDNA screen for reversion-inducing properties (the ability to reverse the transition of transformed malignant cells from flat to round morphology back to a flat morphology) (Takahashi *et al*, 1998). The 971 amino acid protein was found to inhibit the activity of matrix-metalloproteinase-9 (MMP-9)

(Takahashi *et* al, 1998) and subsequently MMP-2 and MT1-MMP (Oh *et al*, 2001). MMPs are a family of enzymes capable of breaking down extra-cellular matrix (ECM) proteins and as such have a role in tissue remodelling during developmental and cancer progression processes. Left unregulated, MMPs are capable of destroying the ECM, leaving cells susceptible to migration. They have been implicated in breast cancer cell invasion, epithelial to mesenchymal transition and motility (Radisky *et al*, 2010).

Many cancers have demonstrated reduced levels of *RECK* and in most cases this associated with bad prognosis (Clark *et al*, 2007). *RECK* has a 766bp CpG island consisting of 73 CpGs and hypermethylation has been described in lung, colon, oral and gastric cancers as a mechanism of down regulation (Chang *et al*, 2007; Cho *et al*, 2007; Long *et al*, 2008; Du *et al*, 2010) and in most cases is associated with bad prognosis. I believe this to be the first example of *RECK* methylation in breast tumours although downregualtion has been shown in both breast cancer cell lines (Figueira *et al*, 2009) and primary tumours (Span *et al*, 2003). In the study by Span *et al*, tumours with low *RECK* expression levels also had significantly lower recurrence free survival and an association with lymph node positive tumours on a multivariate level. Results from this thesis showed methylated tumours had significantly lower expression levels than normal breast tissue and that methylation (measured by the Infinium assay) associates with relapse and DFS. This is promising as measurement of quantitative methylation at this position is relatively easy and could be set up diagnostically.

RECK expression has been shown to be reduced by oncogenic Ras activity upregulating DNMT3b via the ERK signalling pathway and subsequently inducing downregulation of *RECK* by methylation (Chang *et al*, 2006). The relationship between *RECK* methylation and tumourigenesis may therefore be related to Ras activity. In lung cancer, *RECK* hypermethylation has been associated with k-Ras mutations (Chang *et al*, 2007) although not

all samples showing *RECK* methylation also harboured k-Ras mutations suggesting this is not the only possible mechanism responsible for *RECK* hypermethylation in cancer systems. None of the other *RECK* hypermethylation studies have looked at combined k-Ras mutation and *RECK* methylation status therefore it is uncertain how frequent this scenario is. However, it is an interesting example of how upregulation of an oncogene can cause downregulation of a tumour suppressor gene by hypermethylation. Interestingly, it therefore describes a scenario that would make certain tumours susceptible to specific hypermethylation events.

4.5.2.3 ACADL

ACADL (*acyl-coenzyme A dehydrogensase, long chain*) is an 11 exon gene situated at 2q34-q35. It produces a 403 amino acid protein that forms a homotetramer responsible for the first step in breaking down long-chain fatty acids. Three other members are within the family, *ACADS, ACADM* and *ACADVL,* responsible for breaking down short-chain, medium-chain and very long-chain fatty acids respectively. Mutations in *ACADL* cause LCAD deficiency, a syndrome characterised by problems during periods of fasting. Changes in metabolism are common hallmarks of cancer, however, normally this involves a greater level of fatty acid metabolism to create more Acetyl-coA for entering the Krebs cycle, in which case, downregulation of *ACADL* would be advantageous to the cancer cell. Combined with this, *ACADM* mutations have been detected in breast cancer (Sjöblom *et al*, 2006) and *ACADL* has been shown to be reduced in a retinoic-acid-receptor-α dominant negative transgenic mouse model susceptible to fatty acid deposition and liver tumours (Yanagitani *et al*, 2004).

Since the remaining six genes that showed an association with relapse were not analysed further they will be not discussed in detail. However, since they showed a high association between relapse and/or DFS it would be worth analysing them further in the future as

potentially they would be useful prognostic markers. As for *SFRP2*, *RECK* and *ACADL*, the association with methylation at the particular CpG with relapse/DFS needs to be repeated on a separate, larger cohort of samples using the same assay before these results can be considered as true potential markers of relapse/DFS. However, if this were confirmed it would be a relatively easy diagnostic test with highly relevant prognostic value.

4.5.3 Use of Infinium array for candidate gene identification

As a way of identifying candidate hypermethylated genes three of the most highly methylated genes and three multiple positive probe genes were analysed further.

4.5.3.1 Highly Methylated Genes

Of the three highly methylated genes analysed, *ABCA3*, *COX7A1* and *SST*, only the latter two of the three showed re-expression following treatment with 5-azaDC and methylation in corresponding cell lines.

COX7A1

COX7A1 (*cytochrome c oxidase subunit VIIa polypeptide 1 (muscle)*) is a four exon gene situated at 19q13.1 with a 1258bp CpG island containing 85 CpGs. The protein product is 79 amino acids long and functions as a member of the heteromeric complex of cytochrome c oxidase (COX).

Results identified methylation and re-expression in all cell lines, with all lines showing complete methylation and complete re-expression, with no expression observed in any cell line prior to 5-azaDC treatment. Expression of mRNA was observed in the normal non-cancerous breast sample. Whilst the high level of complete methylation and the implication that *COX7A1* is expressed only in muscle tissue point to *COX7A1* methylation being a

mechanism aiding lack of expression in non-muscle tissue, the presence of mRNA within the normal breast sample suggests this may not be the case. Analysis of tumour/normal paired samples detected methylation in 100% of tumour samples. Methylation frequencies of this magnitude are not commonly observed, especially with a relatively unsensitive assay such as CoBRA, again suggesting methylation may be used as a mechanism for suppression of *COX7A1* expression in non-muscle tissue. However, methylation was detected at either an equal (29.4%; 5/19) or lower (70.6%; 12/17) level in corresponding normal tissue. These results are puzzling as they suggest tumour-enhanced methylation and *COX7A1* mRNA expression in breast tissue when previous evidence suggests expression is only present in muscle tissue. Having said this, *COX7A1* remains largely unstudied and perhaps further work on this gene may highlight functions within tissues other than muscle. It should be noted that the majority of corresponding normal samples do show methylation but at a lower level than the corresponding tumour. If assuming *COX7A1* is normally expressed and lowly methylated in normal breast tissue these results suggest hypermethylation may be an early event in breast carcinogenesis or merely an enhancement of the methylation already present.

SST

SST (*somatostatin*) is a two exon gene located at 3q28. It has a CpG island consisting of 265 bp containing 25 CpGs. Hypermethylation of this region has been observed in esophageal, cervical, colon and gastric cancers (Jin *et al*, 2007; Ongenaert *et al*, 2008; Mori *et al*, 2006; Jackson *et al*, 2010). In primary methylated esophageal and gastric tumours reduced expression at the RNA level has also been shown (Jin *et al*, 2007; Jackson *et al*, 2010). In addition, studies of early esophageal lesions suggest *SST* methylation is an early event in this cancer (Jin *et al*, 2007). Results obtained in this thesis show that *SST* is also highly methylated in breast cancer. Methylation was detected in 59% samples by the array and 68.4% by

CoBRA assay on a separate cohort. In adjacent normal tissue, methylation was undetected by CoBRA in 31.6% (6/19) of samples while lower levels of methylation were observed in 31.6% (6/19) of normal samples compared to tumours, with 1 sample showing an equal amount of methylation in the normal as its corresponding tumour. These results suggest *SST* methylation may be an early event in breast cancer development and possibly occurs at a low level early on throughout the tissue. It would be nice to be able to look in breast DCIS samples for *SST* methylation to determine if methylation is present in these early lesions.

SST is a hormone capable of inhibiting secretion of growth factors and has been shown to be able to inhibit cell proliferation. SST interacts with one of five receptors to bring about its effects, at least four of these receptors (SSTRs) have been shown to be expressed on breast tumour cells, and patients that express SSTRs have longer DFS times (reviewed in Watt *et al*, 2008). Therapeutically, SST analogues have been used in clinical trials however no significant improvement has been observed in OS times so far (reviewed in Watt *et al*, 2008). Due to the role of *SST* in breast and other cancers, the status of *SST* hypermethylation coupled with assessment of SSTR receptor status may offer a distinguishing characteristic that could be used to decide which patients may or may not benefit from *SST* targeted treatments.

4.5.3.2 Multiple probe genes

CDKL2

CDKL2 (*cyclin dependent kinase-like 2 (CDC-2-related kinase)*) is a 12 exon gene situated at 4q21.1 with a CpG island of 713bp with 56 CpGs situated upstream of and across the transcription start site. *CDKL2* is largely unstudied and definitive functions are unknown. It is a member of a large family of CDC2-related serine/threonine kinases, known to be involved in cell cycle control. A recent study has also identified *CDKL2* as having a single nucleotide

polymorphism (SNP) that associated with early age diagnosis of breast cancer (Bonifaci *et al*, 2010). A missense mutation has also been identified in *CDKL2* in a large muti-cancer study of primary tumours (Greenman *et al*, 2007) and a truncating mutation has been observed in the breast cancer cell line HCC2218 (Stephens *et al*, 2005). *CDKL2* has shown both tumour specific methylation and re-expression following demethylation, this, combined with the potential function of a CDC2-related kinase and previously described mutational status, suggest CDKL2 is a good candidate for further analysis in breast cancer.

ZNF154

ZNF154 (zinc finger protein 154 (pHZ-92)) is a three exon gene situated at 19q13.4. It has a 328kb CpG island containing 31 CpGs positioned in the first exon and into the first intronic region. Very little is known about *ZNF154* except that it most likely functions in transcriptional regulation. Results from this thesis have shown a high level of methylation (87.5%, 64.3% of which was tumour specific). The level of methylation by CoBRA assay is very high (87.5%) and coupled with the relatively high frequency of methylation in the adjacent normal tissue (35.7%), suggests *ZNF154* methylation may be an early event in breast cancer development, however studies in DCIS samples would need to be carried out to confirm this.

4.5.4 Comparison with the MIRA study (chapter 3)

Comparisons between genes identified in the MIRA chapter and Infinium chapter does not show a large level of agreement. However, it is not particularly surprising that the same targets were not identified in both studies due to large differences in the assays and samples used. The 16 genes analysed by CoBRA in chapter 3 and Infinium in chapter 4 on the same samples were generally in agreement, and the majority that were not were being analysed in

two different regions of their CpG islands. It is interesting that, in general, all genes that were found to be methylated by CoBRA in chapter 3 were picked up at a lower frequency by Infinium, possibly due to the differing levels of criteria between candidate selection, making the Infinium selection more stringent and therefore underestimating the true level of methylation, although this is preferable to overestimation.

The Infinium HumanMethylation27 array represents an affordable method of assaying a relatively large number of patient samples. Separate studies using the same assay are also more comparable due to the consistency between probed regions as oppose to the affinity enrichment approaches. However, the use of a small number of probes per gene, irrelevant of the size of the CpG island, placed only within (or within close proximity to) CpG islands means an underrepresentation of some regions. The newly available HumanMethylation450 array should help to combat some of these issues.

In general, both methods have their advantages and disadvantages and perhaps the best approach is to perform both types of methodology on any given set of samples if enough material is available. This would provide a subset of high confidence candidates identified by both methodologies as well as those targets that may only be detected by one assay due to intrinsic aspects of each methodology.

4.5.5 Final Conclusions

The initial aims of this chapter were to use the Infinium array to create a list of hypermethylated genes and analyse potential candidates, to compare overall methylation levels between tumours with differing clinical characteristics and to identify genes where hypermethylation associates with specific clinical features.

The list of genes produced contains numerous novel potentially methylated genes in breast cancer and further work has highlighted at least two novel genes, *SST* and *CDKL2*, worth investigating further based on methylation alone without taking clinical associations into consideration. There are multiple ways a list such as this can be mined for good candidates. Because this study was primarily interested in finding clinical associations the analysis for good potential tumour suppressor gene candidates was rather superficial. These candidates would have benefited from further functional investigation to demonstrate any potential tumour suppressor activity in breast cancer. It would therefore be beneficial to have analysed the list more comprehensively for good candidates that were not necessarily associated with clinical features. This is something that could be done in the future by myself or others now this work has been published and the gene lists are in the public domain.

The analysis of clinical features on an overall and individual level has produced some interesting and clinically relevant results, in particular, the association of 6 genes with DFS. However, a particular weakness of this study is the use of such a definitive classification of hypermethylation. Although the work done suggests that what was deemed methylated and unmethylated were true, this classification was then used to assess clinical associations. To determine if these associations are real a much larger, separate cohort would need to be analysed using the same, or at least similar assay (such as pyrosequencing or EpiTYPER^{TM}). However, if even one of the 6 genes identified in this study held the association in a large cohort of samples this may be beneficial to developing markers for patient prognosis. If the information had been available, it would have been nice to identify and assess genes whose methylation status associated with drug treatment outcomes. It would have also been beneficial to know the molecular subtype of the samples with regards to luminal A/B etc. rather than just receptor status.

This work has now been published under Hill *et al* (2011) *Cancer Research* **71(**8): 2988-2999. A copy is provided at the back of this thesis.

CHAPTER 5

DNA METHYLATION ANALYSIS OF *RASSF10* IN COMMON SOLID TUMOURS

5.1 ABSTRACT

Frequent methylation of the newest RASSF family member, *RASSF10*, has been described in childhood acute lymphocytic leukaemia (ALL) (Hesson *et al*, 2009) and thyroid cancer (Schagdarsurengin *et al*, 2009) suggesting methylation of *RASSF10*, as is the case for other selected RASSF members, may be a common event in multiple cancers. To continue these studies, the work in this chapter assesses the methylation status of *RASSF10* across a panel of common solid tumours.

Frequent methylation was observed in all cell line types analysed and in lung (25%), colorectal (31%), breast (25%), kidney (25%) and glioma (57%) primary tumours. Due to the higher methylation frequency in gliomas, this study concentrated on further analysis of RASSF10 epigenetic inactivation in this cancer type. The frequency of methylation in glioma tumours differed between WHO grades; grade I astrocytomas were unmethylated, whilst grade II and III astrocytomas had methylation frequencies of 60% and 80% respectively. Grade IV gliobastoma multiforme (GBM) samples showed methylation in 65% and 69% of primary GBM and secondary GBM respectively. In secondary grade IV glioblastoma multiforme (sGBM), *RASSF10* methylation was an independent prognostic factor, associating with worse overall survival and worse progression-free survival. Very little methylation (6%) was found in non-cancerous brain tissue. Glioma cell lines showed re-expression of *RASSF10* at the mRNA and protein level following treatment with the demethylating agent, 5-azaDC.

This work has shown frequent methylation in common solid tumour types, in particular gliomas, and may offer the possibility of developing a prognostic marker for secondary GBM.

5.2 INTRODUCTION

As discussed in chapter one, the *RASSF* gene family represent a well documented family of genes with multiple links to cancer. Biological roles of the RASSFs include affecting RAS behaviour, interaction with the SWH pathway, apoptosis and cell cycle regulation. When coupled with common downregulation in various cancers, *RASSF* family gene members become attractive candidates for analysis. Because of this, the methylation status of *RASSF* family members has been extensively screened in multiple cancers. Many members of the family (*RASSF1A, RASSF2, RASSF4, RASSF6* and *RASSF10*) have now been demonstrated to be downregulated as a result of hypermethylation in various cancer types (Dammann *et al*, 2000; Dammann *et al*, 2001; Hesson *et al*, 2004; Hesson *et al*, 2005; Endoh *et al*, 2005; Eckfield *et al*, 2004; Hesson *et al*, 2009). In addition, the methylation status of *RASSF1A* has been associated with multiple clinical features of various cancers offering diagnostic and prognostic potential (Hesson *et* al, 2007). The most recently identified member, *RASSF10*, has been found to be expressed in multiple tissue types and demonstrate frequent methylation in ALL and thyroid cancers (Hesson *et al*, 2009; Schagdarsurengin *et al*, 2009). Since these represent quite diverse cancers, it seems likely that *RASSF10* methylation may be a common event in many cancer types. The work in this chapter focuses on analysis of *RASSF10* DNA methylation in common solid tumour types.

5.3 AIMS

1. To assess the methylation status of *RASSF10* in common solid tumour cell lines.
2. To assess the methylation status of *RASSF10* in primary tumour types of methylated cell lines and to determine tumour specificity of methylation in positive tumour types.
3. To assess the methylation status of *RASSF10* in relation to clinical features.

5.4 RESULTS

5.4.1 Methylation analysis of solid tumour cell lines

Due to frequent methylation of *RASSF10* in childhood ALL (Hesson *et al*, 2009) and thyroid cancer (Schagdarsurengin *et al*, 2009) the panel of available solid tumour cell lines were analysed at the same region (figure 5.1). Frequent methylation was observed in all cell line types analysed; breast (33%), colorectal (83%), lung (40%), kidney (33%) and glioma (83%) (figure 5.2; table 5.1). For one of the highest methylated cancer cell line types, gliomas, no *RASSF10* expression was observed at the mRNA level of methylated cell lines and re-expression was observed following treatment with the demethylating agent 5-azaDC (figure 5.3). In addition, further work by Dr. Nick Underhill-Day showed re-expression of RASSF10 protein in 5-azaDC treated glioma cell line A172 and expression pre and post treatment in the unmethylated cell line, H4 (figure 5.3). These results show that *RASSF10* is commonly methylated in a range of common solid tumour cell lines and that methylation is capable of suppressing expression of *RASSF10* in gliomas in addition to thyroid cancers and childhood ALL.

5.4.2. Methylation analysis of primary solid tumour samples

All tumour types showing methylation in cell lines were analysed in primary tumours and in addition, because methylation was high in all cell line types analysed, some tumours were also analysed for which cell lines were unavailable. No methylation was observed in medulloblastomas (0/21) and infrequent methylation was observed in ovarian tumour samples (3/18; 17%). More frequent methylation was observed in lung, breast, kidney colorectal and glioma tumours. Lung tumours showed methylation in 4/20 (25%) primary tumour samples and 0/4 (0%) of non-cancerous adjacent lung tissue for the methylated

Figure 5.1_*RASSF10* CoBRA region

A schematic of *RASSF10* gene structure as determined by Hesson *et al*, 2009 is shown in blue with the CpG island below in green. The region analysed by CoBRA is indicated with dashed lines and the sequence displayed underneath. CpGs assessed in the CoBRA region are numbered 1-61 and BstUI sites are in red. Primer locations are underlined.

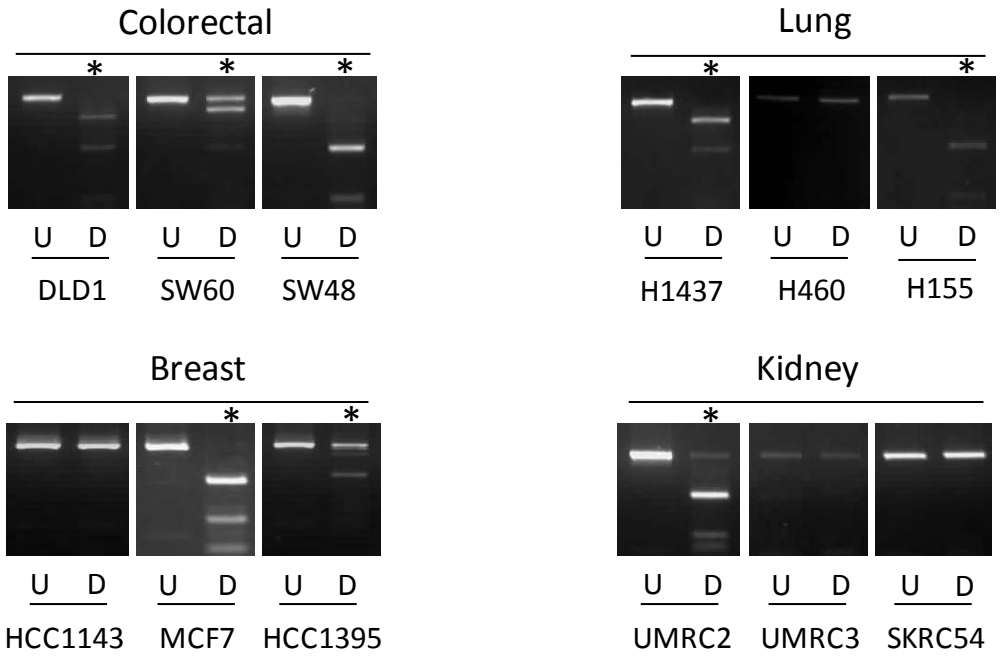


Figure 5.2_*RASSF10* CoBRA results for solid tumour cell lines
CoBRA results are shown with undigested PCR product (U) run next to digested PCR products (D). Three samples each are shown for colorectal, breast, lung and kidney cancer cell lines. * indicates methylated samples

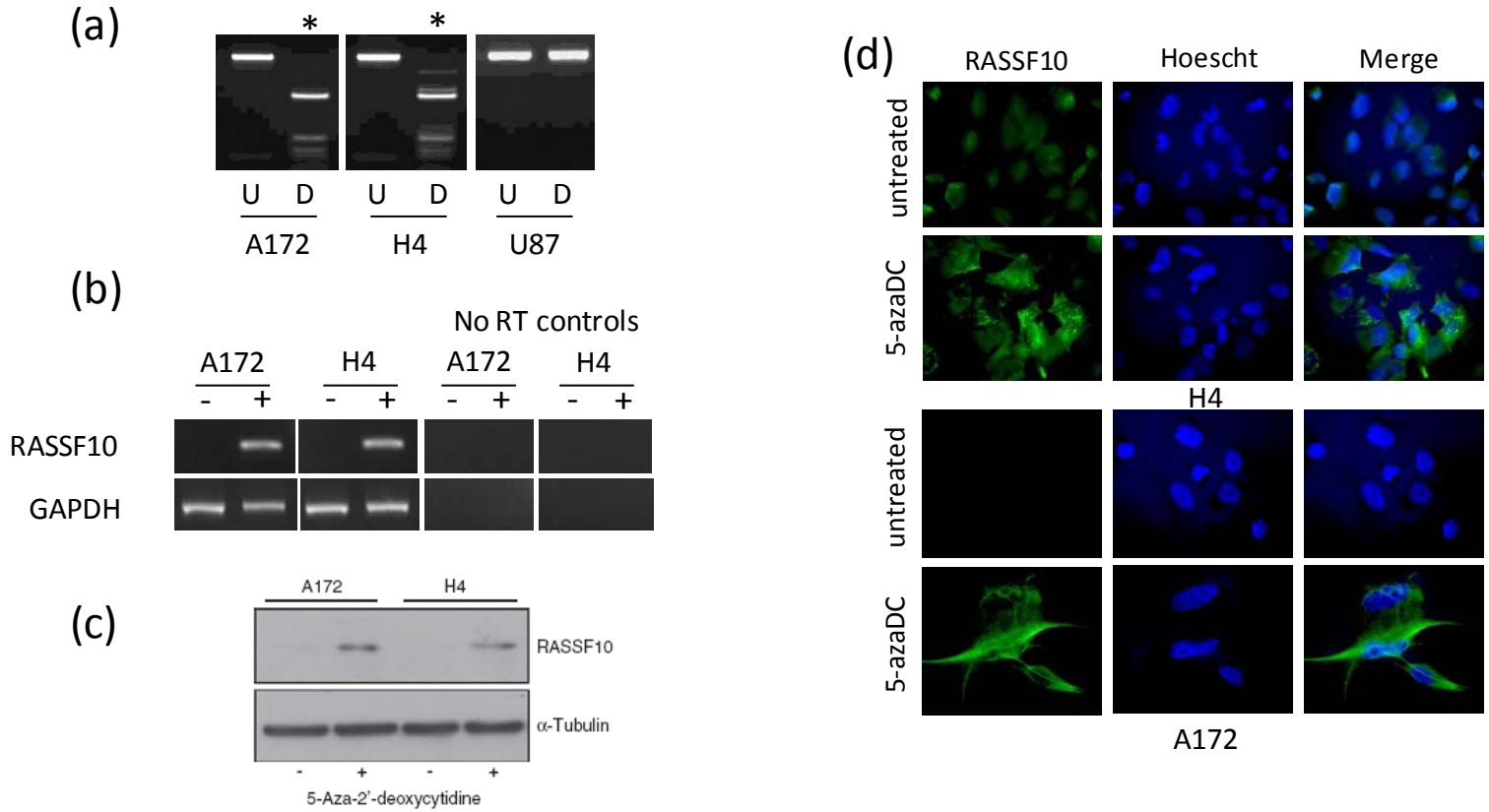


Figure 5.3_Methylation analysis of glioma cell lines

CoBRA results are shown for glioma cell lines with undigested PCR products (U) run next to digested PCR products (D). * indicates methylated samples (a). Expression results are shown pre (-) and post (+) treatment of 5-azaDC. Results are shown for both *RASSF10* and *GAPDH*. No RT controls are also shown for *RASSF10* and *GAPDH* to show there is no DNA contamination as *RASSF10* is a single exon gene (b). (c) RASSF10 protein expression can be seen only after treatment with 5-azaDC by Western blot analysis and (d) Immunoflourescence results are shown for H4 and A172 untreated and treated with 5azaDC. Work done for (c) and (d) was carried out by Dr. Nick Underhill-Day.

samples, therefore showing tumour specific methylation. Breast tumour samples showed tumour specific methylation in 4/16 (25%) tumour normal paired samples (figure 5.4). Kidney tumours showed methylation in 4/16 (25%) and no methylation was seen in 6/8 normal kidney DNA samples. Colorectal tumour samples showed a slightly greater level of methylation in 5/16 (31%) of samples (figure 5.4; table 5.1), unfortunately adjacent normal tissue was unavailable for these samples. The greatest level of methylation in primary tumours was detected in gliomas and therefore further work concentrated on this tumour type.

5.4.3 Methylation analysis of primary glioma samples

Glioma world health organisation (WHO) grades I-III astrocytomas and grade IV glioblastoma multiforme (GBM) samples were analysed. Grade I astrocytomas were unmethylated (0/10), grades II and III astrocytomas were methylated in 6/10 (60%) and 8/10 (80%) respectively (table 5.1). Grade IV GBM can be classified as either primary GBM (pGBM) or secondary GBM (sGBM) depending on whether earlier grade lesions were present that progressed to grade IV GBM (sGBM) or whether the tumour presented at grade IV without early lesions (pGBM). There is evidence that pGBM and sGBM have distinct genetic profiles (Parsons *et al*, 2008; Yan *et al*, 2009) and so have been analysed separately for *RASSF10* methylation status. Both forms of grade IV GBM tumours showed a lower level of methylation than grade III astrocytomas. pGBM showed methylation in 13/20 (65%) of samples and sGBM showed a comparable level of methylation, 9/13 (69%) of samples. Very little methylation was observed in non-cancerous brain tissue (1/15, 6%) (table 5.1). Clone sequencing of glioma samples from different grades shows methylation is widespread throughout the region being analysed and that the CoBRA results are representative of the wider levels of methylation (figures 5.5 and 5.6). For 7 of the 9 methylated sGBM samples, DNA from the early grade lesion (grade II) was also available.

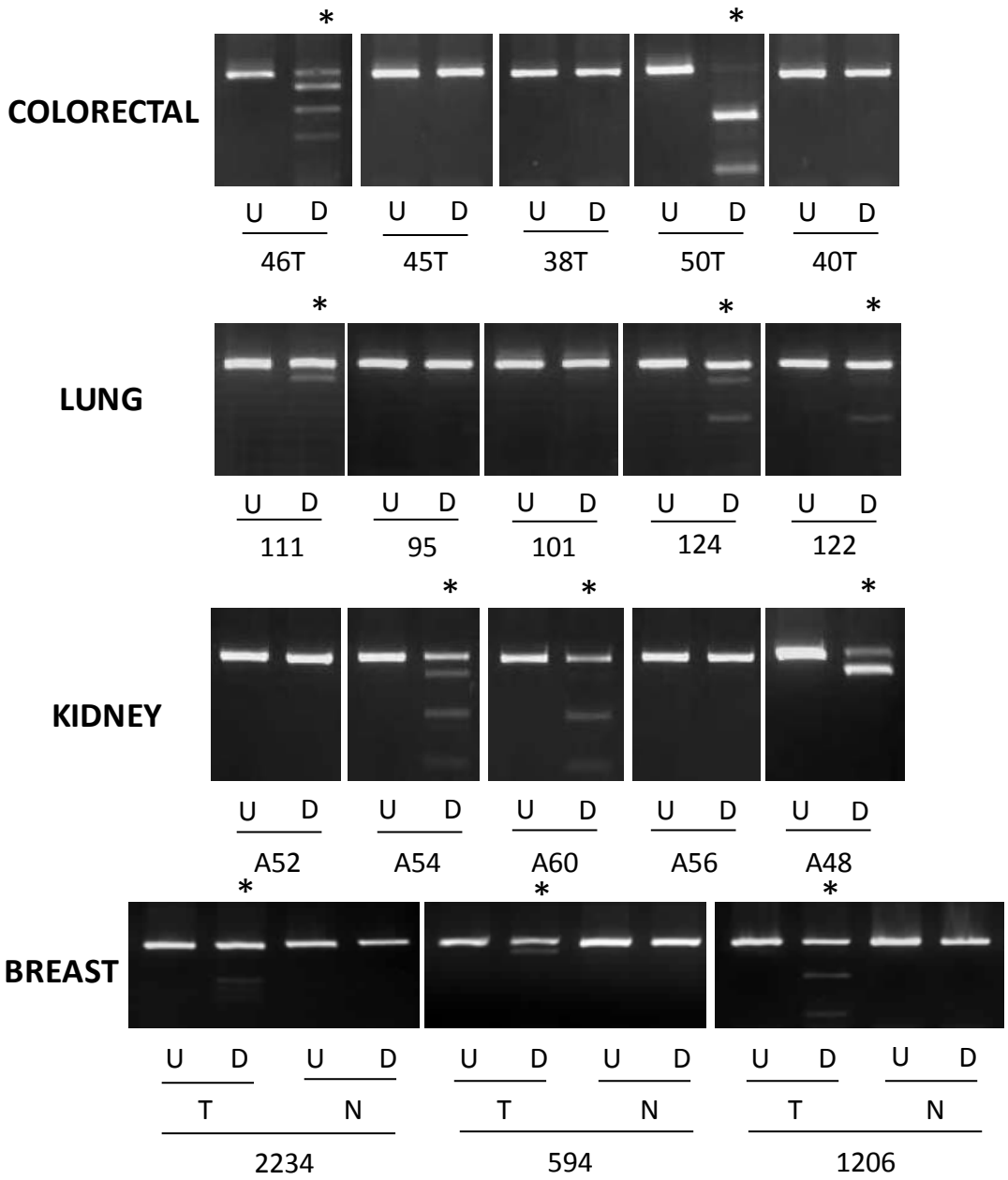


Figure 5.4_CoBRA analysis of primary colorectal, lung, kidney and breast samples

CoBRA results are shown for five each of primary colorectal, lung and kidney tumour samples. Three breast tumour (T) and adjacent normal (N) breast tissue paired samples are shown for breast cancer. All CoBRA results are shown with undigested PCR products (U) run next to digested PCR products (D). * indicates methylated samples.

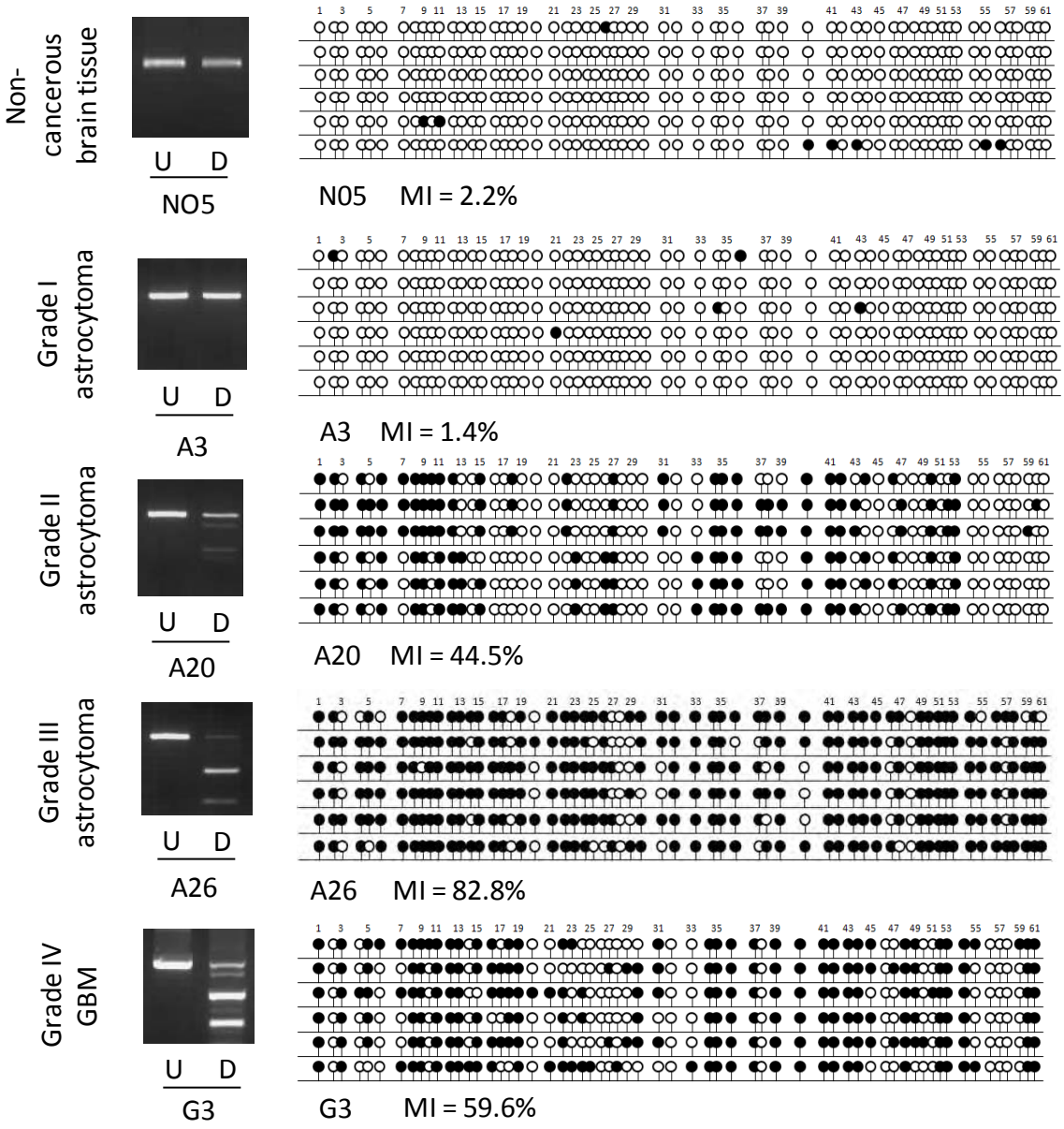


Figure 5.5_CoBRA and sequencing of primary glioma tumour samples and non-cancerous brain tissue I

CoBRA results are shown for five samples, one each of non-cancerous brain tissue, grades I, II and III astrocytomas and grade IV GBM. In all cases, undigested PCR products (U) are run next to digested PCR products (D). Results from clone sequencing are also shown for each sample. Each line represents a single allele, white and black circles represent unmethylated and methylated CpGs respectively. MIs are calculated as a percentage of methylated CpGs out of the total number of CpGs analysed.

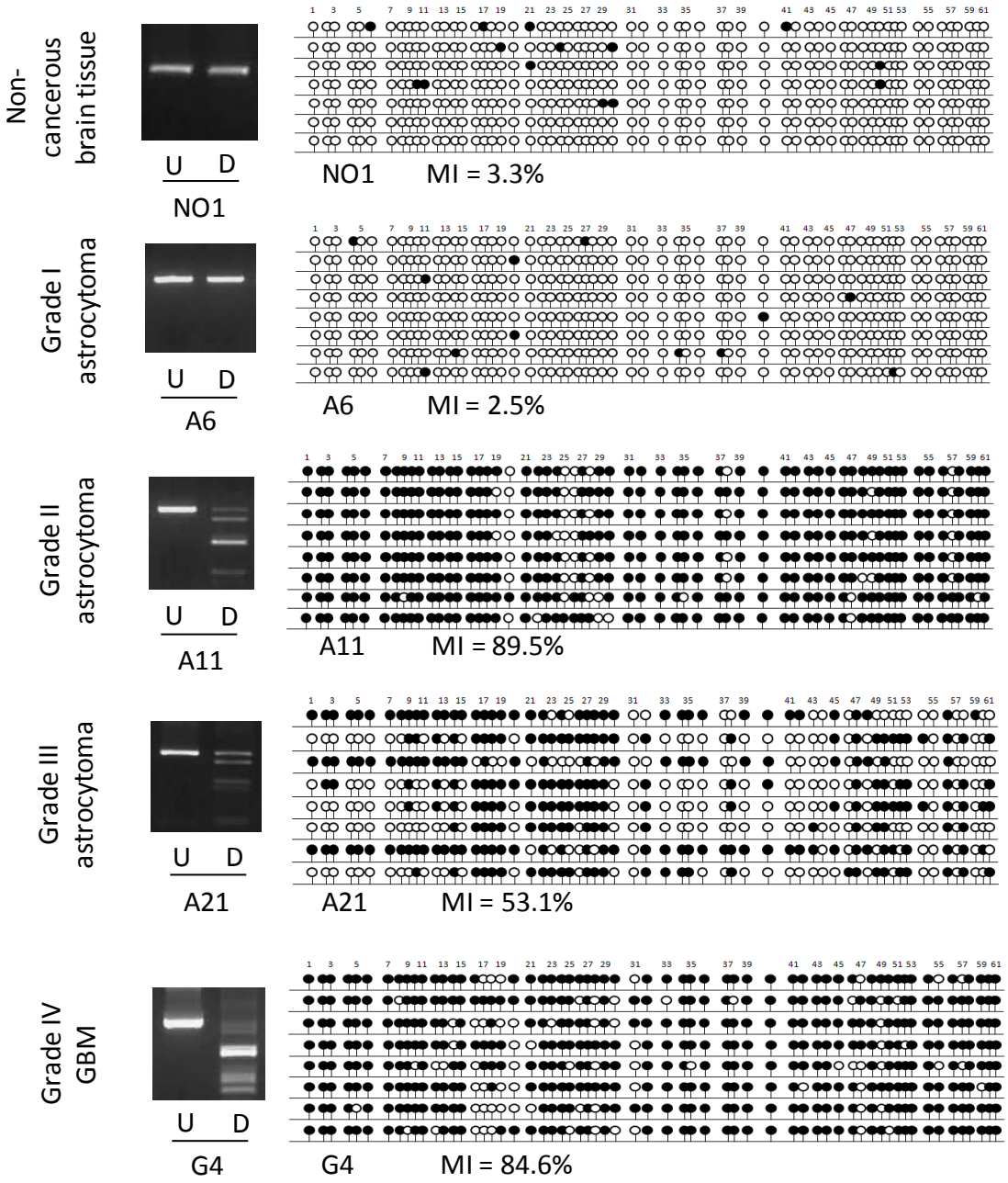


Figure 5.6_CoBRA and sequencing of primary glioma tumour samples and non-cancerous brain tissue II

CoBRA results are shown for five samples, one each of non-cancerous brain tissue, grades I, II and III astrocytomas and grade IV GBM. In all cases, undigested PCR products (U) are run next to digested PCR products (D). Results from clone sequencing are also shown for each sample. Each line represents a single allele, white and black circles represent unmethylated and methylated CpGs respectively. MIs are calculated as a percentage of methylated CpGs out of the total number of CpGs analysed.

Assessment of *RASSF10* methylation status in these early lesions showed methylation in 6 out of the 7 samples, suggesting *RASSF10* methylation is an early event in sGBM gliomagenesis (figure 5.7).

Methylation of another RASSF family member, *RASSF1A*, is a common event in glioma tumours (Hesson *et al*, 2004), as is methylation of O-6-methylguanine-DNA methyltransferase (*MGMT*) (Esteller *et al*, 1999). To determine if *RASSF10* methylation is independent of these events the methylation status of *RASSF1A* and *MGMT* was determined for this cohort of gliomas (only pGBM, not sGBM samples were analysed). *RASSF1A* was methylated in 84% of samples and *MGMT* was methylated in 26% of samples (figure 5.8). No associations were identified between *RASSF10* methylation and either *RASSF1A* or *MGMT*. In addition, mutations in isocitrate dehydrogenase 1 and 2 (*IDH1* and *IDH2*) genes are prevalent in gliomas, particularly sGBM, and associate with good prognosis (Parsons *et al*, 2008; Yan *et al*, 2009). Work done by K Robel in Dr. Dietmar Krex's laboratory (Klinik und Poliklinik für Neurochirurgie, Universitätsklinikum Carl Gustav Carus, Technische Universität Dresden, 01307 Dresden, Germany) identified *IDH1* mutations in 10 of the 20 grade II and III astrocytomas, 4 out of 10 sGBM and 2 out of 20 pGBM samples, no *IDH2* mutations were detected in any samples from this cohort. No associations were observed between *IDH* mutation status and either *RASSF10* or *RASSF1A* methylation.

5.4.4 Association of *RASSF10* methylation with survival

Overall survival (OS) and progression-free survival data was available for glioma tumours used in this study. *RASSF10* methylation in sGBM tumours associated significantly with worse OS and worse PFS (p=0.0318 and p=0.0231 respectively) (figure 5.9). With Cox proportional hazard regression analysis, *RASSF10* methylation and age at first operation were

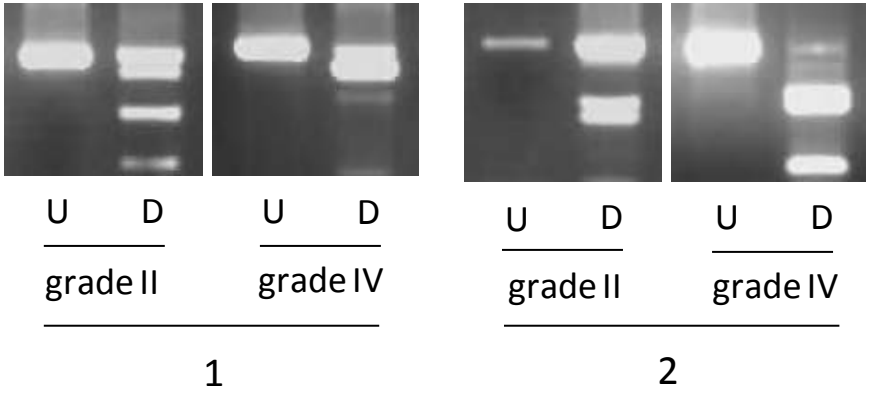


Figure 5.7_*RASSF10* methylation at grade II and IV gliomas
CoBRA results are shown for grade II and grade IV lesions for two patients. Results are shown with undigested PCR products (U) run next to digested PCR products (D).

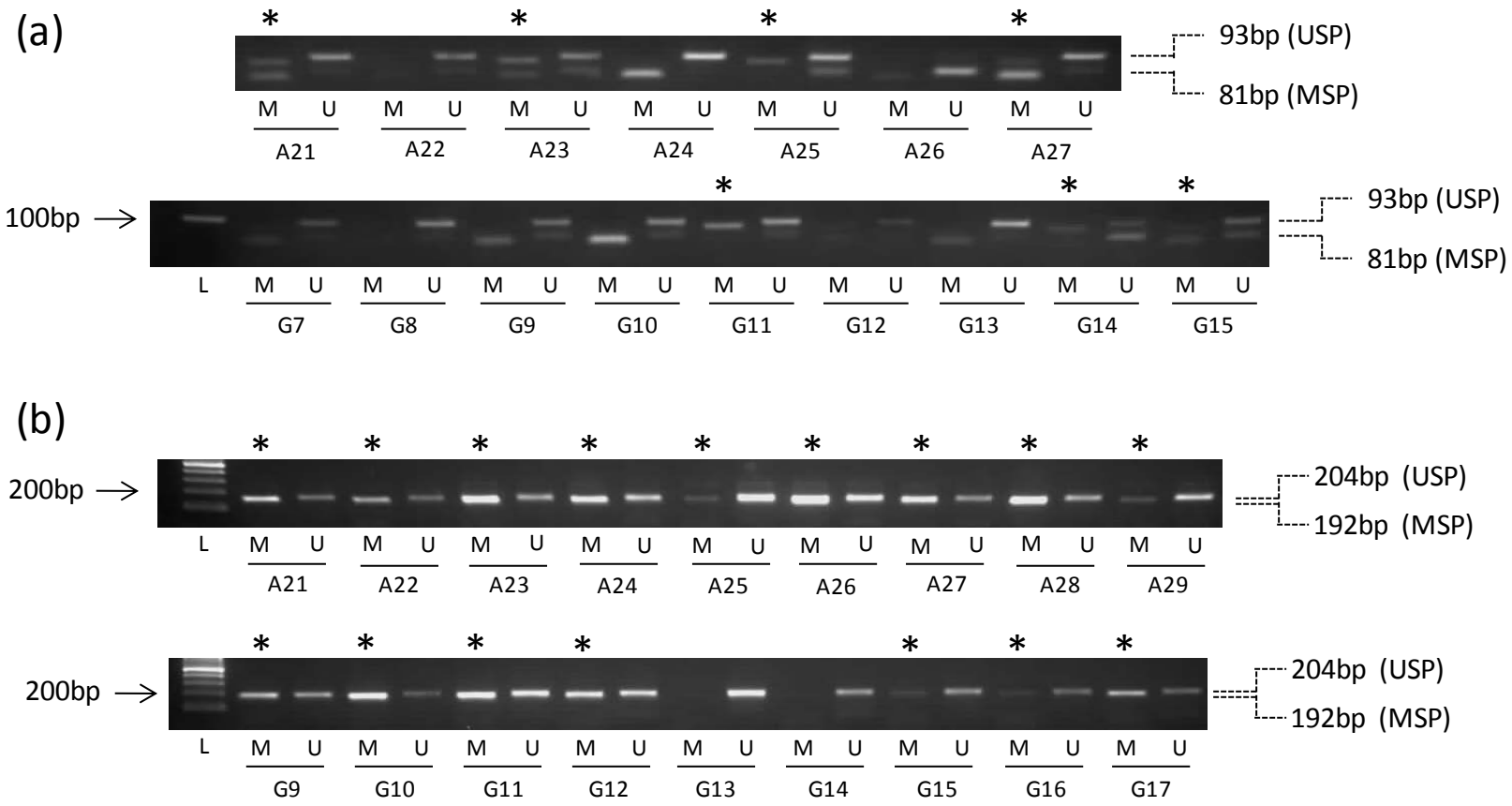


Figure 5.8_MSP analysis of *RASSF1A* and *MGMT*
Selected MSP results are shown for *RASSF1A* (a) and *MGMT* (b). MSP results (M) are run next to USP results (U). In some cases a ladder (L) is also shown.

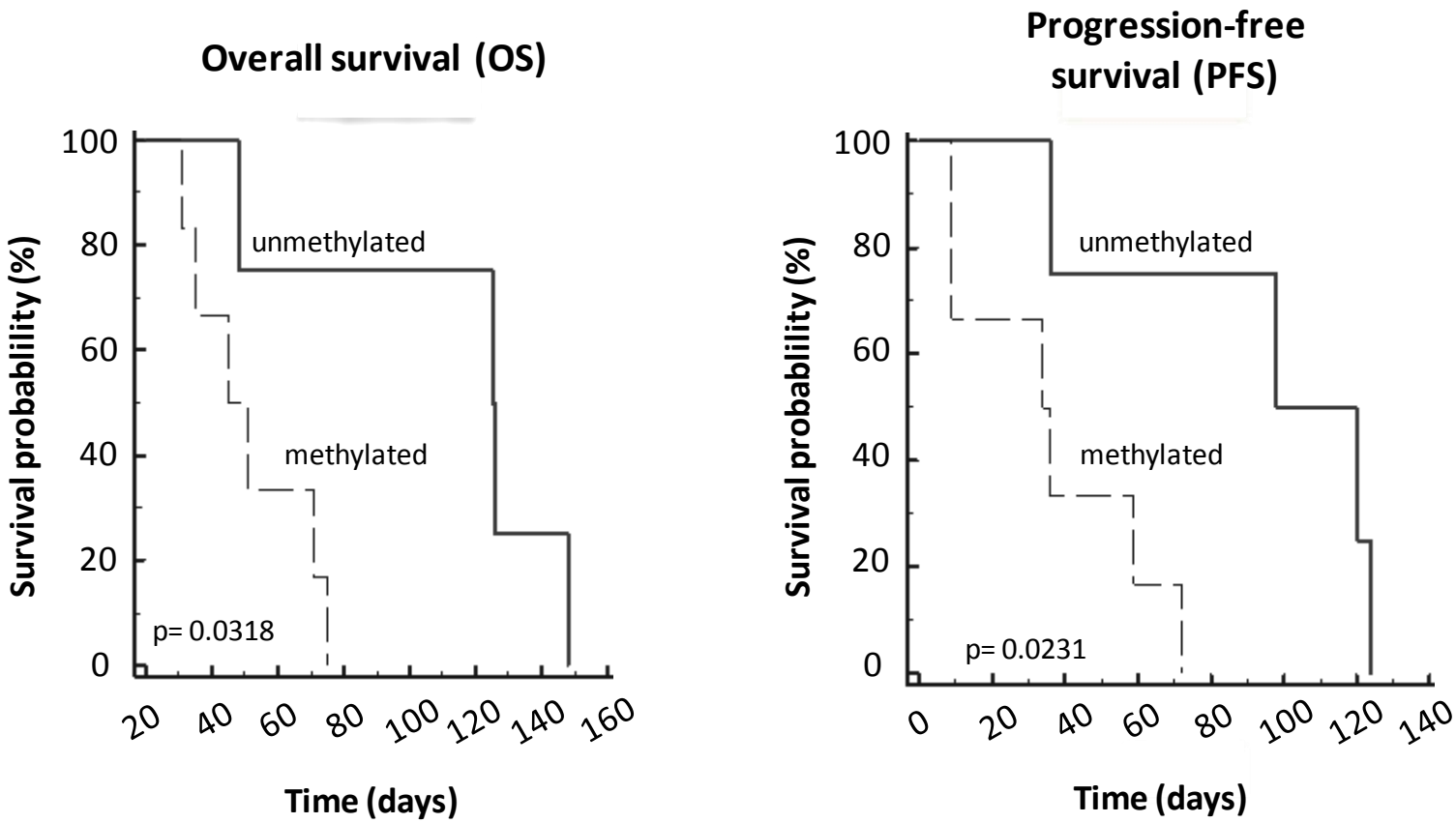


Figure 5.9_Kaplan-Meier analysis for *RASSF10* in sGBM tumours
Graphs are shown for Kaplan-Meier analysis of *RASSF10* methylation in sGBM tumours. An association with overall survival (OS) and progression-free survival (PFS) was observed.

independent prognostic factors for OS (p=0.01952 and p=0.02404 respectively and PFS (p=0.02724 and p=0.3223 respectively). *IDH1* mutations were an independent prognostic factor for OS (p=0.02699) but did not reach significance for PFS (p=0.06723).

5.5 DISCUSSION

Results from this study have added to those by Hesson *et al*, 2009 and Schagdarsurengin *et al*, 2009 to show that *RASSF10* methylation is a common event in many tumour types, particularly gliomas.

5.5.1 *RASSF10* methylation

RASSF10 methylation was initially identified in childhood ALL samples (Hesson *et al*, 2009). Results showed that methylation was highly prevalent in T-ALL (80%) but not B-ALL (17%). Leukaemia cell lines were also highly methylated (100%) and showed re-expression of *RASSF10* following treatment with 5-azaDC. A subsequent study in thyroid cancer identified frequent methylation in both cell lines (100%) and primary tumours (66%) (Schagdarsurengin *et al*, 2009). Expression levels of *RASSF10* mRNA were found to be lower in thyroid cancer cell lines than normal human thyroid tissue. Expression could be re-expressed in thyroid cancer cell lines using 5-azaDC. The authors found differing frequencies of methylation in different subtypes of primary thyroid cancer, the highest being in medullary thyroid carcinoma (100%) and the lowest in undifferentiated thyroid cancer (40%). Due to the neuroendocrine basis of medullary thyroid carcinoma it was suggested by the authors that there may be an association between *RASSF10* methylation and tumours with a neuroendocrine origin. Results from the current study have now shown frequent methylation in lung, colorectal, kidney, breast and glioma tumours, ranging from 25% in lung and breast tumours up to 80% in grade III astrocytomas. Methylation of glioma tumours proved particularly interesting, with high but differing levels of methylation in multiple stages of disease. *RASSF10* methylation appears to occur as an early event and associated with worse OS and PFS in this cohort of sGBM, although the sGBMs available for survival analysis and

those with corresponding early lesions formed only a small cohort and therefore the association with survival and analysis in early and late lesions of the same patient would benefit from being observed in a separate, larger cohort. Table 5.1 shows a summary of all methylation data known to date for *RASSF10*. All data now available on *RASSF10* methylation in various cancers shows methylation is a common event in many cancers, the most frequent being childhood ALLs, thyroid cancers and gliomas. Only medulloblastomas and ovarian cancers have shown no or infrequent methylation. Some non-small cell lung cancers (NSCLCs) show neuroendocrine features whilst small cell lung cancers (SCLCs) do not, based on the higher level of methylation in thyroid cancers with a neuroendocrine origin it would be interesting to see whether SCLC and NSCLC show different frequencies of *RASSF10* methylation. All methylation work in this chapter was carried out by CoBRA and in some cases bisulphite sequencing, however, due to the common nature of *RASSF10* methylation across multiple cancer types the setting up of an MSP or pyrosequencing assay may benefit future work in this area.

5.5.2 *RASSF10* methylation in gliomas

Brain and central nervous system (CNS) tumours account for a relatively small proportion of all cancers diagnosed in the UK (4785 cases out of 309,527; 1.5%) and, due to the generally bad prognosis of these tumours, a slightly larger proportion of cancer related deaths (3674 out of 156,232; 2.4%) (CRUK cancer statistics). The majority of brain and CNS cancers are gliomas, consisting of astrocytomas (including glioblastoma multiforme), ependymomas, oligodendrogliomas and mixed gliomas, which consist of a mixture of the other subtypes. Of these, astrocytomas/glioblastoma multiforme are the most frequent in adults, accounting for more than 50% (Adamson *et al*, 2009). Prognosis is poor for patients diagnosed with glioblastomas (grade IV astrocytomas/glioblastoma multiforme), the median survival time

CANCER	SUBTYPE	METHYLATION FREQUENCY	REFERENCE
Childhood acute lymphocytic leukaemia	cell lines	7/7 (100%)	Hesson *et al*, 2009
	B-ALL	8/51 (16%)	
	T-ALL	23/26 (88%)	
	Pre-B-ALL	0/1 (0%)	
	Unclassified childhood leukaemias	0/5	
Thyroid cancer	cell lines	9/9 (100%)	Schagdarsurengin *et al*, 2009
	follicular thyroid carcinoma	5/10 (50%)	
	papillary thyroid carcinoma	9/12 (75%)	
	medullary thyroid carcinoma	5/5 (100%)	
	unclassified thyroid carcinoma	2/5 (40%)	
Breast cancer	cell lines	33%	This study
	primary tumours	4/16 (25%)	
Lung cancer	cell lines	40%	This study
	NSCLC	4/20 (25%)	
Colorectal cancer	cell lines	83%	This study
	primary tumours	5/16 (31%)	
Kidney cancer	cell lines	33%	This study
	primary tumours	4/16 (25%)	
Glioma	cell lines	83%	This study; Hill *et al*, 2011
	Grade I astrocytomas	0/10 (0%)	
	Grade II astrocytomas	6/10 (60%)	
	Grade III astrocytomas	8/10 (80%)	
	Grade IV pGBM	13/20 (65%)	
	Grade IV sGBM	9/13 (69%)	
Ovarian	primary tumours	3/18 (17%)	This study
Medulloblastomas	primary tumours	0/21 ((0%)	This study

Table 5.1 Methylation frequency of *RASSF10* in multiple cancers
Methylation frequencies of *RASSF10* in all cancers assessed at the time of writing.

after diagnosis being in the range of only 14 months. For this reason, GBM genetic and epigenetic alterations have been extensively studied. GBM tumours are most commonly primary GBM (pGBM), presenting as *de novo* tumours rather than with earlier lesions and are most prevalent in males (Adamson *et al*, 2009). In contrast, secondary GBM (sGBM) tumours account for approximately 5% of GBM cases and present with earlier grade II or III astrocytoma lesions. sGBM are more evenly distributed between males and females and are more prevalent in younger patients (Adamson *et al*, 2009). sGBM generally have a better prognosis than pGBM with greater overall survival times. Genetic alterations in GBM include overexpression of *EGFR* (Watanabe *et al*¸1996), inactivating mutation of *TP53* (Ishii *et al*, 1999) and loss of heterozygosity (LOH) at 10q25 (Rasheed *et al*, 1995). Recent sequencing studies have also identified frequent mutations in the *isocitrate dehydrogenase 1* and *2* genes (*IDH1* and *IDH2*) that associate with good prognosis and are more prevalent in sGBM (Parsons *et al*¸2008; Yan *et al*, 2009).

Hypermethylation is a common event in gliomas and methylation has been described in many genes including *RASSF1A* (Hesson *et al*, 2004), *HIC1* (Li *et al*, 1998), *PTEN* (Baeza *et al*, 2003) and *RB1* (Nakamura *et al*, 2001) amongst many others (reviewed in Hesson *et al*, 2008). Whilst not necessarily the highest methylated gene in gliomas, the most clinically significant is *MGMT*. *MGMT* is a DNA repair enzyme that removes alkyl groups from the O6 position of guanine. Alkylation at this position often results in cross-linking between DNA strands, causing cell death when not repaired. Some cancer drugs exploit this mechanism and cause alkylation of the DNA resulting in cell death. As *MGMT* inhibits cross-linking by repairing alkylation, inactivation of *MGMT* is beneficial for alkylating agents as the cross-linking mechanism is no longer inhibited. Methylation of *MGMT* in gliomas was first described as being prevalent in 40% of gliomas (Esteller *et al*, 1999) and was later shown to

associate with a positive response to alkylating agents (Esteller *et al*, 2000a). The methylation status of *MGMT* is an excellent example of how DNA hypermethylation can be used to aid treatment and prognosis.

There have been numerous studies that have assessed the glioma epigenome, recent studies include Martinez *et al*, 2009; Noushmehr *et al*, 2010 and Wu *et al*, 2010. In the study by Noushmehr *et al*, 2010 as part of a TCGA (the cancer genome atlas) group study, a subset of tumours were identified with a characteristic DNA methylation profile similar to the CpG island methylator phenotype (CIMP) previously observed in colorectal cancers. For this reason, it was termed glioma-CIMP (G-CIMP). G-CIMP tumours were found to associate with the proneural expression group previously identified in another TCGA group study (Verhaak *et al*, 2010), be of lower age at diagnosis and have better OS. *IDH1* mutations were also associated with G-CIMP tumours (Noushmehr *et al*, 2010). The study by Wu *et al*, 2010, used the MIRA assay to assess the epigenome of the tumours used in this chapter (the 30 astrocytomas and 20 pGBM samples) and found many DNA methylation alterations. The study also found that methylation was of increasing prevalence from grade I to III astrocytomas and then dropped slightly in pGBM, in accordance with the findings for *RASSF10*.

The results from this study add to the ever growing list of hypermethylated genes in gliomas, which, when expression is lost, is likely to have a detrimental effect in the glioma cell. In addition, *RASSF10* methylation appears to be an early event in sGBM and could have potential for prognosis, although a separate and much larger cohort of samples needs to be analysed to provide more evidence for this.

5.5.3 *RASSF10* as a tumour suppressor

High methylation of the newest RASSF member, *RASSF10*, in multiple cancers suggests it, like the classical RASSFs, is likely to behave as a tumour suppressor. In addition to the work described in this chapter Dr. Nick Underhill-Day carried out some functional work in cancer cell lines to demonstrate the importance of loss of RASSF10 in the tumourigenic process.

Using an in-house generated RASSF10 specific antibody, RASSF10 was shown to localise primarily within the cytoplasm, with relocation to the nucleus during mitosis (Hill *et al*, 2011). siRNA knockdown of RASSF10 in expressing glioma cell lines showed a significantly increased ability to metabolize resazurin ($p<0.001$), suggesting loss of RASSF10 leads to an increase in the viable cell population. In addition, loss of RASSF10 appeared to lead to an increase in DNA synthesis and thus cell proliferation (Hill *et al*, 2011).

Overexpression of RASSF10 in non-expressing (due to hypermethylation) glioma cell lines U343 and H4 showed reduced colony forming ability compared to their empty vector transfected counterparts ($p=0.0098$ and $p=0.0174$ respectively) (figure 5.10). Knockdown of RASSF10 showed an increased ability to grow in soft agar, showing that RASSF10 can inhibit growth in an anchorage-independent manner. Overexpression of RASSF10 in U343 and H4 cell lines did not show any reduced ability to migrate during a scratch would-healing assay. Taken together, these results suggest RASSF10 is capable of suppressing growth in cancer cells but does not affect cell migration (in glioma cells) (Hill *et al*, 2011). In combination with the hypermethylation of *RASSF10* in many cancer types, these results provide promising evidence towards a tumour suppressor role for *RASSF10*.

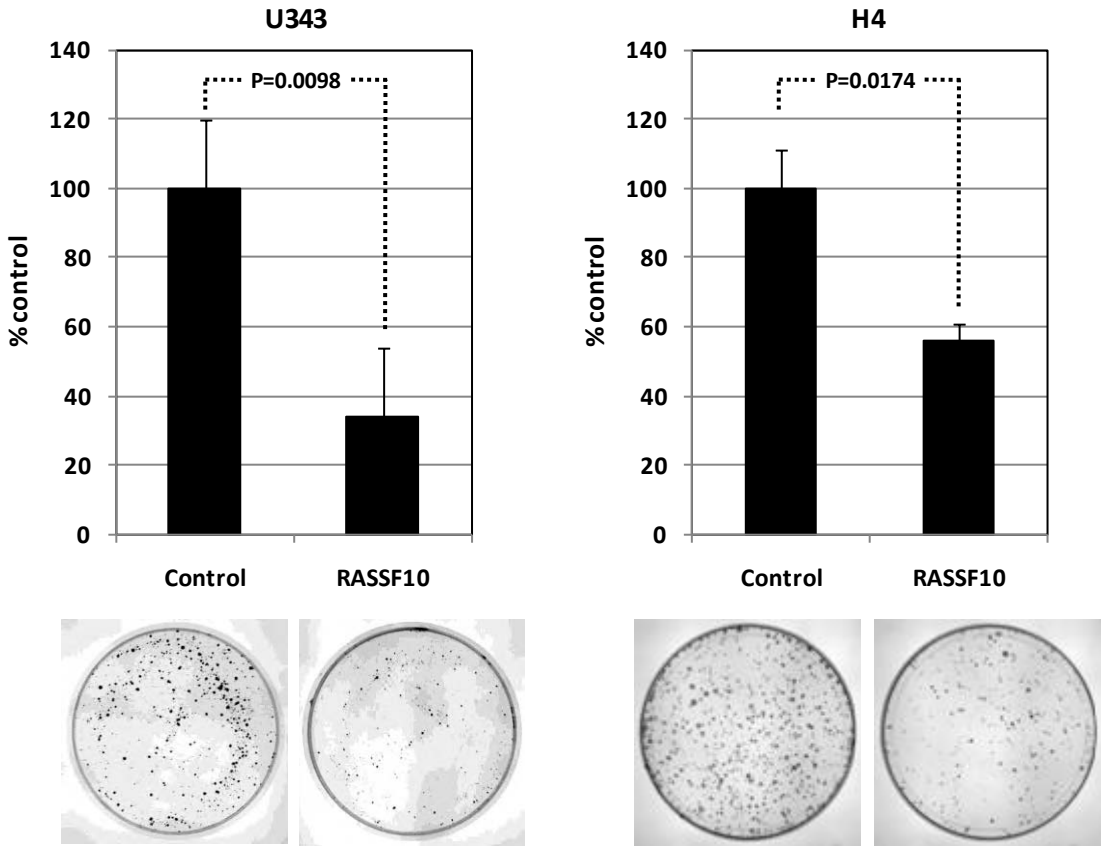


Figure 5.10_RASSF10 colony forming ability

RASSF10 overexpressing U343 and H4 cell lines show reduced colony forming ability compared to control transfected cell lines. Graphs are shown for each cell line with colony forming ability of RASSF10 overexpressing cells as a percentage of control transfected cells. Representative photographs are also shown underneath in each case. This work was carried out by Dr. Nick Underhill-Day and Dr. Mark Morris.

5.5.4 *RASSF10* as a member of the N-terminal RASSFs

The high levels of hypermethylation observed for *RASSF10* in multiple cancers and the tumour suppressor effects of RASSF10 on the growth of cancer cell lines suggests that *RASSF10* is a candidate tumour suppressor gene (Hill *et al*, 2011). Although there is now abundant evidence for the tumour suppressor roles of classical RASSF family members there is much less evidence for the N-terminal members.

Hypermethylation of the N-terminal RASSFs is uncommon. *RASSF7* has a large CpG island that has been assessed for hypermethylation in many cancers but none has been detected (Recino *et al*, 2010). *RASSF8* does not have a CpG island according to UCSC, however, use of EBI CpG plot detects a CpG island upstream of and across the transcription start site. A region of this CpG island has been assessed for methylation in many cancer cell lines however no methylation was detected (Lock *et al*, 2010). Since *RASSF9* does not have a CpG island, *RASSF10* is the only N-terminal RASSF to be hypermethylated in tumours. However, *RASSF8* has been described as a candidate tumour suppressor gene for many reasons (reviewed in Underhill-Day *et al*, 2011). Amongst other reasons, it has been shown to be downregulated in lung adenocarcinoma and able to inhibit anchorage-independent growth when re-introduced into lung cancer cell lines (Falvella *et al*, 2006). In addition, *in vivo* work has shown enhanced cell proliferation in *Xenopus laevis* when overexpressed and solid tumour formation in mice when RASSF8-depleted cells were introduced into SCID (severe combined immunodeficiency) mice (Lock *et al*, 2010). In contrast to *RASSF8*, *RASSF7* has been shown to be upregulated in many cancers including pancreatic ductal carcinoma, islet cell tumours, endometrial cancer and ovarian clear cell carcinoma (Brandt *et al*, 2004; Friess *et al*, 2003; Logsdon *et al*, 2003; Lowe *et al*, 2007; Mutter *et al*, 2001; Tan *et al*, 2009). It is thought the upregulation of *RASSF7* may be a by-product of hypoxic conditions in solid

tumours (Underhill-Day *et al*, 2011; Sherwood *et al*, 2010). RASSF7 has been shown to be required for mitosis and therefore cell cycle progression and cell survival (Sherwood *et al*, 2008) and has been implicated in nectroptosis (Hitomi *et al*, 2008). Little is known about RASSF9 function, although it has been shown to associate with R-RAS, K-RAS and N-RAS (Rodriguez-Viciana *et al*, 2004).

The work described in this chapter has identified frequent methylation of *RASSF10* in many common solid tumour types, in particular gliomas, which has shown a very high frequency of methylation and appears to occur at an early stage of sGBM tumours. Although further work is needed in a much larger cohort, *RASSF10* may provide prognostic value in sGBM tumours. In addition, the further work done by Dr. Nick Underhill-Day has shown tumour suppressor properties of RASSF10. This work adds to the little known information on N-terminal RASSFs, providing further evidence for tumour suppressor properties of N-terminal as well as classical RASSF members.

This work has now been published under Hill *et al*, 2011 *Oncogene* **30**(8):978-989. A copy of this publication is provided at the back of this thesis.

CHAPTER 6

DNA METHYLATION ANALYSIS OF THE NEWLY IDENTIFIED SALVADOR/WARTS/HIPPO PATHWAY MEMBER, *KIBRA*

6.1 ABSTRACT

The Salvador/Warts/Hippo (SWH) pathway was identified in *Drosophila melanogaster* as a control mechanism for organ and tissue size (Tapon *et al*¸ 2002). This function is conserved in mammals and is often deregulated in cancer. Recently, the WW-domain protein gene, *Kibra* was identified as part of the pathway in *D.melogaster* (Yu *et al*, 2010; Genevet *et al*, 2010; Baumgartner *et al*, 2010). The work in this chapter has assessed the methylation status of the human homologue, *KIBRA*, across a range of cancer cell lines. Frequent methylation was detected in leukaemia cell lines but not solid tumour cell lines. Methylation correlated with gene silencing which could be reversed by treatment with the demethylating agent 5-aza-2'-deoxycitidine. Primary childhood acute lymphocytic leukaemia samples showed a high frequency of methylation, predominantly in B-cell acute lymphoblastic leukaemia (B-ALL) rather than T-cell acute lymphoblastic leukaemia (T-ALL) samples (70% compared to <20%; p=0.0019). In addition, methylation associated with the *ETV6/RUNX1* {t(12;21)(p13;q22)} chromosomal translocation (p=0.0082), suggesting that *KIBRA* may play an important role in t(12;21) leukaemogenesis. In B-ALL paired samples at diagnosis and remission *KIBRA* methylation was seen in diagnostic but not in any of the remission samples. In accordance with this, *KIBRA* mRNA expression was only observed in remission samples. *KIBRA* methylation is therefore common in B-ALL and linked to a specific genetic event.

6.2 INTRODUCTION

As discussed in chapter one, the Salvador/Warts/Hippo (SWH) pathway is a controller of tissue and organ size which, when deregulated, can have tumourigenic effects. When active, the upstream members of the pathway function to suppress the activity of the YAP1/TAZ-TEAD transcription factor complex. Mammalian upstream members of the pathway have been shown to act in a tumour suppressor role. Deregulation of these members are involved in neurofibromatosis through mutation of *NF2*, soft tissue sarcoma through hypermethylation of *MST1/2*, melanomas through *MOBKL1A/B*, and astrocytoma and breast cancers through hypermethylation of *LATS1/2* (Rouleau *et al*, 1993; Tapon *et al*, 2002; Lai *et al*, 2005; Seidel *et al*, 2007; Jiang *et al* 2006; Takahashi *et al*, 2005; Kosaka *et al*, 2007; Sasaki *et al*, 2007). In addition, *YAP1* overexpression has been observed in lung, medulloblastomas and oespophageal cancers (Dai *et al*, 2003; Fernandez-L *et al*, 2009; Imoto *et al*, 2009). Recently, a new human member of the SWH pathway was identified, *KIBRA* (*kidney and brain protein*), also known as *WWC1* (*WWC and C2 domain containing 1*), the fly homologue of which has been shown to function in a tumour suppressor capacity in *drosophila melanogaster* (Yu *et al*, 2010; Genevet *et al*, 2010). Since hypermethylation is a common mechanism of disruption of the SWH pathway it is worth analysing this newly identified and previously unstudied in human cancer SWH pathway member. The work in this chapter will analyse the *KIBRA* methylation status across solid and haematological cancer cell lines and subsequent relevant primary tumour samples.

6.3 AIMS

1. To assess the methylation status of *KIBRA* in major solid tumour cancer cell lines and haematological cancer cell lines

2. To analyse the methylation status of *KIBRA* in relevant primary tumours of any of the cancer cell line types shown to be methylated.

3. To access any association between observed methylation and any additional genetic or clinical features, where additional information is available.

6.4 RESULTS

6.4.1 Analysis of *KIBRA* in cancer cell lines

KIBRA has a CpG island of 1165bp containing 113 CpGs situated upstream and inclusive of the transcription start point. A 248bp region was analysed by CoBRA (figure 6.1) in a panel of 7 leukaemia, 7 colorectal, 5 breast, 10 renal, 4 glioma, 5 prostate and 15 lung cancer cell lines. Methylation was infrequent in solid tumours with digestion only observed in 1/7 (14.3%) glioma cell lines (figure 6.2). No methylation was observed in non-cancerous brain samples.

More frequent methylation was found in leukaemia lines with 4/7 (57.1%) showing either complete (3/4) or partial (1/4) methylation (figure 6.3). In addition, no methylation was identified in either control bone marrow or peripheral blood samples (figure 6.3). Analysis of expression pre and post treatment with 5-azaDC in leukaemia cell lines showed re-expression in methylated lines (figure 6.3). These results show frequent methylation of *KIBRA* in leukaemia lines that appears to be capable of repressing expression at the RNA level.

6.4.2 Analysis of *KIBRA* in primary glioma tumour samples

Due to the infrequent levels of methylation in glioma cell lines only high grade primary glioma tumours were initially analysed. Of the 20 grade IV glioblastoma multiforme (GBM) tumour samples analysed none showed methylation (figure 6.4). Coupled with the cell line analysis, these results show that this region of the *KIBRA* CpG island is unmethylated in a wide range of common solid tumours.

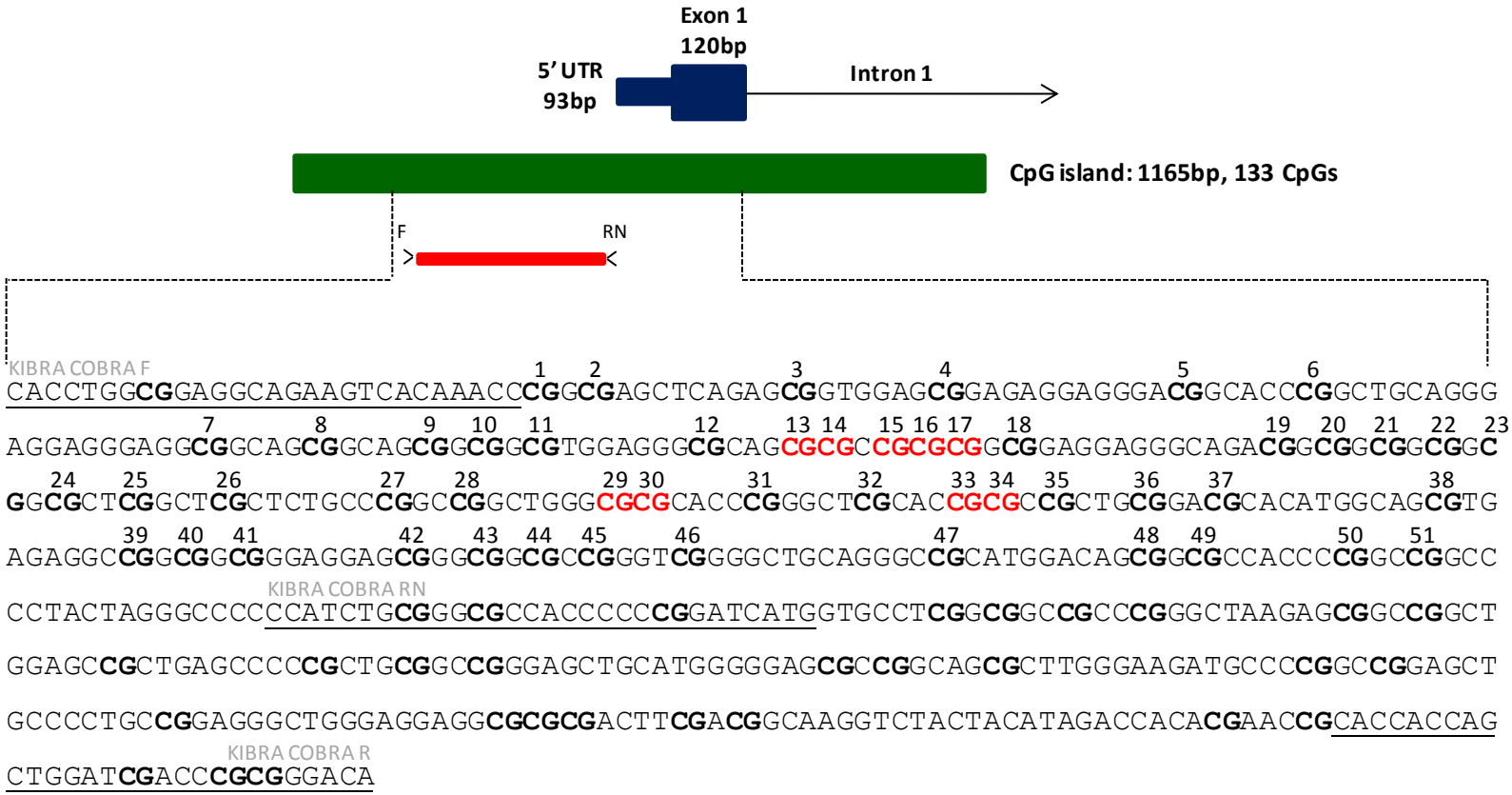


Figure 6.1_*KIBRA* CoBRA region

A schematic of *KIBRA* 5'UTR and exon 1 is shown (blue) with the CpG island underneath (green). The region amplified by CoBRA primers is indicated with dotted lines, the region of which is assessed by BstUI digestion is shown in red. The sequence amplified by CoBRA is displayed underneath with CpGs numbered 1-51 and BstUI sites shown in red text.

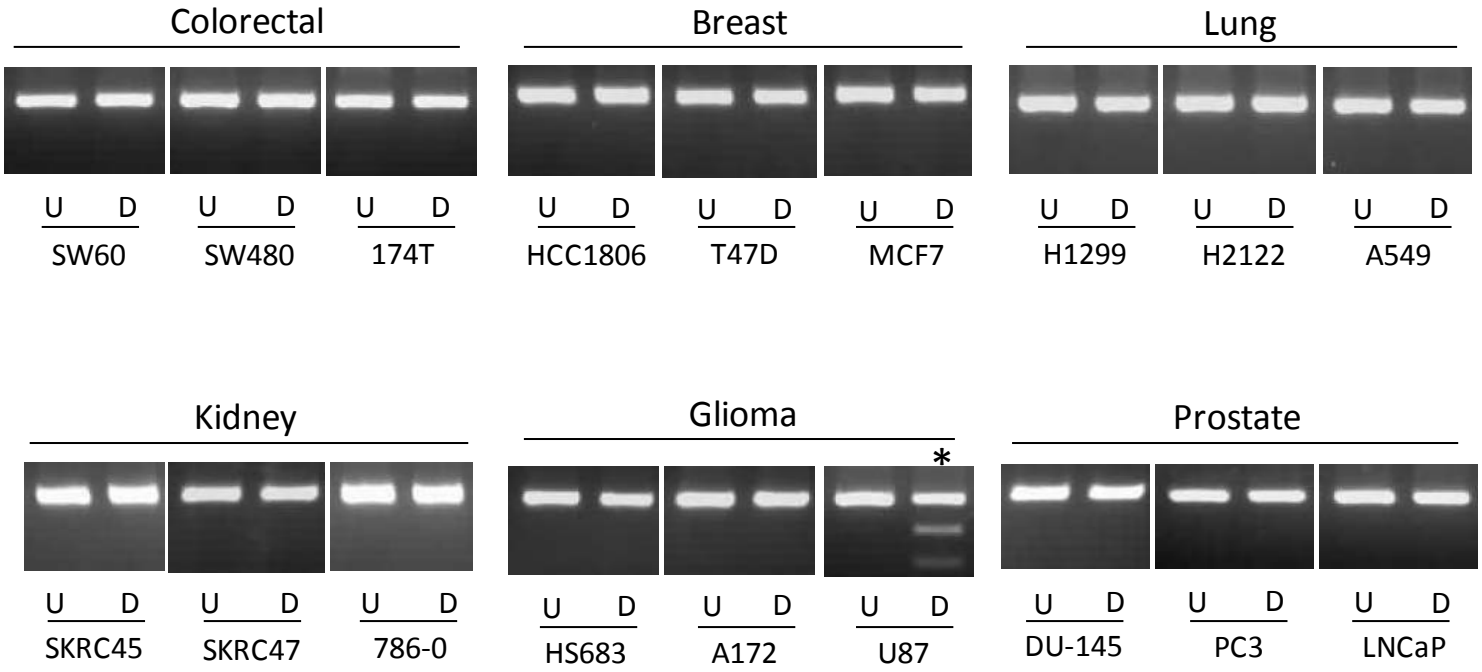


Figure 6.2_CoBRA results for solid tumour cancer cell lines.
CoBRA results are shown for three cell lines for each of the tumour types as indicated. All results are shown with undigested products (U) next to digested products (D). * indicates methylated samples.

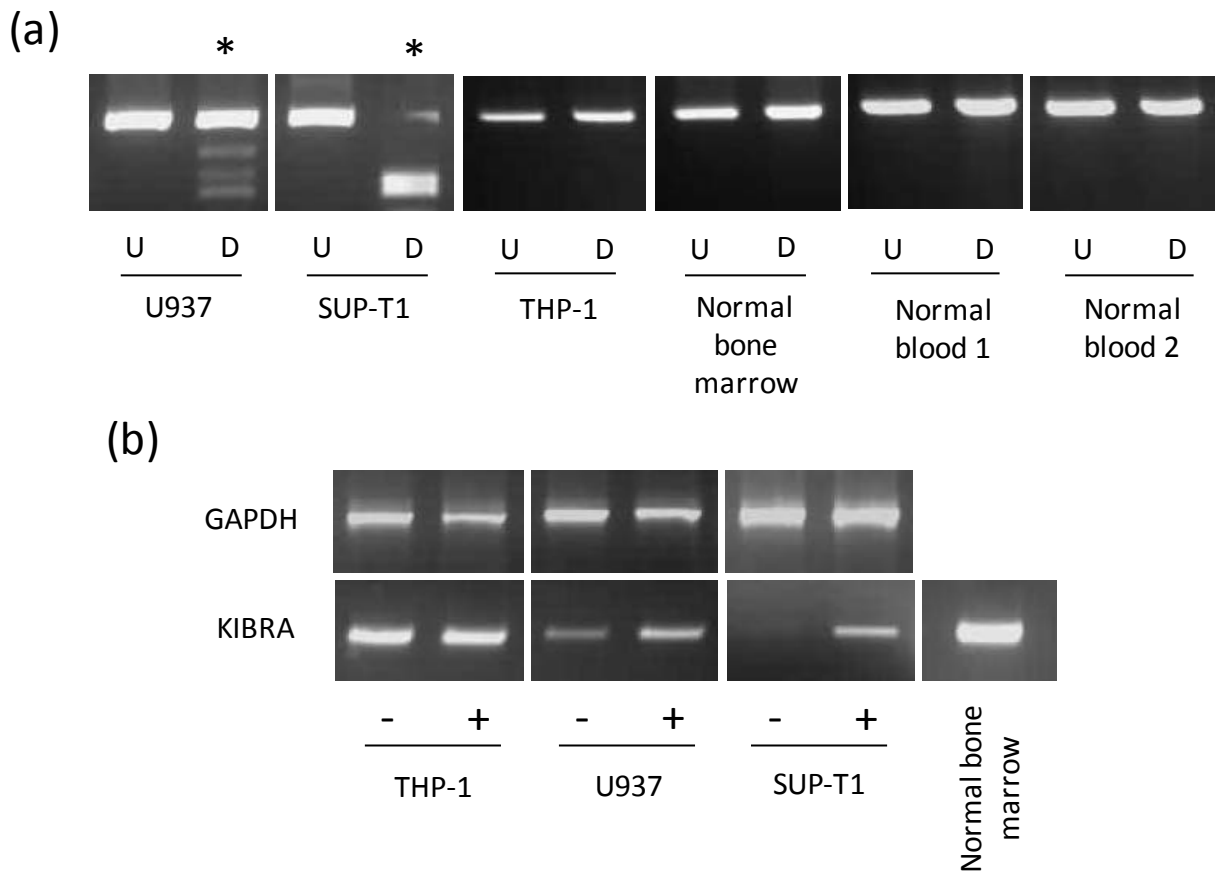


Figure 6.3_CoBRA and expression results for leukaemia cell lines, normal bone marrow and normal blood.
(a) CoBRA results are shown for three leukaemia cell lines, two normal blood samples and one normal bone marrow. Undigested PCR products (U) are run next to digested PCR products (D). (b) Expression results for *KIBRA* and *GAPDH* are shown pre (-) and post (+) treatment with 5-azaDC along with normal bone marrow for *KIBRA*. * indicates methylated samples.

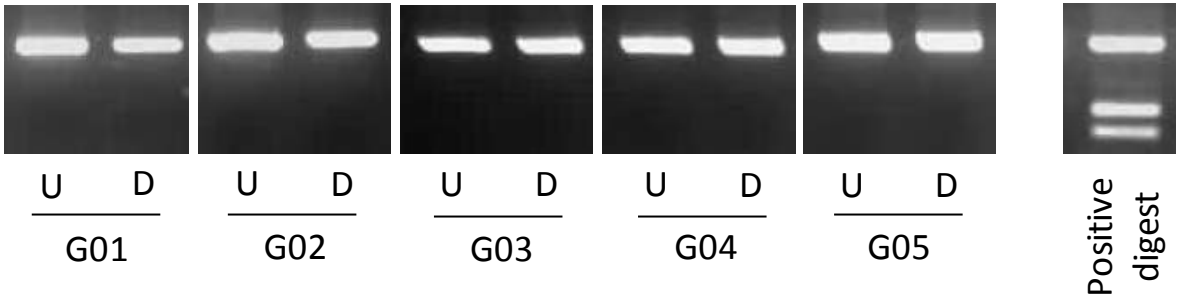


Figure 6.4_CoBRA results for primary glioma samples.
CoBRA results for five grade IV GBM samples are shown with undigested PCR products (U) run next to digested products (D). A positive digest is also shown to demonstrate effective BstUI digestion.

6.4.3 Analysis of *KIBRA* in primary leukaemia samples

A total of 70 childhood acute lymphocytic leukaemia (ALL) primary tumours consisting of 52 B-ALL, 1 pre-B-ALL, 12 T-ALL and 5 unclassified were assessed. A high frequency of methylation was observed overall (38/65, 58.5%) with the greatest level in B-ALL (33/38, 70%) samples compared to T-ALL (2/12, 16.7%) samples (p=0.0019) (figure 6.5). In order to confirm the CoBRA results and assess the pattern of methylation, clone sequencing of bisulphite modified DNA was performed. Results showed the region was highly methylated with little heterogeneity between the samples (figure 6.6). Neither normal blood nor bone marrow showed significant levels of methylation within the region (figure 6.6). B-ALL leukaemias commonly have a translocation between *ETV6/RUNX1* {t(12;21)(p13;q22)} and *KIBRA* methylation strongly associates with this translocation in this cohort of samples (p = 0.0082).

6.4.4 Methylation and expression analysis in diagnosis and remission paired primary B-ALL samples

Five pairs of B-ALL *ETV6/RUNX1* translocation positive samples from the diagnosis and remission stages of disease were analysed for methylation and expression. CoBRA analysis of B-ALL samples showed *KIBRA* methylation at diagnosis but not remission in all five pairs (figure 6.7a). In addition, RT-PCR showed expression of *KIBRA* mRNA at remission, when methylation was no longer present but not at diagnosis, when samples were all methylated (figure 6.7b). These results indicate that *KIBRA* methylation is specific for malignant cells and that methylation is capable of suppressing *KIBRA* mRNA expression in primary samples.

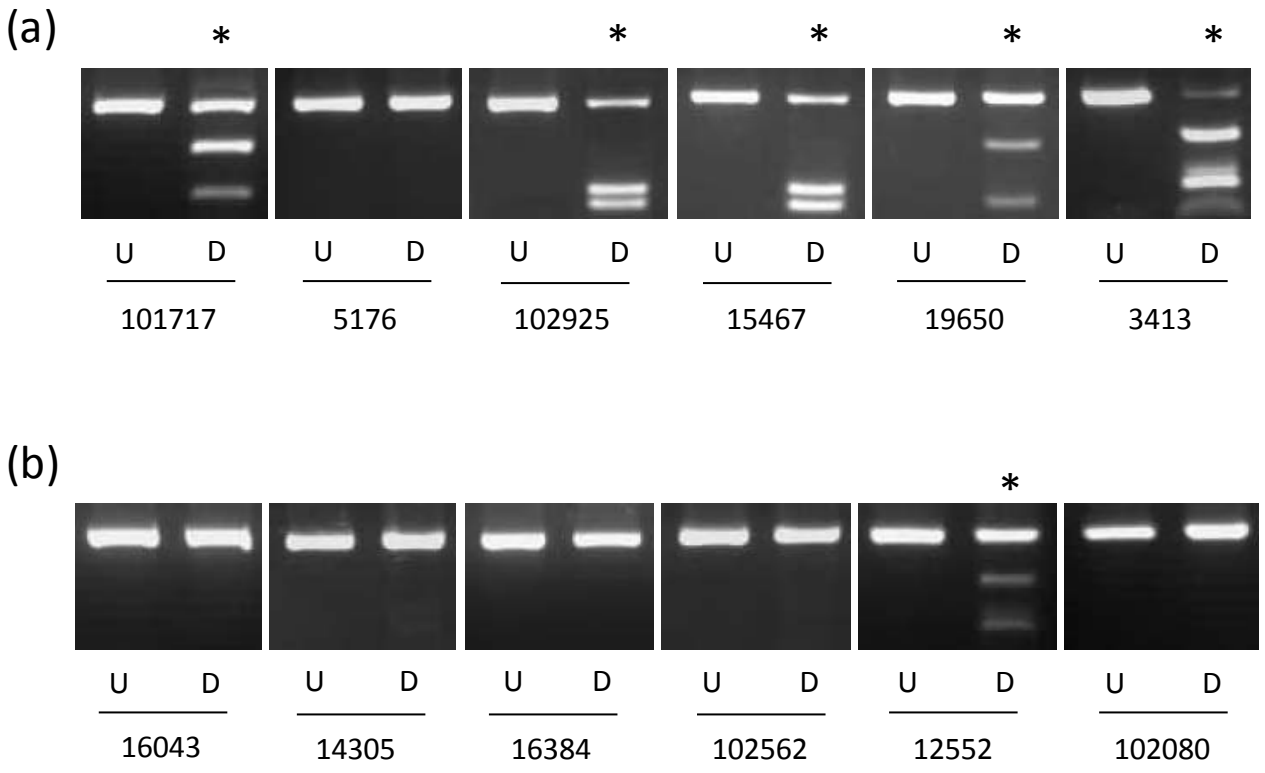


Figure 6.5_CoBRA results for primary ALL samples.
CoBRA results are shown for five primary B-ALL samples (a) and five primary T-ALL samples (b). Undigested PCR products (U) are run next to digested PCR products (D). * indicates methylated samples.

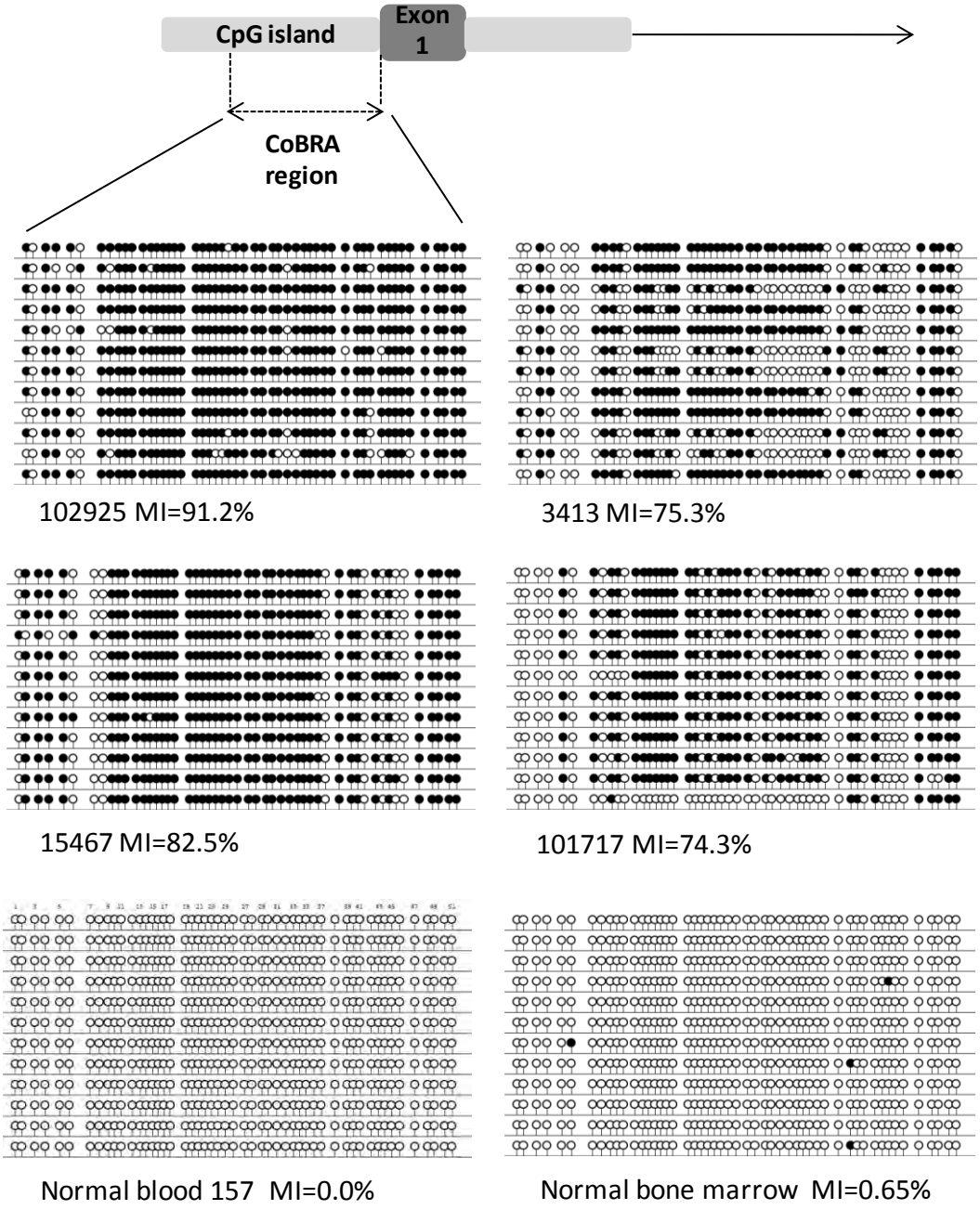


Figure 6.6_Sequencing of primary ALL samples

A schematic of the location of the *KIBRA* CpG island in relation to *KIBRA* exon 1 is shown along with the position of the region being analysed by CoBRA. Clone sequencing results are shown for four B-ALL samples (102935, 3413, 15467 and 101717), one normal blood and one normal bone marrow. Each line represents a single allele, and black and white circles represent methylated and unmethylated CpGs respectively. Methylaton indexes (MIs) were calculated as a percentage of methylated CpGs out of total CpGs analysed.

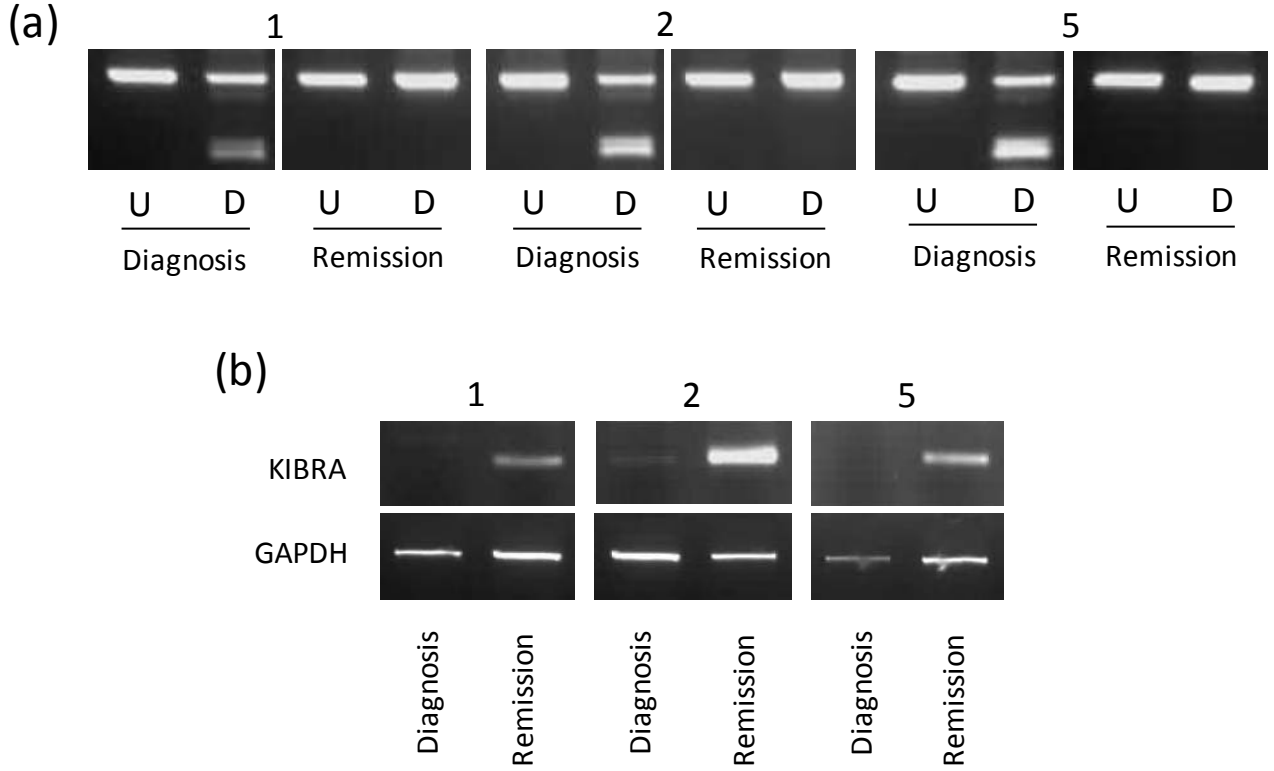


Figure 6.7_CoBRA and expression results for diagnosis and remission B-ALL samples. (a) CoBRA results are shown for three of the five diagnosis and remission paired B-ALL samples. Undigested PCR products (U) are shown next to digested PCR products (D). (b) *KIBRA* expression results are shown for diagnosis and remission samples along with *GAPDH*.

6.5 DISCUSSION

This study has identified frequent methylation of *KIBRA* in childhood acute lymphocytic leukaemia (ALL) (58.5%) that was more frequent in B-ALL than T-ALL samples (70% compared to 16.7% respectively; p=0.0019). Methylation of *KIBRA* was also significantly more prevalent in samples harbouring the *ETV6/RUNX1* translocation (p=0.0082).

6.5.1 *KIBRA* methylation in childhood ALL

In the western world, childhood ALL is the most common form of childhood leukaemia. In the UK, 31% of childhood cancers are leukaemias, of which 79% are of the ALL subtype (CRUK cancer statistics). Despite the disease having high survival rates of up to 80%, childhood ALL still accounts for 30% of all childhood cancer deaths in the UK (CRUK cancer statistics). The highest rates of survival are observed in children aged 1-9 years whilst infants (<1 year) have very high rates of treatment failure (Möricke *et al*, 2005; Pui *et al*, 2008).

ALL is split into two major types depending on the cell type of origin; B-cell ALL (B-ALL) and T-cell ALL (T-ALL). B-ALL cells express B-cell associated antigens such as CD79a, CD19 and HLA-DR whilst T-ALL cells express T-cell associated antigens such as CD3, CD7, CD2 and CD5. The vast majority of childhood ALL cases are B-ALL (80-85%) with the remainder being mostly T-ALL although some cancers are of an undetermined lineage, demonstrating features of both B- and T-cells. Chromosomal aberrations are a common feature of leukaemias with both changes in ploidy and genomic translocations frequent events. ALL patients demonstrating high hyperploidy (51-60 chromosomes per cell) have a favourable outcome, this is common in B-ALL (20-25% of B-ALL patients), whilst patients demonstrating hypodiploidy (less than 40 chromosomes per cell) have a poor prognosis

(Paulsson *et al*, 2009; Nachmann *et al*, 2007). Common translocation events in ALL include the t(12;21)(p13;q22) translocation producing the *ETV6*/*RUNX1* fusion protein, the t(9;22)(q34;q11) translocation (Philadelphia chromosome) producing the *BCR*/*ABL* fusion protein, the t(1;19)(q23;p13.3) translocation producing the *TCF3/PBX1* fusion protein and translocations involving the MLL (*myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)*) gene at 11q23. *BCR*/*ABL*, *TCF3*/*PBX1* and *MLL* translocations occur in 3%, 5% and 5% of childhood ALL cases respectively (Crist *et al*, 1990; Pui *et al*, 2003; Uckun *et al*, 1998). The most frequent translocation in childhood ALL is the *ETV6/RUNX1* translocation [t(12;21)(p13;q22)] (previously known as *TEL/AML1*). The *ETV6*/*RUNX1* translocation is a frequent and early event, occurring in 25% of all B-ALL cases, the translocation is associated with good prognosis and whilst it often occurs prenatally, it is not sufficient to cause childhood ALL alone (Greaves *et al*, 2003; Shurtleff *et al*, 1995; Armstrong and Look, 2005). In its wild type form, *RUNX1* acts as a transcriptional activator of numerous genes including those, such as *IL3* (*interleukin 3*) (Uchida *et al*, 1997), that are involved in haematopoietic development. Transcriptional activation of genes by *RUNX1* is brought about by chromatin modifications, for example, the involvement of *RUNX1* in myeloid differentiation is associated with its interaction with the histone acetyltransferase p300 (Kitabayashi *et al*, 1998). When fused with a portion of *ETV6*, the transcriptional repression domain within wild type *ETV6* causes the normally *RUNX1* activated genes to become transcriptionally repressed (Hiebert *et al*, 1996; Fears *et al*, 1997). *ETV6* mediated repression of genes by the *ETV6*/*RUNX1* fusion protein is achieved through nuclear receptor corepressor (N-CoR) HDAC complexes (Guidez *et al*, 2000).

Hypermethylation is also a common event in haematological cancers. In childhood ALL, numerous genes have been shown to be highly methylated, including, among many others,

KLK10 (*kallikrein related peptidase 10*), *SFRP1* (*secreted frizzled-related protein 1*), *ADAMTS1* (*ADAM metallopeptidase with thrombospondin type 1 motif, 1*) and *CDKN1A* (*cyclin-dependent kinase inhibitor 1A (p21, Cip1*), all methylated in between 40% and 60% of cases (Roman-Gomez *et al*, 2005; Roman-Gomez *et al*, 2002). There have been numerous genome-wide methylation studies identifying further genes hypermethylated in ALL including a family of ephrin receptor tyrosine kinases and genes showing significant differences in methylation status between B-ALL and T-ALL (Dunwell *et al*, 2010; Kuang *et al*, 2008). Genome wide studies have also identified methylation profiles that predict relapse and discriminate between cytogenetic ALL subtypes (Milani *et al*, 2010) and higher levels of methylation in lymphoid lineage leukaemias rather than myeloid malignancies (Martín-Subero *et al*, 2009b). The work in this chapter has identified frequent methylation in the *KIBRA* gene in childhood ALL, which is the first example of *KIBRA* methylation in human cancer. Significantly higher levels of methylation were observed in B-ALL cancers compared to T-ALL cancers (p=0.0019) suggesting a possible role in B-cell leukaemogenesis. However, it is important to note that the controls used for this study consisted of normal blood DNA and normal bone marrow DNA, neither of which are specific for either B or T lineage cells and therefore will not allow for comparison of methylation status between B-cell ALL and a pure sample of normal B-cell DNA, and likewise for T-ALL and T-cell DNA. Despite this, as gene specific hypomethylation is a relatively uncommon event in cancers it is unlikely that the subtype showing the higher frequency of *KIBRA* methylation, B-ALL, at 70.6%, is normally methylated in B-cells. In addition, the resulting increase in expression following methylation removal in cell lines and the methylation and corresponding loss of expression of *KIBRA* in diagnosis but not remission samples suggests *KIBRA* methylation events are likely to have some biological function on the leukaemic cells. *KIBRA* methylation also associated with the

ETV6/RUNX1 translocation event and it has been suggested that *ETV6/RUNX1* leukaemias are more likely to have methylation of multiple genes than those without the translocation (Roman-Gomez *et al*, 2006). *KIBRA* methylation may therefore have a role in *ETV6/RUNX1* leukaemogenesis.

6.5.2 *KIBRA* methylation as a deregulated member of the SWH pathway in haematological cancers

Deregulation of the SWH pathway has been observed in many cancer types including breast, melanomas, neurofibromas, soft tissue sarcomas, and astrocytomas via numerous members of the pathway (Rouleau *et al*, 1993; Lai *et al*, 2005; Seidel *et al*, 2007; Jiang *et al*, 2006; Takahashi *et al*, 2005). This is the first report describing epigenetic inactivation of *KIBRA*, a newly identified upstream member of the SWH pathway in any human cancer. As no methylation was discovered in the panel of solid tumours looked at, inactivation of *KIBRA* by hypermethylation appears to be restricted to haematological cancers, although this cannot be said for certain as there are many types of cancer that have not been studied in this chapter. In particular, it would be interesting to look at neurofibromas as mutations in one of the binding partners of *KIBRA*, *NF2*, is involved in neurofibromatosis type 2 pathogenesis (Rouleau *et al*, 1993).

It has recently been shown that expression of the SWH pathway members *MOBKL2A*, *MOBKL2B* and *LATS1* is downregulated and associates with lower survival in another haematological cancer, mantle cell lymphoma (Hartmann *et al*, 2010). A separate study has shown methylation of *LATS1* in childhood and adulthood ALL (Roman-Gomez *et al*, 2004) suggesting the downregulation observed by Hartmann *et al*, 2010 could be due to methylation. In previous studies, methylation of *large tumour suppressor 2* (*LATS2*) has also

been shown to be methylated in ALL with associated lower expression in primary leukaemia samples. Both childhood and adulthood cases were affected and in both cases lower expression associated with higher relapse and mortality. Unlike *KIBRA*, expression did not return in remission samples (Jiménez-Velasco *et al*, 2005). In addition, previous work form this laboratory identified methylation of *FAT1* and *YAP* in primary childhood ALL samples (Dunwell, 2010). Taken together these results suggest deregulation of the SWH pathway may be a common event in haematological cancers. This may offer insight into the development of haematological cancers.

6.5.3 *KIBRA* as an upstream member of the SWH pathway

KIBRA was recently identified as forming part of the Kibra/Expanded/Merlin complex of the SWH pathway in flies (human homologues KIBRA, FRMD6 and NF2 respectively) (Yu *et al*, 2010; Genevet *et al*, 2010; Baumgartner *et al*, 2010). The complex is thought to be predominantly located at the sub-apical region of the cell and upon binding with both Merlin and Expanded thought to recruit core SWH members to the apical membrane where they become activated. At present, it is unknown how members of the Kibra/Expanded/Merlin complex are regulated. Given the methylation of *KIBRA* in ALL but not common solid tumour types, it would be interesting to see if *KIBRA* behaved differently in haematological cells compared to other cell types.

6.5.4 Conclusions and further work

This study is the first description of *KIBRA* methylation in any human cancer. *KIBRA* methylation was a frequent event in childhood ALL but not in common solid tumours, suggesting haematological specific hypermethylation. Although the aims of this chapter have been met, it would be beneficial to show *KIBRA* expression reduced at the protein as well as

the RNA level and that loss of *KIBRA* results in tumourigenic characteristics in an ALL model. It would also be interesting to see if loss of *KIBRA* resulted in the same phenotypes in cells with and without the *ETV6/RUNX1* translocation, to determine if *KIBRA* loss has a role in t(12;21) leukaemogensis.

This work has now been published under Hill *et al* (2011) *Epigenetics* **6**(3):326-332. A copy of this publication is provided at the back of this thesis.

CHAPTER 7

DISCUSSION

Cancer represents the largest cause of death worldwide, accounting for 7.5 million deaths in 2008 (Jemal *et al*¸ 2011). In the UK, 309 500 people were diagnosed with cancer in 2008, breast and prostate cancer diagnoses being the most common in adult females and males, representing 31% and 24% of all adult female and male cancers respectively (CRUK cancer statistics). Although most commonly presenting in adulthood, cancer can also be a disease of childhood, the most common forms in children in the UK being leukaemias and brain and CNS cancers (CRUK cancer statistics). As a result of lifestyle changes and an aging population, the incidence of many cancers across the developed world has been, and continues to, increase. However, in the developed world, the numbers of deaths from cancer have been decreasing, primarily as a result of better disease understanding. The development of better screening methods for cancers, such as breast cancer through mammography and cervical cancer through cervical screening programs, have helped to both identify cancers earlier and identify patients most at risk who can be monitored more closely. Additionally in the case of cervical cancer, to identify pre-malignant lesions that can be treated before progression to malignancy. The identification of specific genetic alterations in cancers have also led to better disease outcomes through targeted therapies such as the treatment of *ERBB2* overexpressing breast cancers with trastuzumab (Herceptin) (Hudis *et al*¸ 2007) and the recent clinical trial of Vemurafenib in melanoma patients with a specific *BRAF* mutation (V600E) (Chapman *et al*, 2011). Despite the better outcome of a cancer diagnosis, continued identification of good

therapeutic targets and prognostic, diagnostic and treatment outcome markers are essential for benefitting patient outcome.

Changes in DNA methylation are a hallmark of cancer genomes and a common mechanism of tumour suppressor gene inactivation. DNA methylation marks are also amenable to biomarker development, for example, the methylation status of *MGMT* is widely used to predict which glioma patients are most likely to benefit from alkylating dugs such as temolozide. Assessment of DNA methylation changes in cancer genomes can therefore highlight genes that may elucidate deregulated genes and pathways, and potential candidates for biomarker development. To aid in this field of research, the work in this thesis aimed to identify DNA methylation changes in sporadic cancers, presenting two studies aiming to identify genes hypermethylated in sporadic breast cancers and two studies looking at the methylation status of two candidate genes across multiple cancer types.

7.1 Genome wide analysis of sporadic breast cancer

Recent technological advances have developed multiple assays for genome wide analysis of DNA methylation. Two of the most recently developed methods are methylated DNA enrichment approaches and Illumina methylation array platforms. Chapters three and four have used two of the most up to date and comprehensive methods available at the time of study to assess methylation in sporadic breast cancer.

Chapter three used the Methylated CpG Island Recovery Assay (MIRA) in conjunction with CpG island microarrays across five breast cancer cell lines. Sixteen genes were identified that showed frequent methylation in cell lines but not normal breast tissue DNA, of which nine were also frequently methylated in primary tumour samples. Of these nine, five genes demonstrated frequent, tumour specific methylation in a second cohort of sporadic ductal

breast tumour patient samples and re-expression following treatment with the de-methylating drug, 5-azaDC, in methylated cell lines. All five genes, *CIDE-A*, *DBC1*, *EMILIN2*, *FBLN2* and *SALL1*, are functionally good candidate genes to be involved in breast cancer tumourigenesis. Methylation of three of the five genes also showed an association with clinical features; *EMILIN2* methylation associated with ER and PR positive tumours, relapse and worse disease free survival (DFS), whilst *CIDE-A* and *FBLN2* methylation associated with ER or PR positive tumours respectively. In addition, decreased expression was observed in *EMILIN2* methylated patient samples, suggesting a likely functional consequence of *EMILIN2* hypermethylation. Hypermethylation in cancer tissue was novel for *EMILIN2* and *CIDE-A* whilst methylation of the remaining genes, *DBC1*, *FBLN2* and *SALL1* had previously been observed in other cancers (Izumi *et al*, 2005; Dunwell *et al*, 2009; Kuang *et al*, 2008; Tong *et al*, 2010). Analysis of these genes in other solid tumours demonstrated frequent methylation for three of the genes, *SALL1*, *FBLN2* and *EMILIN2*, in prostate cancers (*SALL1* and *FBLN2*), colorectal cancers (*EMILIN2*) and gliomas (*FBLN2*). This study represents the first example of using the methylated CpG island recovery assay (MIRA) in invasive breast cancer and one of the first examples of using methylated DNA enrichment based assays on invasive breast cancer. The study has shown the technique to be successful in identifying interesting methylated genes.

Chapter four used the Illumina Infinium HumanMethylation27 array on a cohort of 39 sporadic ductal breast carcinoma samples. The study demonstrated that global hypermethylation levels in a cohort of sporadic ductal breast tumour samples is significantly higher in relapsing tumours and those positive for ER and PR receptors. The association between high methylation levels and ER/PR status has also been demonstrated in two other studies that were published in the same year as this (Li *et al*, 2010; Holm *et al*, 2010). These

results may indicate a CIMP-like phenotype in this subset of tumours and suggest that ER/PR positive tumours are more susceptible to hypermethylation events, making them the most suitable breast cancer type to be treated with demethylating drugs such as 'Vidaza® and 'Decitibine/Dacogen®' if these treatments ever come to fruition in solid tumour therapy regimes. Whilst the methylation status of individual genes has previously been shown to correlate with relapse in breast cancer, I believe this is the first time a global increase in methylation has been shown to associate with relapse in breast cancer, although no association was observed with disease free survival.

The study was also able to identify individual genes where methylation status correlated with clinical features. Six genes, *RECK*, *SFRP2*, *ITR*, *UGT3A1*, *ACADL* and *UAP1L1*, were identified that showed significant association between methylation and relapse/DFS. Of these six, *RECK*, *SFRP2* and *ACADL* were further characterised, showing tumour specific/enhanced methylation and reduced expression in primary tumour samples (*RECK* and *ACADL* only), suggesting these genes may also have functional consequences as well as being potential candidates with prognostic value. These results will be of particular interest to those in the clinical breast cancer research field. Eighteen genes (*B3GAT1*, *TNFRSF10D*, *GRIA4*, *C1orf114*, *DNAJC6*, *C6orf145*¸ *CORIN*, *LAMA2*, *ITR*, *MCAM*, *C12orf34*, *NPY*, *COL1A2*, *RSNL2*, *ZNF660*, *ADAMTSL1*, *DPP4*, *MAT1A*) were identified as being significantly more frequently methylated in ERPR expressing tumours. Further work on *TNFRSF10D*, *COL1A2* and *C1orf114* demonstrated methylation is capable of reducing expression in cell lines, tumour specific/enhanced methylation for all three genes and significant loss in expression in methylated patient samples for *C1orf114*. The findings for *C1orf114* are particularly interesting as it was also identified in the DCIS MIRA study (Tommasi *et al*, 2009) and therefore represents an uncharacterised gene with strong evidence towards a potential role in

breast cancer, making it an exciting candidate for further work. Analysis of candidate genes irrespective of any possible association with clinical features was also successful in identifying interesting genes; further analysis of the most highly methylated and genes with multiple probes identified four genes, *COX7A1*, *SST*, *CDKL2*, *ZNF154*, that showed correlation between expression and cell line methylation and subsequent specific/enhanced methylation in the tumour normal paired samples. These results have helped to identify further genes involved in breast cancer and given some insight into the global methylation changes of breast cancer.

7.1.1 Experimental Design for Breast Cancer Array Analysis Studies

The research carried out in chapters 3 and 4 utilised two of the most recently developed techniques, an affinity pull down approach (chapter 3) and an array based on direct hybridisation of bisulphite modified DNA (chapter 4). The fundamental design of these two studies exhibited substantial differences.

The MIRA assay was carried out on a small number of breast cancer cell lines whilst the Infinium array was performed on a larger cohort of sporadic breast tumour patient samples. It is the differing array methodologies that ultimately led to the use of different sample types. The MIRA technique requires a large amount of starting material, 5µg, compared to the 0.5µg required for Infinium. Access to 5µg of patient material was simply not possible at the time of study and since cell lines offer the possibility of treatment with 5-azaDC to compare expression in relation to DNA methylation status they represent a useful starting material for this type of study. However, it is known that DNA hypermethylation is both increased in the number of genes and more expansive in terms of the level of methylation within a particular region in cell lines compared to patient tumour samples (Smiraglia *et al*, 2001), meaning

many genes may be identified as frequently methylated in cell lines that are not frequently methylated in primary tumours. Taking this into account, frequent methylation in cell lines was classed as >30% compared to >25% in primary tumours, however, despite this, a frequency of just under half (44%) of genes frequently methylated in breast cancer cell lines were not methylated in primary tumour samples, highlighting the major disadvantage to using cell line DNA as starting material for this study. In comparison, the Infinium array gave an initial quantitative measurement of methylation across a panel of 39 sporadic samples. Due to the number of samples used and the availability of clinical information for these samples, it was possible to use the data directly from the array to identify genes that are associated with particular clinical features. Therefore the use of primary samples conferred an advantage to finding clinically relevant methylation events, offering an advantage over using cell lines in this respect.

When comparing methylation with expression, although only a small number of samples were analysed, the efficiencies of the two arrays appeared to be comparable; 88.9% (8/9) of genes analysed for expression in the MIRA study correlated with methylation whilst 83.3% (10/12) of genes analysed in the Infinium study correlated with methylation, suggesting the use of either platform is successful in identifying methylation events that are capable of affecting gene expression.

The genes analysed by CoBRA in tumour/normal paired samples showed a greater level of specificity if identified through the MIRA study compared to the Infinium study. Whilst both assays showed tumour specific/enhanced methylation in many genes, genes identified through the MIRA assay showed less methylation in the associated normals compared to those identified in via the Infinium study, where many genes showed reduced methylation in adjacent normals compared to methylated tumours rather than having a complete

unmethylated status. This is most likely due to the MIRA assay design encompassing more probes per CpG island than the Infinium array and therefore giving results representative of a larger region.

Use of affinity pull-down methods to isolate methylated DNA is a newly developed technique for assessing global methylation levels. The technique offers the advantage of not requiring bisulphite modification of DNA which may prove particularly useful for modification to a technique able to distinguish between 5-hydroxymethylcytosine (5hmC) and 5-methylcytosine (5mC) in the future. However, there may be in-built biases to the regions of DNA that are most successfully isolated due to protein/antibody binding specificities. A recent study comparing MeDIP and methyl binding domain (MDB) protein affinity enrichment techniques suggested that MeDIP and MBD based techniques do indeed differ between preferred enrichment sites (Nair *et al*¸ 2011). MBD based techniques appear to preferentially pull-down dense CpG regions which may explain the high levels of Homeobox related genes in the MIRA studies (Rauch *et al*, 2006; Tommasi *et al*¸ 2009). Whilst this in no way suggests that the detection of tumour acquired methylation of these genes is not true, it may suggest that the enrichment of these genes within the lists could be artefacts of the technique. The opposite is true for MeDIP based approaches, with preferential enrichment of less dense CpG regions. Additionally, analysis of the MIRA results in chapter three in the high stringency gene list selected genes based on a very high number of probes being methylated which will bias the list towards larger CpG islands, highlighting the potential impact of using different methods of analysis. Nonetheless, the technique allows analysis of larger regions of DNA than the Infinium assays and is not limited by the assessment of specific CpGs by selective probe design.

The Infinium HumanMethylation27 array is designed to generate a quantitative measurement of a single CpG dinucleotide, however, the unmethylated and methylated bead types are designed like USP and MSP primers respectively, making binding of DNA to either bead type dependent on the methylation status of upstream CpGs. The β value therefore gives a quantitative measurement of the CpG when additional CpGs in the immediate vicinity are also either methylated or unmethylated, which, although a minor point, does confer an advantage to the assay if looking for wider methylation levels. As already mentioned, the technique suffers from the limitations imposed by specific probe design as opposed to pull-down of any methylated DNA region. Whilst the MeDIP and MBD based techniques are often hybridised to microarrays, they lend themselves to sequencing based methods of analysis of isolated methylated fragments. This, when sequencing methods are more cost effective, will allow for the next step of methylome analysis.

In summary, the method of use to assess global methylation levels depends greatly on the aims of the study. If the identification of clinically relevant genes is the main aim of a project then the Infinium assay is best suited due to the quantitative nature and amenability to the development of a diagnostic assay. If novel candidate genes are required then one of the methylated DNA enrichment assays are perhaps more suited due to the increased number of CpG islands in the array covering more features such as uncharacterised genes and miRNAs, however, a larger number of samples and the use of primary patient material would be more appropriate. The newly available Illumina Infinium HumanMethylation 450k array, covering a much wider selection of CpGs within gene related CpG islands, CpGs within CpG islands unrelated to genes, CpGs outside of CpG islands and CpGs related to other features such as miRNAs compared to the 27K array, will bridge the gap between higher coverage CpG island microarrays that are used with methylated DNA enrichment approaches and the lower

coverage direct bisulphite modified approaches. Although the use of different methodologies is useful due to a likely increase in the chances of identifying novel findings as any inherent biases in each method will be in part counteracted by using a different method, for the purposes of clinically relevant research, it may be advantageous for similar, more comparable platforms to be used by different groups. This will allow results to be pooled together to highlight the most promising candidates for further analysis as potential biomarkers. Having said this, for true comparisons to be made between studies, the method of data analysis as well as the array methodology needs to be more comparable and a more conformed approach to data analysis would perhaps benefit this field at present. Large cancer genome research consortiums, such as the TCGA (The Cancer Genome Research Atlas), are currently carrying out methylation studies on a large number of samples for those cancers with the greatest overall health impact, either because of uniformally bad prognosis, such as gliomas, or because of the sheer number of individuals affected by the disease, such as breast cancers. The TCGA are currently performing these analyses using the Infinium HumanMethylation27 array. Data is now publically available for a number of these studies although publications of actual findings have not yet been produced in most cases. It will be interesting to see if results from smaller, similar studies are replicated in these large cohorts. Those that demonstrate replication in these large studies will be those candidates showing the greatest promise as potential biomarkers. However, these large studies may make it more desirable to smaller groups to obtain data using either different methodologies or using a different, more unusual cohort of samples.

7.2 Candidate genes analysis of *RASSF10* and *KIBRA* in multiple cancer types

Work in this thesis also analysed the methylation status of two candidate genes, *RASSF10* and *KIBRA*, both of which produced positive results.

RASSF10 hypermethylation was demonstrated in multiple common solid tumours, adding weight to it being a candidate tumour suppressor gene. This is exciting for researchers in the field of RASSF family members, particularly as it is a member of the N-terminal RASSFs, which, apart from *RASSF8*, show little evidence to behave as tumour suppressor genes like many of the classical RASSFs. As a newly characterised gene, it will also be interesting to see results from functional studies in addition to those described in Hill *et al*, (2011). The methylation work of *RASSF10* focused on the highest methylated tumour type, gliomas, which showed a high level of methylation in grade III and IV tumours. A correlation was also observed between *RASSF10* methylation and overall survival times in secondary grade IV gliomas. Methylation was also observed in the early lesions of secondary gliomas, suggesting methylation of *RASSF10* is likely to be an early event. These results will be of interest to those in the basic science field of gliomagenesis and those investigating possible markers for clinical prognosis.

Analysis of *KIBRA* methylation status across multiple cancer types identified frequent methylation in childhood ALL samples that correlated with the most common translocation within this type of leukaemia. This, as the first demonstration of hypermethylation of this newly identified member of the SWH pathway, will be of great interest to those working in the SWH pathway field. Not only does it indicate a possible tumour suppressor role for *KIBRA* in mammalian systems, the haematological cancer specific nature of hypermethylation may suggest *KIBRA* has differing functions in different cell types. The cancer specific nature of methylation and the association of *KIBRA* methylation with the *ETV6*/*RUNX1* translocation will also be of interest to those working within the field of childhood ALL.

The work here shows that a candidate gene approach for the identification of novel hypermethylated genes is successful when good candidates are chosen, for example, as in this

case when genes are known to be part of a particularly important pathway or family. This approach is a much cheaper way of identifying novel methylated genes and, due to the nature of the approach, when positive, will undoubtedly identify functionally relevant genes. The recent developments of genome wide techniques should therefore not stop small candidate genes studies that can produce just as interesting and relevant results.

7.3 Final Conclusions

Through the use of genome wide studies on breast cancer cell lines and clinically characterised sporadic breast tumour patient samples and candidate gene approaches, the research presented in this thesis has identified numerous genes hypermethylated in various cancers. Some of these genes, such as *EMILIN2* and *C1orf114* represent excellent candidates for further analysis in breast cancer. Other genes, such as *RECK* and *ACADL*, represent genes that warrant further analysis on a multi-variant level and in a much larger cohort to ascertain whether these targets would be useful as prognostic tests. Results from the candidate gene analyses have identified *RASSF10* and *KIBRA* as new genes involved in gliomas and childhood ALL respectively which is exciting for those in the respective cancer fields but also the *RASSF* family and SWH pathway fields.

	CELL LINE	ORGAN	TISSUE	TUMOUR STAGE	DISEASE	AGE OF PATIENTS AT COLLECTION	SEX	OTHER INFORMATION
BREAST	HCC1806	mammary gland; breast	not avaiable	TNM stage IIB, grade 2	primary acantholytic squamous cell carcinoma	60	female	-
	HCC1395	mammary gland; breast	duct	TNM stage I, grade 3	primary ductal carcinoma	43	female	-
	HCC1419	mammary gland; breast	not available	TNM stage I, grade 3	primary ductal carcinoma	42	female	-
	HCC1143	mammary gland; breast	duct	TNM stage IIA, grade 3	primary ductal carcinoma	52	female	-
	HCC1937	mammary gland; breast	duct	TNM stage IIB, grade 3	primary ductal carcinoma	23	female	-
	MCF-7	mammary gland; breast	not available	not available	adenocarcinoma	69	female	derived from metastatic site: pleural effusion
	T-47D	mammary gland; breast	duct	not available	ductal carcinoma	54	female	derived from metastatic site: pleural effusion
	HTB19	mammary gland; breast	not available	not available	carcinoma	74	female	-
	MDA-MB-231	mammary gland; breast	not available	not available	adenocarcinoma	51	female	derived from metastatic site: pleural effusion

	CELL LINE	ORGAN	TISSUE	TUMOUR STAGE	DISEASE	AGE OF PATIENTS AT COLLECTION	SEX	OTHER INFORMATION
LEUKAEMIA	JURKAT	not available	not available	not available	acute T cell leukemia	not available	male	-
	NALM6	peripheral blood	peripheral blood	not available	acute lymphoblastic leukemia	19	not available	-
	U937	pleural effusion	pleural effusion	not available	histiocytic lymphoma	37	male	-
	SUP-T1	pleural effusion	pleural effusion	not available	lymphoblastic leukemia	8	male	-
	CEM	peripheral blood	peripheral blood	not available	acute lymphoblastic leukemia	4	female	-
	DND41	peripheral blood	peripheral blood	not available	T-acute lymphoblastic leukaemia	13	male	-
	THP-1	peripheral blood	peripheral blood	not available	acute monocytic leukemia	1	male	-

	CELL LINE	ORGAN	TISSUE	TUMOUR STAGE	DISEASE	AGE OF PATIENTS AT COLLECTION	SEX	OTHER INFORMATION
GLIOMA	Hs683	brain	brain	not available	glioma	76	male	-
	U343	brain	brain	not available	glioblastoma	not available	male	-
	H4	brain	brain	not available	neuroglioma	37	male	-
	U87	brain	brain	classified as grade IV as of 2007	glioblastoma; astrocytoma	44	female	-
	A172	brain	brain	not available	glioblastoma	53	male	-

Appendix A. Cell line details for major cancer cell lines assessed in this study.
Cell line details are given for the major cancers assessed during this study (breast, leukaemia and glioma). In each case available data is given with regards to organ, tissue, tumour stage, disease, age of patients at collection and sex.

GENE	PRIMER	PRIMER SEQUENCE (5'-3')	ANNEALING TEMPERATURE (°C)		PROGRAM	PRODUCT SIZE (bp)	
			PRIMARY REACTION	SECONDARY REACTION		PRIMARY REACTION	SECONDARY REACTION
ABT1	F	GAGATTTGATYGGGAATTTGAGAATAATT	54	54	STANDARD	473	416
	RN	ACRAAATATAACCCAAATACACAATACCTAA					
	R	TCCRACCTCRCCATAAACRCTAAAAAAATT					
BARHL2	F	TATTYGTGATTGTGGATATYGGGATTAT	53	53	STANDARD	625	494
	RN	CCCRTAACCRAATTTCTAAACTTTCTTAA					
	R	AACCTAACACCRAACRACAACRATAA					
BUB3	F	GGTAGAGYGTTAGTTTTTAGAAAATTATATTTTTT	54	54	TOUCHDOWN	480	348
	RN	TTCCCTCAAAACACTCRCAAACTAAA					
	R	CAATTACRAACCCRAAAAATCCCTAA					
CCNJ	F	GGTTTAATTGGATTGGTTATTGTAGTG	54	54	STANDARD	425	311
	FN	TAATYGAAGYGTTAGGTTAGTTTAGGT					
	R	CCCCRCTCATACTACRAACAATAA					
CD44	F	GGAGGTATTGYGTTATTTAGGGTAAGATT	56	56	STANDARD	521	435
	RN	ATATCAAAAAAAACAAACCCACACAAAAARCAARCA					
	R	AATTAAAARCTCCATCAAAATAACAAATCRCCTA					
CDH24	F	GGTTTTTTYGTAGTGGGTTTYGAATTTTTTATT	55	55	STANDARD	438	367
	FN	TTGGTAGTTYGYGTTTTYGGGTGTAT					
	R	CCRCRACTCCRCTACAAATCTAAA					
CENPJ	F	TGGGAGTTTGGGGYGAYGTTTTATT	54	54	STANDARD	456	401
	FN	TTTYGTTAYGTYGTTYGTTTAYGTYGATTAT					
	R	AACCCCRAAATCCRAACAAAATCCTA					
CIDE-A	F	AGGATTGGTTATGATTTTAATTGTTTYGATT	53	53	STANDARD	315	247
	RN	ACCTAACTAACRCAAAAAAAAACRTTTTAA					
	R	CATAATCCCRAACRACCTCCATAA					
DBC1	F	AATATATTAGTATTATATTAAGTTTTTATT	52	52	STANDARD	448	360
	FN	TAGTTTTTTGAGTTTTAGGTATAGGAATTT					
	R	TAAATATAAACCAAACTACTAAAAACCAAA					
DDEF2	F	ATTTTTYGGTAGGGGGYGAGTTTYGTAT	54	54	STANDARD	622	494
	FN	TTTYGGYGGGGYGTATTTTTATAYGTATT					
	R	AACRCRAAAACCRCRCRAAACCTAA					
DLGAP1	F	GGGTAATAGAGTAAGATTTTGTAAAGTGTAATATT	55	55	STANDARD	542	437
	FN	TTGGAGTTTGATTTTYGGGTGTTTATGATT					
	R	CACRACCRTACRCTTATACTACAAAATATA					

EMILIN2	F	GAATGAGTTATTGGAAATGTTGTAAYGTAGATT	54	56	STANDARD	506	432
	RN	AAACACRCRAAACCAAAACCRTCTAAACTA					
	R	AAATACCCRAACTACRAACCRACRTAA					
EPSTI1	F	TAGGATTGTAGTAATTTTGYGGGTTAATYGTTTATT	54	54	STANDARD	390	273
	FN	TTAAGTTAGTTTAAGTTAATATTTTGGGTTGTTAATT					
	R	AAATTCACCACTCTATTACRAATATTCATAA					
FBLN2	F	GGATTTTAATTTTTGGTTTTGTTTTGTGGG	58	58	STANDARD	591	557
	FN	AGTGTTAGGTTTGGTATTTAGAGGTTAGGT					
	R	ACCTCTATACCCRCATCTATACACTAA					
FOXF2	F	TTTYGGGGTTTAGGTYGYGGTTTTAT	56	56	STANDARD	345	283
	RN	AAAAACCCARCTAAARCAAARCCRCAAAARCA					
	R	CTCRATAATCATCTAAAACCCRAAAACRAAA					
FOXQ1	F	AATTGGGGAGTYGTTATTATTTYGGTT	54	54	STANDARD	480	354
	RN	CTCRTTAAAAAACCCAAAAAAAATTCTTCACTAA					
	R	CRATTCRACRTTAAAACTTTAAAATAAATATATA					
GSC	F	GTTAAATTGGTTTTGTTTGTTATTTGTATATTATT	54	54	TOUCHDOWN	362	287
	RN	AAAACRCRCCRAAAACAAAACCTTAAAATA					
	R	AACRACTAAAATATTATCRATACTAAACATACTAA					
KLF11	F	TAGGGTTTTAGGTTTYGTTTTTYGTTTYGATT	54	54	TOUCHDOWN	435	354
	RN	AACRCAACAACRACCRCACRTAAACTA					
	R	TCRTCTAAACCTACRAAATCCRACRTATA					
KLF13	F	TTTTTTAGTYGGGGGTTAAGGTTGTTATTAT	54	54	STANDARD	410	297
	RN	TAACCRAAAACRAAACAAACRCCATTAA					
	R	AATTACAAACRACRACRACTACRAAAAACTA					
LYPD5	F	TYGTTATTAGTTTATTGTTTTTATTTTYGTAYGTAT	54	54	STANDARD	568	297
	FN	TATTYGGAGTATAGGTTATAATTTTTAGAAATT					
	R	TCACCATTCRCACRATAAAACACATCCTTAA					
MAD1L1	F	TTTTGGTTTAGTTTTYGGTTTYGAGTTT	54	54	STANDARD	374	274
	FN	TAGATTYGGTTTTTTAYGGTTYGATTGTT					
	R	AAATCCTAAAACACRAAATCTAAATCRTTATA					
NRXN2	F	TTTAGTTTYGYGTTGYGYGAAGATT	53	53	STANDARD	424	246
	RN	AACRAACAACRCAACCAATACAAAAA					
	R	CCRAAAAAAATACTCCRCRAATCAAA					
ONECUT1	F	GAGGGGYGATGTTTTTAAGTATAGTATTTT	57	57	STANDARD	552	390
	FN	TYGTAYGTGTTYGTAGTTTTTTGGTTATTT					
	R	AAAAATCCRCRCCRTTAAAAAAACRCAATA					

PAX9	F	GAATTAAAGYGGGTTTTAGATAGTAATAGGTT	58	58	STANDARD	582	512
	FN	GTTYGTAGTTYGTGGGGTTGATTTTAT					
	R	CAACTAAATTTCCCTAAAACTACAACRACAAATA					
PHF2	F	TTTTTYGYGTATATTAGYGTTTTGAGTT	56	56	STANDARD	371	293
	RN	AATACACRAACACCRTCRCCATATTA					
	R	ACCTACCRTAAAACCAATCCTTACAA					
POU4F1	F	AAAGGTGTTTATTTTTAGATAATAGTTGGATT	53	53	STANDARD	614	535
	FN	TATGTTYGGGTAGTTATAGGTGTATTATT					
	R	CATAACAAAATAAAACTACTTACTATTCATAA					
SALL1	F	TATTTTTTATTYGGGGYGGTTTAATTT	54	54	STANDARD	626	507
	RN	TAACCCAATAAACCRACCRCRAAACTAA					
	R	ACRATTAATCATAAAAAAACTCTTTAAAAAACRATTAA					
SALL3	F	TYGTATTYGGGTTTYGTTATAGTYGTATT	53	53	STANDARD	432	263
	RN	AACCRCRCATAAAACTAAACAATTAAACTAATA					
	R	CCTCRTCCRACTTAAAATACTAAAACTTAA					
SESN3	F	TTTTGTTTAGAAAGGAAGGTTTAGAGAAT	53	53	STANDARD	369	309
	RN	TAAACTCCAATAAACACAAAACTCCTTA					
	R	AAAACAAATAATATCATAATACACAATCCTATAA					
SIM2	F	GTATTTATTTGAGTGTTAYGTAAGTTATTTTAATAT	53	53	STANDARD	497	384
	RN	ACACRCACRCRACAAAAACTACTAA					
	R	ATCTAATCAAACACTAAAACRATAACACTAA					
TP53INP1	F	TAGGGYGGTTTYGGGGTTTTYGTAT	56	56	STANDARD	537	389
	RN	CCAACCRACAACRAAACRACCCTAA					
	R	CCRCCRAAAACCRACCCACTAAATA					

Appendix B.I_CoBRA primers used in the MIRA study

Primer sequences are shown for all genes analysed by CoBRA in the MIRA study (chapter 3). In each case forward (F), reverse (R) and nested (either forward nested (FN) or reverse nested (RN)) primer sequences are shown. Annealing temperatures and product sizes are given for primary and secondary PCR reactions. Also shown are whther the PCR was carried out using a the standard CoBRA PCR program or the touchdown CoBRA PCR program.

GENE	PRIMER	PRIMER SEQUENCE 5'-3'	ANNEALING TEMPERATURE (°C)	PRODUCT SIZE (bp)
CIDE-A	F	ATCACAGACTAAGCGAGTCC	56	364
	R	TAAGGCAGCCGATGAAGTCC		
COMP	F	ACTGCAGGAAACCAACGCG	56	474
	R	GAGTTGGGGACGGAGTTATG		
DBC1	F	CACCGTCAAGGATTTACAACCAG	56	258
	R	GTCCATATACATGGTCAAAGCCTCC		
EMILIN2	F	CGCCAGGAACAAGAACTGG	56	465
	R	CCTGGGCTCTGAGAATTGG		
EPSTI1	F	GAGCTAAACCGGTTCACC	56	178
	R	GCCTTCATTTTTTGGAGTTCAG		
FBLN2	F	AATGGACCCTGCAAGCAGGTGTGCA	56	266
	R	CAGGTGAGTGCCTTGTAGCAG		
SALL1	F	TCGTGCATGTTTCTCCTCAG	56	334
	R	GCTGCTGCATACTGATTCCA		
SESN3	F	AAAAGAATCAGAGTGTCTCAAC	56	260
	R	CCATGCGCAACATGTAAAAC		
SIM2s	F	AGTATATGCTGGACATGTCC	56	219
	R	ATAGGGTCTTCTCGATCAGG		
SIM2l	F	AGCTAAAAATCCTCCAGAGC	56	296
	R	TGGTAGTTGAGCAGCACGAAGG		

Appendix B.II_Expression primer sets used in the MIRA gene analysis
Forward (F) and reverse (R) primer sequences are shown for each gene analysed for expression in chapter 3. Annealing temperatures and product sizes are also shown.

	Accession	Gene Symbol	GeneName	SystematicName	Probe Location	Total number of cell lines >10	no. Probes
1	ref|NM_005069|ref|S	SIM2	single-minded homolog 2 (Drosophila)	chr21:036991933-036991977	PROMOTER	5	39
2	ref|NM_005068|ref|S	SIM1	single-minded homolog 1 (Drosophila)	chr6:101022497-101022556	PROMOTER	5	34
3	ref|NM_001426|ref|E	EN1	engrailed homeobox 1	chr2:119331833-119331877	PROMOTER	5	27
4	ref|NM_003735|ref|P	PCDHGA12	protocadherin gamma subfamily A, 12	chr5:140787476-140787520	PROMOTER	5	26
5	ref|NM_145285|ref|N	NKX2-3	NK2 transcription factor related, locus 3 (Drosophila)	chr10:101279919-101279968	PROMOTER	5	24
6	ref|NM_006194|ref|P	PAX9	paired box gene 9	chr14:036197711-036197755	PROMOTER	5	24
7	ref|NM_005523|ref|H	HOXA11	homeobox A11	chr7:027198283-027198327	PROMOTER	5	23
8	ref|NM_012183|ref|F	FOXD3	forkhead box D3	chr1:063558671-063558715	PROMOTER	5	20
9	ref|NM_001604|ref|P	PAX6	paired box gene 6 (aniridia, keratitis)	chr11:031798336-031798380	PROMOTER	5	20
10	ref|NM_032772|ref|Z	ZNF503	zinc finger protein 503	chr10:076837683-076837727	PROMOTER	5	20
11	ref|NM_001079668|r	TITF1	thyroid transcription factor 1	chr14:036061739-036061783	PROMOTER	5	18
12	ref|NM_021240|ref|D	DMRT3	doublesex and mab-3 related transcription factor 3	chr9:000960899-000960943	PROMOTER	5	16
13	ref|NM_001485|ref|G	GBX2	gastrulation brain homeobox 2	chr2:236742543-236742602	PROMOTER	5	16
14	ref|NM_032391|ref|P	PRAC	small nuclear protein PRAC	chr17:044156171-044156225	PROMOTER	5	16
15	ref|NM_020063|ref|B	BARHL2	BarH-like 2 (Drosophila)	chr1:090965044-090965088	PROMOTER	5	14
16	ref|NM_138281|ref|D	DLX4	distal-less homeobox 4	chr17:045398069-045398113	PROMOTER	5	14
17	ref|NM_173849|ref|G	GSC	goosecoid	chr14:094307384-094307428	PROMOTER	5	14
18	ref|NM_000325|ref|P	PITX2	paired-like homeodomain transcription factor 2	chr4:111769654-111769708	PROMOTER	5	12
19	ref|NM_005519|ref|H	HMX2	homeobox (H6 family) 2	chr10:124892389-124892434	PROMOTER	5	11
20	ref|NM_014360|ref|N	NKX2-8	NK2 transcription factor related, locus 8 (Drosophila)	chr14:036123094-036123152	PROMOTER	5	11
21	ref|NM_020766|ref|P	PCDH19	protocadherin 19	chrX:099549877-099549926	PROMOTER	5	11
22	ref|NM_000738|ref|C	CHRM1	cholinergic receptor, muscarinic 1	chr11:062450280-062450324	PROMOTER	5	10
23	ref|NM_004122|ref|G	GHSR	growth hormone secretagogue receptor	chr3:173650117-173650167	PROMOTER	5	10
24	ref|NM_016358|ref|I	IRX4	iroquois homeobox protein 4	chr5:001938805-001938849	PROMOTER	5	10
25	ref|NM_004498|ref|O	ONECUT1	one cut domain, family member 1	chr15:050871053-050871098	PROMOTER	5	10
26	ref|NM_032109|ref|O	OTP	orthopedia homeobox	chr5:076971045-076971092	PROMOTER	5	10
27	ref|NM_018902|ref|P	PCDHA11	protocadherin alpha 11	chr5:140222721-140222765	PROMOTER	5	10
28	ref|NM_000209|ref|P	PDX1	pancreatic and duodenal homeobox 1	chr13:027389600-027389645	PROMOTER	5	10
29	ref|NM_017977|ref|A	AIM1L	absent in melanoma 1-like	chr1:026559303-026559347	PROMOTER	5	9
30	ref|NM_006898|ref|H	HOXD3	homeobox D3	chr2:176736231-176736280	PROMOTER	5	9
31	mirna|hsa-mir-183|m	hsa-mir-183		chr7:129209466-129209510	PROMOTER	5	9
32	ref|NM_198289|ref|C	CIDEA	cell death-inducing DFFA-like effector a	chr18:012244652-012244696	PROMOTER	5	8
33	ref|NM_001989|ref|E	EVX1	even-skipped homeobox 1	chr7:027245650-027245694	PROMOTER	5	8
34	ref|NM_001452|ref|F	FOXF2	forkhead box F2	chr6:001330099-001330143	PROMOTER	5	8
35	ref|NM_006562|ref|L	LBX1	ladybird homeobox 1	chr10:102986147-102986205	PROMOTER	5	8
36	ref|NM_172337|ref|O	OTX2	orthodenticle homeobox 2	chr14:056346176-056346229	PROMOTER	5	8
37	ref|NM_007374|ref|S	SIX6	sine oculis homeobox homolog 6 (Drosophila)	chr14:060043824-060043875	PROMOTER	5	8
38	ref|NM_199425|ref|V	VSX1	visual system homeobox 1 homolog, CHX10-like (zebrafish)	chr20:025011822-025011866	PROMOTER	5	8
39	ref|NM_006557|ref|D	DMRT2	doublesex and mab-3 related transcription factor 2	chr9:001035959-001036013	PROMOTER	5	7
40	ref|NM_153038|ref|F	FLJ32447	hypothetical protein LOC151278	chr2:222870183-222870240	PROMOTER	5	7
41	ref|NM_021973|ref|H	HAND2	heart and neural crest derivatives expressed 2	chr4:174689199-174689245	PROMOTER	5	7
42	ref|NM_019558|ref|H	HOXD8	homeobox D8	chr2:176697833-176697892	PROMOTER	5	7
43	ref|NM_018918|ref|P	PCDHGA5	protocadherin gamma subfamily A, 5	chr5:140724058-140724102	PROMOTER	5	7
44	ref|NM_005413|ref|S	SIX3	sine oculis homeobox homolog 3 (Drosophila)	chr2:045013603-045013647	PROMOTER	5	7
45	ref|NM_000729|ref|C	CCK	cholecystokinin	chr3:042281937-042281983	PROMOTER	5	6
46	ref|NM_000522|ref|H	HOXA13	homeobox A13	chr7:027211864-027211923	PROMOTER	5	6
47	ref|NM_019102|ref|H	HOXA5	homeobox A5	chr7:027150867-027150926	PROMOTER	5	6
48	ref|NM_005946|ref|M	MT1A	metallothionein 1A	chr16:055229819-055229863	PROMOTER	5	6
49	ref|NM_004387|ref|N	NKX2-5	NK2 transcription factor related, locus 5 (Drosophila)	chr5:172596396-172596446	PROMOTER	5	6
50	ref|NM_020298|ref|A	ABCC9	ATP-binding cassette, sub-family C (CFTR/MRP), member 9	chr12:021986684-021986733	PROMOTER	5	5
51	ref|NM_021570|ref|B	BARX1	BarH-like homeobox 1	chr9:095762026-095762085	PROMOTER	5	5
52	ref|NM_001267|ref|C	CHAD	chondroadherin	chr17:045901905-045901949	PROMOTER	5	5
53	ref|NM_022659|ref|E	EBF2	early B-cell factor 2	chr8:025965385-025965440	PROMOTER	5	5
54	ref|NM_152393|ref|K	KBTBD5	kelch repeat and BTB (POZ) domain containing 5	chr3:042701982-042702026	PROMOTER	5	5

55	ref|NM_018916|ref|P	PCDHGA3	protocadherin gamma subfamily A, 3	chr5:140703684-140703740	PROMOTER	5	5
56	ref|NM_018920|ref|P	PCDHGA7	protocadherin gamma subfamily A, 7	chr5:140742524-140742569	PROMOTER	5	5
57	ref|NM_006237|ref|P	POU4F1	POU domain, class 4, transcription factor 1	chr13:078080099-078080144	PROMOTER	5	5
58	ref|NM_013435|ref|R	RAX	retina and anterior neural fold homeobox	chr18:055091952-055091996	PROMOTER	5	5
59	ref|NM_002968|ref|S	SALL1	sal-like 1 (Drosophila)	chr16:049745400-049745459	PROMOTER	5	5
60	ref|NM_080717|ref|T	TBX5	T-box 5	chr12:113330972-113331031	PROMOTER	5	5
61	ref|NM_138700|ref|T	TRIM40	tripartite motif-containing 40	chr6:030203273-030203321	PROMOTER	5	5
62	ref|NM_152343|ref|C	C17orf46	chromosome 17 open reading frame 46	chr17:040695617-040695661	PROMOTER	5	4
63	ref|NM_152377|ref|C	C1orf87	chromosome 1 open reading frame 87	chr1:060312136-060312180	PROMOTER	5	4
64	ref|NM_012261|ref|C	C20orf103	chromosome 20 open reading frame 103	chr20:009435912-009435968	PROMOTER	5	4
65	ref|NM_022778|ref|C	CCDC21	coiled-coil domain containing 21	chr1:026424092-026424138	PROMOTER	5	4
66	ref|NM_004364|ref|C	CEBPA	CCAAT/enhancer binding protein (C/EBP), alpha	chr19:038486768-038486812	PROMOTER	5	4
67	ref|NM_004474|ref|F	FOXD2	forkhead box D2	chr1:047671802-047671846	PROMOTER	5	4
68	ref|NM_018926|ref|P	PCDHGB6	protocadherin gamma subfamily B, 6	chr5:140767594-140767653	PROMOTER	5	4
69	ref|NM_005986|ref|S	SOX1	SRY (sex determining region Y)-box 1	chr13:111760308-111760354	PROMOTER	5	4
70	ref|NM_031912|ref|S	SYT15	synaptotagmin XV	chr10:046391257-046391303	PROMOTER	5	4
71	ref|NM_152476|ref|Z	ZNF560	zinc finger protein 560	chr19:009470275-009470319	PROMOTER	5	4
72	ref|NM_182703|ref|A	ANKDD1A	ankyrin repeat and death domain containing 1A	chr15:062984784-062984834	PROMOTER	5	3
73	ref|NM_001007793|r	BUB3	BUB3 budding uninhibited by benzimidazoles 3 homolog (yeast)	chr10:124901023-124901070	PROMOTER	5	3
74	ref|NM_133483|ref|G	GEFT	RAC/CDC42 exchange factor	chr12:056290473-056290517	PROMOTER	5	3
75	ref|NM_030661|ref|H	HOXA3	homeobox A3	chr7:027128948-027129007	PROMOTER	5	3
76	ref|NM_022658|ref|H	HOXC8	homeobox C8	chr12:052685570-052685629	PROMOTER	5	3
77	ref|NM_145805|ref|I	ISL2	ISL2 transcription factor, LIM/homeodomain, (islet-2)	chr15:074414745-074414789	PROMOTER	5	3
78	ref|NM_030915|ref|L	LBH	limb bud and heart development homolog (mouse)	chr2:030307042-030307088	PROMOTER	5	3
79	ref|NM_005583|ref|L	LYL1	lymphoblastic leukemia derived sequence 1	chr19:013076587-013076631	PROMOTER	5	3
80	ref|NM_005461|ref|M	MAFB	v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)	chr20:038752812-038752856	PROMOTER	5	3
81	ref|NM_005947|ref|M	MT1B	metallothionein 1B	chr16:055234863-055234907	PROMOTER	5	3
82	ref|NM_004350|ref|R	RUNX3	runt-related transcription factor 3	chr1:025131262-025131309	PROMOTER	5	3
83	ref|NM_171999|ref|S	SALL3	sal-like 3 (Drosophila)	chr18:074838411-074838469	PROMOTER	5	3
84	ref|NM_005627|ref|S	SGK	serum/glucocorticoid regulated kinase	chr6:134539389-134539433	PROMOTER	5	3
85	ref|NM_003181|ref|T	T	T, brachyury homolog (mouse)	chr6:166503036-166503084	PROMOTER	5	3
86	ref|NM_173084|ref|T	TRIM59	tripartite motif-containing 59	chr3:161650924-161650968	PROMOTER	5	3
87	ref|NM_138931|ref|B	BCL6	B-cell CLL/lymphoma 6 (zinc finger protein 51)	chr3:188938175-188938234	PROMOTER	5	2
88	ref|NM_057158|ref|D	DUSP4	dual specificity phosphatase 4	chr8:029263392-029263437	PROMOTER	5	2
89	ref|NM_005226|ref|E	EDG3	endothelial differentiation, sphingolipid G-protein-coupled receptor, 3	chr9:090795739-090795798	PROMOTER	5	2
90	ref|NM_005228|ref|E	EGFR	epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)	chr7:055053802-055053856	PROMOTER	5	2
91	ref|NM_183412|ref|F	FBXO44	F-box protein 44	chr1:011631797-011631841	PROMOTER	5	2
92	ref|NM_021032|ref|F	FGF12	fibroblast growth factor 12	chr3:193609142-193609201	PROMOTER	5	2
93	ref|NM_006735|ref|H	HOXA2	homeobox A2	chr7:027109815-027109859	PROMOTER	5	2
94	ref|NM_017550|ref|M	MIER2	mesoderm induction early response 1, family member 2	chr19:000296298-000296342	PROMOTER	5	2
95	ref|NM_198504|ref|P	PAQR9	progestin and adipoQ receptor family member IX	chr3:144165959-144166018	PROMOTER	5	2
96	ref|NM_001025300|r	RAB12	RAB12, member RAS oncogene family	chr18:008598945-008598989	PROMOTER	5	2
97	ref|NM_017793|ref|R	RPP25	ribonuclease P 25kDa subunit	chr15:073038585-073038631	PROMOTER	5	2
98	ref|NM_003030|ref|S	SHOX2	short stature homeobox 2	chr3:159304168-159304212	PROMOTER	5	2
99	ref|NM_001080508|r	TBX18	T-box 18	chr6:085539442-085539486	PROMOTER	5	2
100	ref|NM_019110|ref|Z	ZKSCAN4-C6orf194	zinc finger with KRAB and SCAN domains 4	chr6:028334977-028335021	DIVERGENT_PROMOTER	5	2
101	ref|NM_014292|ref|C	CBX6	chromobox homolog 6	chr22:037598992-037599036	PROMOTER	5	1
102	ref|NM_001793|ref|C	CDH3	cadherin 3, type 1, P-cadherin (placental)	chr16:067234397-067234441	PROMOTER	5	1
103	ref|NM_018451|ref|C	CENPJ	centromere protein J	chr13:024404315-024404359	PROMOTER	5	1
104	ref|NM_001824|ref|C	CKM	creatine kinase, muscle	chr19:050519039-050519083	PROMOTER	5	1
105	ref|NM_130773|ref|C	CNTNAP5	contactin associated protein-like 5	chr2:124498611-124498655	PROMOTER	5	1
106	ref|NM_005251|ref|F	FOXC2	forkhead box C2 (MFH-1, mesenchyme forkhead 1)	chr16:085157085-085157133	PROMOTER	5	1
107	ref|NM_002196|ref|I	INSM1	insulinoma-associated 1	chr20:020294171-020294230	PROMOTER	5	1
108	ref|NM_006437|ref|P	PARP4	poly (ADP-ribose) polymerase family, member 4	chr13:023985096-023985155	PROMOTER	5	1
109	ref|NM_018935|ref|P	PCDHB15	protocadherin beta 15	chr5:140601943-140601992	PROMOTER	5	1
110	ref|NM_003736|ref|P	PCDHGB4	protocadherin gamma subfamily B, 4	chr5:140747497-140747552	PROMOTER	5	1
111	ref|NM_018927|ref|P	PCDHGB7	protocadherin gamma subfamily B, 7	chr5:140777305-140777349	PROMOTER	5	1

112	ref|NM_080392|ref|F	PTP4A2	protein tyrosine phosphatase type IVA, member 2	chr1:032183113-032183164	PROMOTER	5	1
113	ref|NM_133327|ref|S	SEMA6B	sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6B	chr19:004517343-004517402	PROMOTER	5	1
114	ref|NM_001043|ref|S	SLC6A2	solute carrier family 6 (neurotransmitter transporter, noradrenalin), member 2	chr16:054248005-054248049	PROMOTER	5	1
115	ref|NM_003081|ref|S	SNAP25	synaptosomal-associated protein, 25kDa	chr20:010146077-010146134	PROMOTER	5	1
116	ref|NM_053053|ref|T	TADA1L	transcriptional adaptor 1 (HFI1 homolog, yeast)-like	chr1:165120176-165120221	PROMOTER	5	1
117	ref|NM_014383|ref|Z	ZBTB32	zinc finger and BTB domain containing 32	chr19:040887064-040887108	PROMOTER	5	1
118	ref|NM_018008|ref|F	FEZF2	FEZ family zinc finger 2	chr3:062334646-062334701	PROMOTER	4	13
119	ref|NM_004789|ref|L	LHX2	LIM homeobox 2	chr9:125811708-125811767	PROMOTER	4	13
120	ref|NM_006361|ref|H	HOXB13	homeobox B13	chr17:044165697-044165753	PROMOTER	4	11
121	ref|NM_021926|ref|A	ALX4	aristaless-like homeobox 4	chr11:044294929-044294975	PROMOTER	4	10
122	ref|NM_020649|ref|C	CBX8	chromobox homolog 8 (Pc class homolog, Drosophila)	chr17:075387953-075388002	PROMOTER	4	10
123	ref|NM_002509|ref|N	NKX2-2	NK2 transcription factor related, locus 2 (Drosophila)	chr20:021451830-021451882	PROMOTER	4	10
124	ref|NM_003269|ref|N	NR2E1	nuclear receptor subfamily 2, group E, member 1	chr6:108592291-108592339	PROMOTER	4	10
125	ref|NM_152380|ref|T	TBX15	T-box 15	chr1:119337255-119337299	PROMOTER	4	10
126	ref|NM_003948|ref|C	CDKL2	cyclin-dependent kinase-like 2 (CDC2-related kinase)	chr4:076775087-076775131	PROMOTER	4	9
127	ref|NM_002398|ref|M	MEIS1	Meis homeobox 1	chr2:066506899-066506948	PROMOTER	4	9
128	ref|NM_003222|ref|T	TFAP2C	transcription factor AP-2 gamma (activating enhancer binding protein 2 gamma)	chr20:054635918-054635962	PROMOTER	4	9
129	ref|NM_002148|ref|H	HOXD10	homeobox D10	chr2:176685824-176685868	PROMOTER	4	8
130	ref|NM_014562|ref|O	OTX1	orthodenticle homeobox 1	chr2:063129400-063129453	PROMOTER	4	8
131	ref|NM_002052|ref|G	GATA4	GATA binding protein 4	chr8:011595682-011595741	PROMOTER	4	7
132	ref|NM_018896|ref|C	CACNA1G	calcium channel, voltage-dependent, T type, alpha 1G subunit	chr17:045991351-045991395	PROMOTER	4	6
133	ref|NM_032040|ref|C	CCDC8	coiled-coil domain containing 8	chr19:051608720-051608764	PROMOTER	4	6
134	ref|NM_012182|ref|F	FOXB1	forkhead box B1	chr15:058083339-058083393	PROMOTER	4	6
135	ref|NM_000817|ref|G	GAD1	glutamate decarboxylase 1 (brain, 67kDa)	chr2:171379884-171379928	PROMOTER	4	6
136	ref|NM_004821|ref|H	HAND1	heart and neural crest derivatives expressed 1	chr5:153842538-153842588	PROMOTER	4	6
137	ref|NM_024014|ref|H	HOXA6	homeobox A6	chr7:027154040-027154084	PROMOTER	4	6
138	ref|NM_024337|ref|I	IRX1	iroquois homeobox protein 1	chr5:003645038-003645097	PROMOTER	4	6
139	ref|NM_001004317|r	LIN28B	lin-28 homolog B (C. elegans)	chr6:105507526-105507570	PROMOTER	4	6
140	ref|NM_031439|ref|S	SOX7	SRY (sex determining region Y)-box 7	chr8:010627804-010627861	PROMOTER	4	6
141	ref|NM_003433|ref|Z	ZNF132	zinc finger protein 132	chr19:063643415-063643459	PROMOTER	4	6
142	ref|NM_000697|ref|A	ALOX12	arachidonate 12-lipoxygenase	chr17:006839961-006840005	PROMOTER	4	5
143	ref|NM_001878|ref|C	CRABP2	cellular retinoic acid binding protein 2	chr1:154942970-154943014	PROMOTER	4	5
144	ref|NM_004496|ref|F	FOXA1	forkhead box A1	chr14:037137657-037137701	PROMOTER	4	5
145	ref|NM_004473|ref|F	FOXE1	forkhead box E1 (thyroid transcription factor 2)	chr9:099654302-099654347	PROMOTER	4	5
146	ref|NM_001039132|r	ICAM4	intercellular adhesion molecule 4 (Landsteiner-Wiener blood group)	chr19:010258491-010258538	PROMOTER	4	5
147	ref|NM_145260|ref|O	OSR1	odd-skipped related 1 (Drosophila)	chr2:019424640-019424688	PROMOTER	4	5
148	ref|NM_175738|ref|R	RAB37	RAB37, member RAS oncogene family	chr17:070178761-070178805	PROMOTER	4	5
149	ref|NM_000346|ref|S	SOX9	SRY (sex determining region Y)-box 9 (campomelic dysplasia, autosomal sex-reversal)	chr17:067623636-067623680	PROMOTER	4	5
150	ref|NM_003199|ref|T	TCF4	transcription factor 4	chr18:051408312-051408371	PROMOTER	4	5
151	ref|NM_012138|ref|A	AATF	apoptosis antagonizing transcription factor	chr17:032377352-032377411	PROMOTER	4	4
152	ref|NM_003658|ref|B	BARX2	BarH-like homeobox 2	chr11:128749686-128749730	PROMOTER	4	4
153	ref|NM_174978|ref|C	C14orf39	chromosome 14 open reading frame 39	chr14:060022669-060022713	PROMOTER	4	4
154	ref|NM_005257|ref|G	GATA6	GATA binding protein 6	chr18:018000454-018000504	PROMOTER	4	4
155	ref|NM_024016|ref|H	HOXB8	homeobox B8	chr17:044052345-044052398	PROMOTER	4	4
156	ref|NM_006680|ref|M	ME3	malic enzyme 3, NADP(+)-dependent, mitochondrial	chr11:086061223-086061267	PROMOTER	4	4
157	ref|NM_000278|ref|P	PAX2	paired box gene 2	chr10:102487511-102487558	PROMOTER	4	4
158	ref|NM_178012|ref|T	TUBB2B	tubulin, beta 2B	chr6:003177225-003177269	PROMOTER	4	4
159	ref|NM_018338|ref|W	WDR52	WD repeat domain 52	chr3:114643014-114643058	PROMOTER	4	4
160	ref|NM_024967|ref|Z	ZNF556	zinc finger protein 556	chr19:002810253-002810297	PROMOTER	4	4
161	ref|NM_024833|ref|Z	ZNF671	zinc finger protein 671	chr19:062930878-062930926	PROMOTER	4	4
162	ref|NM_139056|ref|A	ADAMTS16	ADAM metallopeptidase with thrombospondin type 1 motif, 16	chr5:005192805-005192849	PROMOTER	4	3
163	ref|NM_014272|ref|A	ADAMTS7	ADAM metallopeptidase with thrombospondin type 1 motif, 7	chr15:076891365-076891409	PROMOTER	4	3
164	ref|NM_020926|ref|B	BCOR	BCL6 co-repressor	chrX:039851378-039851422	PROMOTER	4	3
165	ref|NM_133468|ref|B	BMPER	BMP binding endothelial regulator	chr7:033910092-033910151	PROMOTER	4	3
166	ref|NM_021179|ref|C	C1orf114	chromosome 1 open reading frame 114	chr1:167663430-167663482	PROMOTER	4	3
167	ref|NM_024007|ref|E	EBF1	early B-cell factor 1	chr5:158464250-158464297	PROMOTER	4	3
168	ref|NM_005895|ref|G	GOLGA3	golgi autoantigen, golgin subfamily a, 3	chr12:131924571-131924617	PROMOTER	4	3

169	ref|NM_003535|ref|H	HIST1H3J	histone cluster 1, H3j	chr6:027966601-027966660	PROMOTER	4	3
170	ref|NM_018953|ref|H	HOXC5	homeobox C5	chr12:052711417-052711463	PROMOTER	4	3
171	ref|NM_022123|ref|N	NPAS3	neuronal PAS domain protein 3	chr14:032472605-032472653	PROMOTER	4	3
172	ref|NM_175747|ref|O	OLIG3	oligodendrocyte transcription factor 3	chr6:137860078-137860122	PROMOTER	4	3
173	ref|NM_004248|ref|P	PRLHR	prolactin releasing hormone receptor	chr10:120345135-120345179	PROMOTER	4	3
174	ref|NM_000312|ref|P	PROC	protein C (inactivator of coagulation factors Va and VIIIa)	chr2:127890158-127890202	PROMOTER	4	3
175	ref|NM_022454|ref|S	SOX17	SRY (sex determining region Y)-box 17	chr8:055528894-055528938	PROMOTER	4	3
176	ref|NM_020417|ref|T	TBX20	T-box 20	chr7:035263668-035263712	PROMOTER	4	3
177	ref|NM_020064|ref|B	BARHL1	BarH-like 1 (Drosophila)	chr9:134445103-134445147	PROMOTER	4	2
178	ref|NM_152732|ref|C	C6orf206	chromosome 6 open reading frame 206	chr6:043720609-043720653	PROMOTER	4	2
179	ref|NM_053056|ref|C	CCND1	cyclin D1	chr11:069161969-069162013	PROMOTER	4	2
180	ref|NM_152890|ref|C	COL24A1	collagen, type XXIV, alpha 1	chr1:086394736-086394792	PROMOTER	4	2
181	ref|NM_030594|ref|C	CPEB1	cytoplasmic polyadenylation element binding protein 1	chr15:081114628-081114686	PROMOTER	4	2
182	ref|NM_004952|ref|E	EFNA3	ephrin-A3	chr1:153309976-153310020	PROMOTER	4	2
183	ref|NM_001451|ref|F	FOXF1	forkhead box F1	chr16:085098933-085098977	PROMOTER	4	2
184	ref|NM_001463|ref|F	FRZB	frizzled-related protein	chr2:183440000-183440049	PROMOTER	4	2
185	ref|NM_145657|ref|G	GSH1	GS homeobox 1	chr13:027261690-027261742	PROMOTER	4	2
186	ref|NM_003532|ref|H	HIST1H3E	histone cluster 1, H3e	chr6:026333291-026333350	PROMOTER	4	2
187	ref|NM_173860|ref|H	HOXC12	homeobox C12	chr12:052630142-052630192	PROMOTER	4	2
188	ref|NM_025251|ref|K	KIAA1688	KIAA1688 protein	chr8:145815654-145815698	PROMOTER	4	2
189	ref|NM_005372|ref|M	MOS	v-mos Moloney murine sarcoma viral oncogene homolog	chr8:057192711-057192763	PROMOTER	4	2
190	ref|NM_006162|ref|N	NFATC1	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1	chr18:075254420-075254464	PROMOTER	4	2
191	ref|NM_001008564|r	NUPL1	nucleoporin like 1	chr13:024773519-024773578	PROMOTER	4	2
192	ref|NM_006902|ref|P	PRRX1	paired related homeobox 1	chr1:168897122-168897167	PROMOTER	4	2
193	ref|NM_004586|ref|R	RPS6KA3	ribosomal protein S6 kinase, 90kDa, polypeptide 3	chrX:020196152-020196196	PROMOTER	4	2
194	ref|NM_144665|ref|S	SESN3	sestrin 3	chr11:094605243-094605287	PROMOTER	4	2
195	ref|NM_032291|ref|S	SGIP1	SH3-domain GRB2-like (endophilin) interacting protein 1	chr1:066771265-066771309	PROMOTER	4	2
196	ref|NM_205848|ref|S	SYT6	synaptotagmin VI	chr1:114498640-114498684	PROMOTER	4	2
197	ref|NM_030788|ref|T	TM7SF4	transmembrane 7 superfamily member 4	chr8:105411302-105411354	PROMOTER	4	2
198	ref|NM_000262|ref|N	NAGA-FAM109B	N-acetylgalactosaminidase, alpha-	chr22:040799884-040799928	DIVERGENT_PROMOTER	4	2
199	ref|NM_017879|ref|N	ZNF416-ZIK1	zinc finger protein 416	chr19:062786492-062786536	DIVERGENT_PROMOTER	4	2
200	ref|NM_005502|ref|A	ABCA1	ATP-binding cassette, sub-family A (ABC1), member 1	chr9:106730640-106730699	PROMOTER	4	1
201	ref|NM_001117|ref|A	ADCYAP1	adenylate cyclase activating polypeptide 1 (pituitary)	chr18:000894464-000894520	PROMOTER	4	1
202	ref|NM_130768|ref|A	ASZ1	ankyrin repeat, SAM and basic leucine zipper domain containing 1	chr7:116854930-116854974	PROMOTER	4	1
203	ref|NM_001166|ref|B	BIRC2	baculoviral IAP repeat-containing 2	chr11:101723042-101723086	PROMOTER	4	1
204	ref|NM_024600|ref|C	C16orf30	chromosome 16 open reading frame 30	chr16:001523904-001523948	PROMOTER	4	1
205	ref|NM_031448|ref|C	C19orf12	chromosome 19 open reading frame 12	chr19:034907436-034907495	PROMOTER	4	1
206	ref|NM_032648|ref|C	C1orf90	chromosome 1 open reading frame 90	chr1:032478591-032478635	PROMOTER	4	1
207	ref|NM_207406|ref|C	CCDC4	coiled-coil domain containing 4	chr4:041849665-041849720	PROMOTER	4	1
208	ref|NM_014516|ref|C	CNOT3	CCR4-NOT transcription complex, subunit 3	chr19:059333090-059333134	PROMOTER	4	1
209	ref|NM_030820|ref|C	COL21A1	collagen, type XXI, alpha 1	chr6:056220441-056220485	PROMOTER	4	1
210	ref|NM_001848|ref|C	COL6A1	collagen, type VI, alpha 1	chr21:046216961-046217005	PROMOTER	4	1
211	ref|NM_024027|ref|C	COLEC11	collectin sub-family member 11	chr2:003620325-003620369	PROMOTER	4	1
212	ref|NM_000499|ref|C	CYP1A1	cytochrome P450, family 1, subfamily A, polypeptide 1	chr15:072806266-072806318	PROMOTER	4	1
213	ref|NM_019074|ref|D	DLL4	delta-like 4 (Drosophila)	chr15:039006706-039006765	PROMOTER	4	1
214	ref|NM_001038493|r	DLX1	distal-less homeobox 1	chr2:172654359-172654403	PROMOTER	4	1
215	ref|NM_015268|ref|D	DNAJC13	DnaJ (Hsp40) homolog, subfamily C, member 13	chr3:133618764-133618808	PROMOTER	4	1
216	ref|NM_001039350|r	DPP6	dipeptidyl-peptidase 6	chr7:153215050-153215094	PROMOTER	4	1
217	ref|NM_152403|ref|E	EGFLAM	EGF-like, fibronectin type III and laminin G domains	chr5:038293528-038293572	PROMOTER	4	1
218	ref|NM_173160|ref|F	FXYD4	FXYD domain containing ion transport regulator 4	chr10:043177643-043177697	PROMOTER	4	1
219	ref|NM_005754|ref|G	G3BP1	GTPase activating protein (SH3 domain) binding protein 1	chr5:151130455-151130503	PROMOTER	4	1
220	ref|NM_002040|ref|G	GABPA	GA binding protein transcription factor, alpha subunit 60kDa	chr21:026029451-026029496	PROMOTER	4	1
221	ref|NM_002729|ref|H	HHEX	hematopoietically expressed homeobox	chr10:094438777-094438824	PROMOTER	4	1
222	ref|NM_014212|ref|H	HOXC11	homeobox C11	chr12:052645983-052646032	PROMOTER	4	1
223	ref|NM_021193|ref|H	HOXD12	homeobox D12	chr2:176672237-176672281	PROMOTER	4	1
224	ref|NM_016257|ref|H	HPCAL4	hippocalcin like 4	chr1:039929862-039929906	PROMOTER	4	1
225	ref|NM_024336|ref|I	IRX3	iroquois homeobox protein 3	chr16:052879099-052879158	PROMOTER	4	1

226	ref|NM_003597|ref|N	KLF11	Kruppel-like factor 11	chr2:010100230-010100280	PROMOTER	4	1
227	ref|NM_016269|ref|L	LEF1	lymphoid enhancer-binding factor 1	chr4:109312434-109312491	PROMOTER	4	1
228	ref|NM_001007125|r	LOC198437	bA299N6.3	chr20:062186247-062186303	PROMOTER	4	1
229	ref|NM_014628|ref|N	MAD2L1BP	MAD2L1 binding protein	chr6:043705522-043705581	PROMOTER	4	1
230	ref|NM_012301|ref|N	MAGI2	membrane associated guanylate kinase, WW and PDZ domain containing 2	chr7:078921596-078921655	PROMOTER	4	1
231	ref|NM_016060|ref|N	MED31	mediator of RNA polymerase II transcription, subunit 31 homolog (S. cerevisiae)	chr17:006498960-006499004	PROMOTER	4	1
232	ref|NM_177524|ref|N	MEST	mesoderm specific transcript homolog (mouse)	chr7:129913135-129913194	PROMOTER	4	1
233	ref|NM_022746|ref|N	MOSC1	MOCO sulphurase C-terminal domain containing 1	chr1:219026573-219026632	PROMOTER	4	1
234	ref|NM_001584|ref|N	MPPED2	metallophosphoesterase domain containing 2	chr11:030562310-030562356	PROMOTER	4	1
235	ref|NM_002440|ref|N	MSH4	mutS homolog 4 (E. coli)	chr1:076035068-076035114	PROMOTER	4	1
236	ref|NM_000256|ref|N	MYBPC3	myosin binding protein C, cardiac	chr11:047333481-047333525	PROMOTER	4	1
237	ref|NM_020443|ref|N	NAV1	neuron navigator 1	chr1:199883708-199883755	PROMOTER	4	1
238	ref|NM_002500|ref|N	NEUROD1	neurogenic differentiation 1	chr2:182256025-182256072	PROMOTER	4	1
239	ref|NM_007361|ref|N	NID2	nidogen 2 (osteonidogen)	chr14:051606531-051606590	PROMOTER	4	1
240	ref|NM_177400|ref|N	NKX6-2	NK6 transcription factor related, locus 2 (Drosophila)	chr10:134452116-134452160	PROMOTER	4	1
241	ref|NM_178864|ref|N	NPAS4	neuronal PAS domain protein 4	chr11:065944601-065944647	PROMOTER	4	1
242	ref|NM_005285|ref|N	NPBWR1	neuropeptides B/W receptor 1	chr8:054014637-054014696	PROMOTER	4	1
243	ref|NM_012226|ref|P	P2RX2	purinergic receptor P2X, ligand-gated ion channel, 2	chr12:131697163-131697207	PROMOTER	4	1
244	ref|NM_016223|ref|P	PACSIN3	protein kinase C and casein kinase substrate in neurons 3	chr11:047165175-047165227	PROMOTER	4	1
245	ref|NM_013942|ref|P	PAX3	paired box gene 3 (Waardenburg syndrome 1)	chr2:222876845-222876895	PROMOTER	4	1
246	ref|NM_020992|ref|P	PDLIM1	PDZ and LIM domain 1 (elfin)	chr10:097045210-097045260	PROMOTER	4	1
247	ref|NM_006211|ref|P	PENK	proenkephalin	chr8:057521511-057521555	PROMOTER	4	1
248	ref|NM_002741|ref|P	PKN1	protein kinase N1	chr19:014403711-014403770	PROMOTER	4	1
249	ref|NM_153819|ref|R	RASGRP2	RAS guanyl releasing protein 2 (calcium and DAG-regulated)	chr11:064270043-064270087	PROMOTER	4	1
250	ref|NM_213594|ref|R	RFX4	regulatory factor X, 4 (influences HLA class II expression)	chr12:105498750-105498794	PROMOTER	4	1
251	ref|NM_007218|ref|R	RNF139	ring finger protein 139	chr8:125555865-125555924	PROMOTER	4	1
252	ref|NM_003013|ref|S	SFRP2	secreted frizzled-related protein 2	chr4:154933182-154933241	PROMOTER	4	1
253	ref|NM_000451|ref|S	SHOX	short stature homeobox	chrY:000501132-000501175	PROMOTER	4	1
254	ref|NM_000340|ref|S	SLC2A2	solute carrier family 2 (facilitated glucose transporter), member 2	chr3:172229035-172229094	PROMOTER	4	1
255	ref|NM_018018|ref|S	SLC38A4	solute carrier family 38, member 4	chr12:045511388-045511446	PROMOTER	4	1
256	ref|NM_032034|ref|S	SLC4A11	solute carrier family 4, sodium bicarbonate transporter-like, member 11	chr20:003166524-003166574	PROMOTER	4	1
257	ref|NM_018967|ref|S	SNTG1	syntrophin, gamma 1	chr8:050985264-050985312	PROMOTER	4	1
258	ref|NM_006946|ref|S	SPTBN2	spectrin, beta, non-erythrocytic 2	chr11:066253004-066253054	PROMOTER	4	1
259	ref|NM_005078|ref|T	TLE3	transducin-like enhancer of split 3 (E(sp1) homolog, Drosophila)	chr15:068180002-068180048	PROMOTER	4	1
260	ref|NM_033285|ref|T	TP53INP1	tumor protein p53 inducible nuclear protein 1	chr8:096031583-096031627	PROMOTER	4	1
261	ref|NM_006953|ref|U	UPK3A	uroplakin 3A	chr22:044059391-044059435	PROMOTER	4	1
262	ref|NM_153453|ref|V	VGLL2	vestigial like 2 (Drosophila)	chr6:117691542-117691586	PROMOTER	4	1
263	ref|NM_003451|ref|Z	ZNF177	zinc finger protein 177	chr19:009334494-009334553	PROMOTER	4	1
264	ref|NM_032423|ref|Z	ZNF528	zinc finger protein 528	chr19:057592881-057592925	PROMOTER	4	1
265	ref|NM_014630|ref|Z	ZNF592	zinc finger protein 592	chr15:083092136-083092182	PROMOTER	4	1
266	ref|NM_033410|ref|Z	ZNF764	zinc finger protein 764	chr16:030480317-030480361	PROMOTER	4	1
267	ref|NM_138740|ref|N	C1orf43-UBAP2L	chromosome 1 open reading frame 43	chr1:152460153-152460197	DIVERGENT_PROMOTER	4	1
268	ref|NM_021259|ref|N	TMEM8-LOC441743	transmembrane protein 8 (five membrane-spanning domains)	chr16:000372852-000372911	DIVERGENT_PROMOTER	4	1

Appendix B.III_High stringency gene list

This table shows the list of genes identified through the high stringency analysis of the MIRA array. Gene names, accession numbers, gene product names, CpG location (systematic name), location of the probe with respect the promoter or divergent promoter classification, the number of cell lines with >10 fold difference between tumour and normal and the number of probes meeting this criteria for each gene are shown.

GENE	PRIMER	PRIMER SEQUENCE (5'-3')	ANNEALING TEMPERATURE (°C)		PROGRAM	PRODUCT SIZE (bp)	
			PRIMARY REACTION	SECONDARY REACTION		PRIMARY REACTION	SECONDARY REACTION
ACADL	F	GGGTTTTTAGAGTTATGTTTGGTTAATTTTATTATT	55	55	TOUCHDOWN	484	207
	FN	GGAYGTTAAAAAGGTTTAGAGTTGTTTYGA					
	RN	AACRATCAACTAAACRTCCACCTATAATA					
	R	AAACRTACRACCATATCCRAAACACAAA					
C1orf114	F	AGAGATAGTTATAAGAGGGAAATTTTATAATTAATAT	55	55	TOUCHDOWN	564	389
	RN	CCTCTTACCTATATAAAATACRACRACTAA					
	R	CCCACCAATACCAAACTCRATAAT					
CDKL2	F	AAGGTAGAAAGGYGTTTTTTGAYGAGAATT	56	56	TOUCHDOWN	479	376
	FN	GGTTTTGAGTTAATTATGGTTYGTGAYGAT					
	R	ACACTTAAAAAAAACCACRAACCRCATACTA					
CLDN6	F	GTTTGAGGGYGTTTYGGTTTATTYGATTAT	57	57	TOUCHDOWN	423	254
	FN	GYGTTTAATTTTTGGGAGTTTYGAAGGATT					
	R	ACCTAACRCCCTCACCCTAAATAAATCTAA					
COL1A2	F	GGAGTTTTTTATTTTATAAGTGGTTTATAGGGT	56	56	TOUCHDOWN	577	494
	FN	AATTTTGTTTATGTTGGGGTTGTAGAGTAT					
	R	TACAAACTCCTTATATCRCAAAACCCCTAA					
COX7A1	F	GAGTTGGAATAYGGTTTTATGYGTTTTTTTAT	55	55	TOUCHDOWN	553	475
	RN	AAATAAAAAACTCCRACCCAACCCATAA					
	R	CTATTTAACRAAAAATCCCAAACCCTAAAAATAA					
RECK	F	ATTTTGTTTAYGTTYGGYGATTTYGGGATT	57	57	TOUCHDOWN	375	259
	FN	TATTTATYGATAYGGGTTTTTTTTTYGGTATTGATT					
	R	ATCCCRCCCCCRAAAAACAAAATTACTA					
SFRP2	F	TTTATTTATGTTTGGTAATTTAGTAGAAATTTYGGAT	55	55	TOUCHDOWN	450	339
	RN	TCCRAAAAAAACTAATCACTACTTTCTAAATCTAA					
	R	ACCCRAAAAAACAAAATAAACCRTTAACTAA					
SFRP5	F	GTTYGAGTYGTTGTAYGGTYGTTTTTAT	56	56	TOUCHDOWN	380	350
	FN	AAGTYGTYGTAGTGTTTTGATATTTTTGTYGATT					
	R	ACRACRATACCAAAATCRTTATCCAAAAAAAACTTATAA					

SIPA1	F	GTTGTTTTAGGGYGTAAAGAGATGATATT	54	55	TOUCHDOWN	307	249
	RN	AAAACRCAACAACRAAAATTCCAAACCTAA					
	R	CTTCRCCTAAAAAACRAAACCATACTAA					
SST	F	TTTAAAATAGAGGGAGAYGGTTGAGAGTAT	56	56	TOUCHDOWN	361	299
	RN	CCRAAAAACRTCRAAAAAATCTCCTTACCTC					
	R	ACTAAACATCCCTTACRTCCAAACCAATAA					
TNFRSF10D	F	ATGGGTTTTTAGTGTTTTTTTTGGAGTTTA	55	55	TOUCHDOWN	597	494
	FN	ATTYGGAATTAAATYGTAGGTTTTGGGTTA					
	R	CAAATAARATATTCCCCAACCAAAA					
ZNF154	F	GGYGGTTATTTTAGTTTTTTTGAGGTGTGTTTAT	55	55	TOUCHDOWN	567	408
	FN	ATAGGGATYGTTGTTTTYGGGTTTTTATT					
	R	CATCCTCACAATAACAATACAACCCTAAA					

Appendix C.I_CoBRA primers used in the Infinium study

Primer sequences are shown for all genes analysed by CoBRA/sequencing in the Infinium study (chapter 4). In each case forward (F), reverse (R) and nested (either forward nested (FN) or reverse nested (RN)) primer sequences are shown. Annealing temperatures and product sizes are given for primary and secondary PCR reactions. Also shown are whether the PCR was carried out using the standard CoBRA PCR program or the touchdown CoBRA PCR program.

GENE	PRIMER	SEQUENCE (5'-3')	ANNEALING TEMPERATURE (°C)	PRODUCT SIZE (bp)
ABCA3	F	ATCATCACCTCCCACAGCAT	56	333
	R	GTAGTCGTCCACGCCGTACT		
ACADL	F	TTGGCAAAACAGTTGCTCAC	56	271
	R	ATTGGCTGAACTCTGGCATC		
C1orf114	F	AATTCAGCGAGCAAAGGAAA	56	330
	R	CGAAAAGGTTTGGCTTCTGA		
CDKL2	F	TCAAATGGATGGATTTGCTG	56	367
	R	GCTTGGATTCCTTGTGTGGT		
COL1A2	F	CCTGGCTTAAAGGGACACAA	56	305
	R	GCTGGTCAGCCCTGTAGAAG		
COX7A1	F	TGGACAGAGGAGGACTACGC	56	369
	R	AGGCCAGCGTTTATTGACAC		
LAMA2	F	TACCCGAAGAATTGGTCCAG	56	363
	R	ACCCCAGCATCATAGACAGC		
RECK	F	CCGAGTGTGCTTCTGTCAAG	56	398
	R	GGAAATGATGAGGGCAGAGA		
SFRP2	F	AGGACAACGACCTTTGCATC	56	290
	R	CAAGCTGTCTTTGAGCCACA		
SST	F	AGTTTGACCAGCCACTCTCC	56	231
	R	CCATAGCCGGGTTTGAGTTA		
TNFRSF10D	F	CAGGAAATCCAAGGTCAGGA	56	265
	R	AGCCTGCCTCATCTTCTTCA		
ZNF154	F	ACGAAAGCGTCTAAGCCTTG	56	324
	R	TTTTTCTCCAAGGTGCTGCT		

Appendix C.II_Expression primer sets used in the Infinium gene analysis
Forward (F) and reverse (R) primer sequences are shown for each gene analysed for expression in chapter 4. Annealing temperatures and product sizes are also shown.

	SYMBOL	Illumina ID	Dif ≥0.4	β ≥0.5	Meth+	CHR	PRODUCT	CPG_ISLAND
1	INPP5B	cg10784030	24	29	**29 (74.4%)**	1	inositol polyphosphate-5-phosphatase; 75kDa	FALSE
2	ABCA3	cg00949442	23	27	**27 (69.2%)**	16	ATP-binding cassette; sub-family A member 3	TRUE
3	COX7A1	cg24335895	19	27	**27 (69.2%)**	19	cytochrome c oxidase subunit VIIa polypeptide 1 (muscle) precursor	TRUE
4	C12orf34	cg02351381	17	24	**24 (61.5%)**	12	hypothetical protein LOC84915	FALSE
5	SST	cg02164046	13	23	**23 (59.0%)**	3	somatostatin preproprotein	TRUE
6	INPP5B	cg02309273	13	22	**22 (56.4%)**	1	inositol polyphosphate-5-phosphatase; 75kDa	FALSE
7	ZNF154	cg21790626	22	17	**22 (56.4%)**	19	zinc finger protein 154 (pHZ-92)	TRUE
8	ZFP42	cg06274159	18	20	**20 (51.3%)**	4	zinc finger protein 42	TRUE
9	GPC2	cg18691434	18	20	**20 (51.3%)**	7	glypican 2	TRUE
10	CDKN2A	cg09099744	16	20	**20 (51.3%)**	9	cyclin-dependent kinase inhibitor 2A isoform 3	TRUE
11	C1orf114	cg08047907	15	20	**20 (51.3%)**	1	hypothetical protein LOC57821	TRUE
12	C1orf114	cg13958426	20	20	**20 (51.3%)**	1	hypothetical protein LOC57821	TRUE
13	TMPRSS6	cg21271753	11	19	**19 (48.7%)**	22	transmembrane protease; serine 6	FALSE
14	KCNC3	cg06572160	10	19	**19 (48.7%)**	19	Shaw-related voltage-gated potassium channel protein 3	TRUE
15	ACOT11	cg10266490	19	19	**19 (48.7%)**	1	thioesterase; adipose associated isoform BFIT1	FALSE
16	CPT1C	cg21604803	19	15	**19 (48.7%)**	19	carnitine palmitoyltransferase 1C	TRUE
17	FLRT2	cg17410236	19	19	**19 (48.7%)**	14	fibronectin leucine rich transmembrane protein 2	TRUE
18	PPGB	cg19067730	15	18	**18 (46.2%)**	20	protective protein for beta-galactosidase	TRUE
19	DDAH2	cg18055007	13	18	**18 (46.2%)**	6	dimethylarginine dimethylaminohydrolase 2	TRUE
20	ZNF177	cg09643544	11	18	**18 (46.2%)**	19	zinc finger protein 177	TRUE
21	SIX6	cg19456540	18	18	**18 (46.2%)**	14	sine oculis homeobox homolog 6	TRUE
22	PPGB	cg08260891	16	17	**17 (43.6%)**	20	protective protein for beta-galactosidase	TRUE
23	ZNF154	cg08668790	14	17	**17 (43.6%)**	19	zinc finger protein 154 (pHZ-92)	TRUE
24	CLDN6	cg07384961	14	17	**17 (43.6%)**	16	claudin 6	TRUE
25	VSX1	cg23097006	10	17	**17 (43.6%)**	20	visual system homeobox 1 protein isoform a	TRUE
26	FCGRT	cg15528736	10	17	**17 (43.6%)**	19	Fc fragment of IgG; receptor; transporter; alpha	TRUE
27	PDGFRB	cg12727795	17	17	**17 (43.6%)**	5	platelet-derived growth factor receptor beta precursor	TRUE
28	AJAP1	cg17525406	17	17	**17 (43.6%)**	1	transmembrane protein SHREW1	TRUE
29	H2AFY	cg24628744	14	16	**16 (41.0%)**	5	H2A histone family; member Y isoform 2	TRUE
30	B3GAT1	cg11038843	14	16	**16 (41.0%)**	11	beta-1;3-glucuronyltransferase 1	TRUE
31	PRKAR1B	cg13577076	13	16	**16 (41.0%)**	7	protein kinase; cAMP-dependent; regulatory; type I; beta	TRUE
32	DNAJC6	cg09082287	11	16	**16 (41.0%)**	1	DnaJ (Hsp40) homolog; subfamily C; member 6	TRUE
33	CD9	cg08519905	10	16	**16 (41.0%)**	12	CD9 antigen	TRUE
34	ACVR1	cg09499849	9	16	**16 (41.0%)**	2	activin A type I receptor precursor	FALSE
35	KCNS2	cg03561565	9	16	**16 (41.0%)**	8	potassium voltage-gated channel; delayed-rectifier; subfamily S; member 2	TRUE
36	SLC5A7	cg16232126	6	16	**16 (41.0%)**	2	solute carrier family 5 (choline transporter); member 7	TRUE
37	SYN2	cg15873301	14	15	**15 (38.5%)**	3	synapsin II isoform IIa	TRUE
38	POU4F2	cg24199834	12	15	**15 (38.5%)**	4	POU domain; class 4; transcription factor 2	TRUE
39	RSNL2	cg01777397	12	15	**15 (38.5%)**	2	restin-like 2	TRUE
40	DPP4	cg12335708	12	15	**15 (38.5%)**	2	dipeptidylpeptidase IV	TRUE
41	ZNF660	cg22598028	11	15	**15 (38.5%)**	3	zinc finger protein 660	TRUE
42	LHCGR	cg12351433	10	15	**15 (38.5%)**	2	luteinizing hormone/choriogonadotropin receptor precursor	TRUE
43	CTPS	cg15043057	10	15	**15 (38.5%)**	1	CTP synthase	TRUE
44	TCF15	cg22449114	9	15	**15 (38.5%)**	20	basic helix-loop-helix transcription factor 15	TRUE
45	UCHL1	cg24715245	9	15	**15 (38.5%)**	4	ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase)	TRUE
46	C1orf104	cg22234962	9	15	**15 (38.5%)**	1	hypothetical protein LOC284618	TRUE
47	ZNF560	cg05221167	15	10	**15 (38.5%)**	19	zinc finger protein 560	TRUE
48	POU4F1	cg08097882	15	13	**15 (38.5%)**	13	POU domain; class 4; transcription factor 1	TRUE
49	HOXD8	cg15520279	15	9	**15 (38.5%)**	2	homeobox D8	TRUE
50	TMEM22	cg02672493	6	15	**15 (38.5%)**	3	transmembrane protein 22	TRUE
51	FOXD3	cg22815110	13	14	**14 (35.9%)**	1	forkhead box D3	TRUE
52	CYTL1	cg01114088	13	14	**14 (35.9%)**	4	cytokine-like 1	TRUE
53	ANGPTL5	cg19531130	12	14	**14 (35.9%)**	11	angiopoietin-like 5	FALSE
54	ADHFE1	cg01988129	11	14	**14 (35.9%)**	8	alcohol dehydrogenase; iron containing; 1	TRUE
55	SLC13A3	cg18468842	11	14	**14 (35.9%)**	20	solute carrier family 13 member 3 isoform a	TRUE
56	CCND2	cg17580045	11	14	**14 (35.9%)**	12	cyclin D2	TRUE
57	SLC9A3	cg02748539	9	14	**14 (35.9%)**	5	solute carrier family 9 (sodium/hydrogen exchanger); isoform 3	TRUE
58	BTG3	cg14380517	9	14	**14 (35.9%)**	21	B-cell translocation gene 3	TRUE
59	MAT1A	cg19423196	9	14	**14 (35.9%)**	10	methionine adenosyltransferase I; alpha	FALSE
60	CDH9	cg12864235	8	14	**14 (35.9%)**	5	cadherin 9; type 2 preproprotein	FALSE
61	LAMA2	cg19774122	14	13	**14 (35.9%)**	6	laminin alpha 2 subunit precursor	TRUE
62	GRIN3A	cg18794577	14	14	**14 (35.9%)**	9	glutamate receptor; ionotropic; N-methyl-D-aspartate 3A	TRUE
63	GRIA1	cg08578734	14	12	**14 (35.9%)**	5	glutamate receptor; ionotropic; AMPA 1	FALSE
64	ZNF577	cg16731240	14	13	**14 (35.9%)**	19	zinc finger protein 577	TRUE
65	DEPDC2	cg13652336	14	14	**14 (35.9%)**	8	DEP domain containing 2 isoform a	TRUE
66	SCUBE3	cg00347904	14	13	**14 (35.9%)**	6	signal peptide; CUB domain; EGF-like 3	TRUE
67	SYDE1	cg04981492	7	14	**14 (35.9%)**	19	synapse defective 1; Rho GTPase; homolog 1	TRUE
68	UGT3A2	cg07084163	5	14	**14 (35.9%)**	5	UDP glycosyltransferase 3 family; polypeptide A2	TRUE
69	IHH	cg25908985	12	13	**13 (33.3%)**	2	Indian hedgehog homolog	TRUE
70	PCDHGB7	cg23563234	10	13	**13 (33.3%)**	5	protocadherin gamma subfamily B; 7 isoform 1 precursor	TRUE
71	ADAMTSL1	cg16714091	10	13	**13 (33.3%)**	9	ADAMTS-like 1 isoform 2	TRUE
72	MCAM	cg21096399	10	13	**13 (33.3%)**	11	melanoma cell adhesion molecule	TRUE
73	PRKCB1	cg05436658	9	13	**13 (33.3%)**	16	protein kinase C; beta isoform 2	TRUE
74	ITR	cg09582042	9	13	**13 (33.3%)**	13	intimal thickness-related receptor	TRUE
75	TNNI3	cg18838701	9	13	**13 (33.3%)**	19	troponin I; cardiac	FALSE

	SYMBOL	Infin. Probe	Dif 0.4	β 0.5	Meth+	CHR	PRODUCT	CPG_ISLAND
76	GUCY2D	cg25465406	13	13	**13 (33.3%)**	17	guanylate cyclase 2D; membrane (retina-specific)	TRUE
77	ZNF132	cg13877915	13	8	**13 (33.3%)**	19	zinc finger protein 132 (clone pHZ-12)	TRUE
78	COL1A2	cg25300386	13	11	**13 (33.3%)**	7	alpha 2 type I collagen	TRUE
79	PRR3	cg21264055	13	12	**13 (33.3%)**	6	proline-rich protein 3	TRUE
80	MYO3A	cg08441170	13	11	**13 (33.3%)**	10	myosin IIIA	TRUE
81	ST8SIA2	cg20339230	13	13	**13 (33.3%)**	15	ST8 alpha-N-acetyl-neuraminide alpha-2;8-sialyltransferase 2	TRUE
82	PCDHAC1	cg12629325	13	12	**13 (33.3%)**	5	protocadherin alpha subfamily C; 1 isoform 1 precursor	TRUE
83	NPY	cg05158615	3	13	**13 (33.3%)**	7	neuropeptide Y	TRUE
84	UAP1L1	cg04582938	7	13	**13 (33.3%)**	9	UDP-N-acteylglucosamine pyrophosphorylase 1-like 1	TRUE
85	SFRP2	cg23207990	7	13	**13 (33.3%)**	4	secreted frizzled-related protein 2 precursor	TRUE
86	C5orf4	cg10257049	7	13	**13 (33.3%)**	5	hypothetical protein LOC10826 isoform 2	FALSE
87	EFNA2	cg11885098	6	13	**13 (33.3%)**	19	ephrin-A2 precursor	TRUE
88	RPIP8	cg13102585	2	13	**13 (33.3%)**	17	RaP2 interacting protein 8	TRUE
89	ACOT8	cg08101264	11	12	**12 (30.8%)**	20	peroxisomal acyl-CoA thioesterase 1 isoform b	FALSE
90	SH3TC2	cg01965939	11	12	**12 (30.8%)**	5	SH3 domain and tetratricopeptide repeats 2	FALSE
91	NRIP2	cg05194726	10	12	**12 (30.8%)**	12	nuclear receptor interacting protein 2	FALSE
92	RALBP1	cg15982419	10	12	**12 (30.8%)**	18	ralA binding protein 1	TRUE
93	SFRP4	cg08261094	10	12	**12 (30.8%)**	7	secreted frizzled-related protein 4	TRUE
94	HIST1H4L	cg17651821	10	12	**12 (30.8%)**	6	H4 histone family; member K	FALSE
95	SLC8A2	cg22123464	9	12	**12 (30.8%)**	19	solute carrier family 8 member 2	FALSE
96	TRIM58	cg07533148	9	12	**12 (30.8%)**	1	tripartite motif-containing 58	TRUE
97	BRAF	cg21043558	8	12	**12 (30.8%)**	7	v-raf murine sarcoma viral oncogene homolog B1	FALSE
98	SLC13A5	cg16652063	8	12	**12 (30.8%)**	17	solute carrier family 13 (sodium-dependent citrate transporter); member 5	TRUE
99	GSTP1	cg22224704	8	12	**12 (30.8%)**	11	glutathione transferase	TRUE
100	FERD3L	cg25691167	12	11	**12 (30.8%)**	7	nephew of atonal 3	TRUE
101	ENPP2	cg06193397	12	11	**12 (30.8%)**	8	ectonucleotide pyrophosphatase/phosphodiesterase 2 (autotaxin)	FALSE
102	HOXA7	cg23432345	12	12	**12 (30.8%)**	7	homeobox protein A7	TRUE
103	CART	cg23300372	12	12	**12 (30.8%)**	5	cocaine- and amphetamine-regulated transcript	TRUE
104	SFRP5	cg05937453	12	11	**12 (30.8%)**	10	secreted frizzled-related protein 5	TRUE
105	TNFRSF10D	cg22783363	12	10	**12 (30.8%)**	8	tumor necrosis factor receptor superfamily; member 10d precursor	TRUE
106	RUSC2	cg15271616	12	11	**12 (30.8%)**	9	RUN and SH3 domain containing 2	TRUE
107	IL11RA	cg21504624	7	12	**12 (30.8%)**	9	interleukin 11 receptor; alpha isoform 1 precursor	FALSE
108	GABRR1	cg00333528	6	12	**12 (30.8%)**	6	gamma-aminobutyric acid (GABA) receptor; rho 1	FALSE
109	KCNMB2	cg08812936	7	12	**12 (30.8%)**	3	calcium-activated potassium channel beta 2 subunit	FALSE
110	SLC2A14	cg05521696	7	12	**12 (30.8%)**	12	glucose transporter 14	TRUE
111	SCARA3	cg13351161	3	12	**12 (30.8%)**	8	scavenger receptor class A; member 3 isoform 2	TRUE
112	UGT3A1	cg23317501	5	12	**12 (30.8%)**	5	UDP glycosyltransferase 3 family; polypeptide A1	FALSE
113	TNFRSF10D	cg11947493	6	12	**12 (30.8%)**	8	tumor necrosis factor receptor superfamily; member 10d precursor	TRUE
114	HOXD9	cg10957151	6	12	**12 (30.8%)**	2	homeobox D9	TRUE
115	KIF5A	cg04270799	10	11	**11 (28.2%)**	12	kinesin family member 5A	TRUE
116	CCK	cg16864658	10	11	**11 (28.2%)**	3	cholecystokinin preproprotein	TRUE
117	EPHA5	cg18420965	9	11	**11 (28.2%)**	4	ephrin receptor EphA5 isoform b	TRUE
118	SUMO3	cg21053323	9	11	**11 (28.2%)**	21	small ubiquitin-like modifier protein 3	TRUE
119	EDNRB	cg23316360	8	11	**11 (28.2%)**	13	endothelin receptor type B isoform 2	TRUE
120	CDO1	cg12880658	8	11	**11 (28.2%)**	5	cysteine dioxygenase; type I	TRUE
121	CNTNAP2	cg16254309	8	11	**11 (28.2%)**	7	cell recognition molecule Caspr2 precursor	TRUE
122	WDR69	cg21577049	11	10	**11 (28.2%)**	2	hypothetical protein LOC164781	TRUE
123	CCND2	cg12382902	11	10	**11 (28.2%)**	12	cyclin D2	TRUE
124	HOXD9	cg14991487	11	11	**11 (28.2%)**	2	homeobox D9	TRUE
125	FAM57A	cg25186143	11	11	**11 (28.2%)**	17	family with sequence similarity 57; member A	TRUE
126	RAB37	cg12448933	11	11	**11 (28.2%)**	17	RAB37; member RAS oncogene family isoform 1	TRUE
127	RASSF2	cg19614321	11	8	**11 (28.2%)**	20	Ras association domain family 2 isoform 2	FALSE
128	CD34	cg19591881	11	10	**11 (28.2%)**	1	CD34 antigen isoform b	FALSE
129	RECK	cg12717594	11	10	**11 (28.2%)**	9	RECK protein precursor	TRUE
130	FLI1	cg17872757	11	8	**11 (28.2%)**	11	Friend leukemia virus integration 1	TRUE
131	PLAU	cg01078276	11	10	**11 (28.2%)**	10	urokinase plasminogen activator preproprotein	TRUE
132	SCRL	cg01857260	11	11	**11 (28.2%)**	19	hypothetical protein LOC126123	TRUE
133	CBR1	cg00695416	11	7	**11 (28.2%)**	21	carbonyl reductase 1	TRUE
134	TMEM74	cg15042080	11	9	**11 (28.2%)**	8	hypothetical protein LOC157753	TRUE
135	ALPL	cg20645065	11	11	**11 (28.2%)**	1	tissue non-specific alkaline phosphatase precursor	TRUE
136	TMEFF1	cg22775000	11	9	**11 (28.2%)**	9	transmembrane protein with EGF-like and two follistatin-like domains 1	TRUE
137	ZNF532	cg12406559	11	11	**11 (28.2%)**	18	zinc finger protein 532	TRUE
138	SCN2B	cg13348944	2	11	**11 (28.2%)**	11	sodium channel; voltage-gated; type II; beta	FALSE
139	CNN1	cg12854483	7	11	**11 (28.2%)**	19	calponin 1; basic; smooth muscle	TRUE
140	ISL1	cg21410991	3	11	**11 (28.2%)**	5	islet-1	TRUE
141	C6orf145	cg24549507	5	11	**11 (28.2%)**	6	hypothetical protein LOC221749	TRUE
142	NOS1	cg03538436	4	11	**11 (28.2%)**	12	nitric oxide synthase 1 (neuronal)	TRUE
143	HDAC9	cg08285151	6	11	**11 (28.2%)**	7	histone deacetylase 9 isoform 1	FALSE
144	EYA4	cg20330472	5	11	**11 (28.2%)**	6	eyes absent 4 isoform c	TRUE
145	ICAM2	cg10275770	5	11	**11 (28.2%)**	17	intercellular adhesion molecule 2 precursor	FALSE
146	CORIN	cg26018901	4	11	**11 (28.2%)**	4	corin	TRUE
147	TBX20	cg02008154	1	11	**11 (28.2%)**	7	T-box transcription factor TBX20	TRUE
148	WRN	cg10709021	5	11	**11 (28.2%)**	8	Werner syndrome protein	TRUE
149	PTPRO	cg27196745	9	10	**10 (25.6%)**	12	receptor-type protein tyrosine phosphatase O isoform b precursor	TRUE
150	HSD17B6	cg21922731	9	10	**10 (25.6%)**	12	3-hydroxysteroid epimerase	FALSE

	SYMBOL	Infin. Probe	Dif 0.4	β 0.5	Meth+	CHR	PRODUCT	CPG_ISLAND
151	CLEC11A	cg13152535	8	10	**10 (25.6%)**	19	stem cell growth factor precursor	TRUE
152	KCNA6	cg26162582	8	10	**10 (25.6%)**	12	potassium voltage-gated channel; shaker-related subfamily; member 6	TRUE
153	C21orf56	cg07747299	8	10	**10 (25.6%)**	21	hypothetical protein LOC84221	TRUE
154	SCARF2	cg14785479	8	10	**10 (25.6%)**	22	scavenger receptor class F; member 2 isoform 2	TRUE
155	TCERG1L	cg03943081	10	10	**10 (25.6%)**	10	transcription elongation regulator 1-like	TRUE
156	PAQR9	cg00970325	10	3	**10 (25.6%)**	3	progestin and adipoQ receptor family member IX	TRUE
157	CCND2	cg16994506	10	9	**10 (25.6%)**	12	cyclin D2	TRUE
158	PDPN	cg18877506	10	10	**10 (25.6%)**	1	lung type-I cell membrane-associated glycoprotein isoform a	TRUE
159	SIPA1	cg25361844	10	10	**10 (25.6%)**	11	signal-induced proliferation-associated protein 1	TRUE
160	LRFN5	cg04784672	10	7	**10 (25.6%)**	14	leucine rich repeat and fibronectin type III domain containing 5	TRUE
161	KCNA1	cg26590537	10	6	**10 (25.6%)**	12	potassium voltage-gated channel; shaker-related subfamily; member 1	TRUE
162	HOXB13	cg21842478	10	10	**10 (25.6%)**	17	homeo box B13	TRUE
163	LOC389458	cg01870826	10	10	**10 (25.6%)**	7	hypothetical protein LOC389458	TRUE
164	RBP1	cg23363832	10	10	**10 (25.6%)**	3	retinol binding protein 1; cellular	TRUE
165	PRKCQ	cg11638200	10	4	**10 (25.6%)**	10	protein kinase C; theta	TRUE
166	CDKL2	cg14988503	10	8	**10 (25.6%)**	4	cyclin-dependent kinase-like 2	TRUE
167	GYPC	cg13901526	10	6	**10 (25.6%)**	2	glycophorin C isoform 1	TRUE
168	KIAA1944	cg13234863	10	10	**10 (25.6%)**	12	hypothetical protein LOC121256	TRUE
169	TF	cg24879335	10	10	**10 (25.6%)**	3	transferrin	TRUE
170	AEBP1	cg02126753	10	8	**10 (25.6%)**	7	adipocyte enhancer binding protein 1 precursor	TRUE
171	FOXL2	cg17503456	10	10	**10 (25.6%)**	3	forkhead box L2	TRUE
172	GSTP1	cg04920951	10	9	**10 (25.6%)**	11	glutathione transferase	TRUE
173	APC	cg16970232	10	7	**10 (25.6%)**	5	adenomatosis polyposis coli	TRUE
174	GSTM2	cg03070194	10	6	**10 (25.6%)**	1	glutathione S-transferase M2	FALSE
175	CARD11	cg26937500	7	10	**10 (25.6%)**	7	caspase recruitment domain family; member 11	TRUE
176	NPY	cg12614105	6	10	**10 (25.6%)**	7	neuropeptide Y	TRUE
177	CDX2	cg02055963	5	10	**10 (25.6%)**	13	caudal type homeo box transcription factor 2	TRUE
178	HHIP	cg13749822	6	10	**10 (25.6%)**	4	hedgehog-interacting protein	TRUE
179	NEUROD1	cg22359606	7	10	**10 (25.6%)**	2	neurogenic differentiation 1	TRUE
180	LRAT	cg23587449	6	10	**10 (25.6%)**	4	lecithin retinol acyltransferase	TRUE
181	PRAP1	cg03743584	5	10	**10 (25.6%)**	10	proline-rich acidic protein 1	FALSE
182	CYB5R1	cg18275051	7	10	**10 (25.6%)**	1	NAD(P)H:quinone oxidoreductase type 3; polypeptide A2	TRUE
183	RELN	cg17923358	6	10	**10 (25.6%)**	7	reelin isoform a	TRUE
184	OLFM2	cg00208967	2	10	**10 (25.6%)**	19	olfactomedin 2	TRUE
185	GPR63	cg26174752	6	10	**10 (25.6%)**	6	G protein-coupled receptor 63	TRUE
186	COL5A2	cg22774472	8	9	**9 (23.1%)**	2	alpha 2 type V collagen preproprotein	FALSE
187	GNG4	cg02780849	8	9	**9 (23.1%)**	1	guanine nucleotide binding protein (G protein); gamma 4	TRUE
188	LIN28	cg16530429	8	9	**9 (23.1%)**	1	lin-28 homolog	TRUE
189	PTPRO	cg09126273	8	9	**9 (23.1%)**	12	receptor-type protein tyrosine phosphatase O isoform d precursor	TRUE
190	PHACTR3	cg20357628	8	9	**9 (23.1%)**	20	phosphatase and actin regulator 3 isoform 2	TRUE
191	EPHA10	cg07447922	9	9	**9 (23.1%)**	1	EPH receptor A10 isoform 2	TRUE
192	ZNF560	cg04062391	9	6	**9 (23.1%)**	19	zinc finger protein 560	TRUE
193	SLC2A5	cg24480859	9	5	**9 (23.1%)**	1	solute carrier family 2 (facilitated glucose/fructose transporter); member 5	FALSE
194	BACE2	cg16334795	9	9	**9 (23.1%)**	21	beta-site APP-cleaving enzyme 2 isoform A preproprotein	FALSE
195	MRPS21	cg05191071	9	9	**9 (23.1%)**	1	mitochondrial ribosomal protein S21	TRUE
196	GRIA1	cg17020834	9	7	**9 (23.1%)**	5	glutamate receptor; ionotropic; AMPA 1	FALSE
197	CCND2	cg02765328	9	4	**9 (23.1%)**	12	cyclin D2	TRUE
198	VIPR2	cg18349835	9	6	**9 (23.1%)**	7	vasoactive intestinal peptide receptor 2	TRUE
199	GRIA4	cg19343464	9	8	**9 (23.1%)**	11	glutamate receptor; ionotrophic	TRUE
200	SFRP1	cg13398291	9	8	**9 (23.1%)**	8	secreted frizzled-related protein 1	TRUE
201	LAMA2	cg20640433	9	9	**9 (23.1%)**	6	laminin alpha 2 subunit precursor	TRUE
202	NEF3	cg18267374	9	9	**9 (23.1%)**	8	neurofilament 3 (150kDa medium)	TRUE
203	KCNC3	cg17838026	9	6	**9 (23.1%)**	19	Shaw-related voltage-gated potassium channel protein 3	TRUE
204	LMOD1	cg09636671	9	5	**9 (23.1%)**	1	leiomodin 1 (smooth muscle)	TRUE
205	FZD10	cg23054883	9	8	**9 (23.1%)**	12	frizzled 10	TRUE
206	HAND1	cg00405677	9	5	**9 (23.1%)**	5	basic helix-loop-helix transcription factor HAND1	FALSE
207	CCND2	cg13801381	9	7	**9 (23.1%)**	12	cyclin D2	TRUE
208	ACADL	cg09068528	9	8	**9 (23.1%)**	2	acyl-Coenzyme A dehydrogenase; long chain precursor	TRUE
209	RYR2	cg11657808	9	9	**9 (23.1%)**	1	ryanodine receptor 2	TRUE
210	KIT	cg27154163	9	5	**9 (23.1%)**	4	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog precursor	TRUE
211	SEMA3B	cg22037648	9	6	**9 (23.1%)**	3	semaphorin 3B isoform 2 precursor	TRUE
212	POU4F3	cg18482268	9	7	**9 (23.1%)**	5	POU domain; class 4; transcription factor 3	TRUE
213	FOXL2	cg14312526	9	9	**9 (23.1%)**	3	forkhead box L2	TRUE
214	HOXB4	cg14458834	9	8	**9 (23.1%)**	17	homeo box B4	TRUE
215	C6orf206	cg04600618	7	9	**9 (23.1%)**	6	hypothetical protein LOC221421	TRUE
216	SOX14	cg16428251	4	9	**9 (23.1%)**	3	SRY-box 14	TRUE
217	RAB11FIP4	cg10149836	3	9	**9 (23.1%)**	17	RAB11 family interacting protein 4 (class II)	TRUE
218	ZDHHC3	cg25766774	5	9	**9 (23.1%)**	3	DHHC1 protein	FALSE
219	CX36	cg21053529	6	9	**9 (23.1%)**	15	connexin-36	TRUE
220	LYPD5	cg12768605	5	9	**9 (23.1%)**	19	LY6/PLAUR domain containing 5	TRUE
221	PTDSS1	cg16177163	2	9	**9 (23.1%)**	8	phosphatidylserine synthase 1	FALSE
222	C6orf188	cg23741330	5	9	**9 (23.1%)**	6	hypothetical protein LOC254228	FALSE
223	PHACTR3	cg20674577	4	9	**9 (23.1%)**	20	phosphatase and actin regulator 3 isoform 2	TRUE
224	HCA112	cg26385222	6	9	**9 (23.1%)**	7	hepatocellular carcinoma-associated antigen 112	TRUE
225	ZNF659	cg18267381	5	9	**9 (23.1%)**	3	zinc finger protein 659	TRUE

	SYMBOL	Infin. Probe	Dif 0.4	β 0.5	Meth+	CHR	PRODUCT	CPG_ISLAND
226	FOXA2	cg07572341	4	9	9 (23.1%)	20	forkhead box A2	TRUE
227	MYO3A	cg23771603	6	9	9 (23.1%)	10	myosin IIIA	TRUE
228	CENTG2	cg04856858	6	9	9 (23.1%)	2	centaurin; gamma 2	TRUE
229	POU3F1	cg21243096	2	9	9 (23.1%)	1	POU domain; class 3; transcription factor 1	TRUE
230	LDHD	cg03991512	4	9	9 (23.1%)	16	D-lactate dehydrogenase isoform 1 precursor	FALSE
231	GRID2	cg13549845	4	9	9 (23.1%)	4	glutamate receptor; ionotropic; delta 2	TRUE
232	DCC	cg01839464	4	9	9 (23.1%)	18	deleted in colorectal carcinoma	TRUE
233	THSD3	cg07965823	7	9	9 (23.1%)	14	thrombospondin; type I domain containing 3 isoform 3	TRUE
234	IGFBP7	cg03876618	6	9	9 (23.1%)	4	insulin-like growth factor binding protein 7	TRUE
235	KLF11	cg02983451	8	8	8 (20.5%)	2	Kruppel-like factor 11	TRUE
236	GHSR	cg11812218	8	8	8 (20.5%)	3	growth hormone secretagogue receptor isoform 1a	TRUE
237	KCNC4	cg05675373	8	5	8 (20.5%)	1	Shaw-related voltage-gated potassium channel protein 4 isoform a	TRUE
238	SCN3B	cg15457899	8	7	8 (20.5%)	11	voltage-gated sodium channel beta-3 subunit (scn3b gene)	TRUE
239	HIF3A	cg02879662	8	8	8 (20.5%)	19	hypoxia-inducible factor-3 alpha isoform b	TRUE
240	RUNX3	cg06377278	8	6	8 (20.5%)	1	runt-related transcription factor 3 isoform 2	TRUE
241	CMTM3	cg16335762	8	5	8 (20.5%)	16	chemokine-like factor superfamily 3 isoform a	TRUE
242	CDKL2	cg24432073	8	8	8 (20.5%)	4	cyclin-dependent kinase-like 2	TRUE
243	HDAC10	cg01120165	8	6	8 (20.5%)	22	histone deacetylase 10	TRUE
244	PLTP	cg26220350	8	6	8 (20.5%)	20	phospholipid transfer protein isoform b precursor	TRUE
245	DFNA5	cg24805239	8	5	8 (20.5%)	7	deafness; autosomal dominant 5 protein	TRUE
246	SOCS2	cg04797323	8	8	8 (20.5%)	12	suppressor of cytokine signaling-2	TRUE
247	BDNF	cg27351358	8	7	8 (20.5%)	11	brain-derived neurotrophic factor isoform a preproprotein	TRUE
248	CLDN11	cg20924286	8	7	8 (20.5%)	3	claudin 11	TRUE
249	CPNE8	cg23495733	8	8	8 (20.5%)	12	copine VIII	TRUE
250	PODN	cg23092823	8	8	8 (20.5%)	1	podocan	TRUE
251	KIRREL2	cg09479015	8	5	8 (20.5%)	19	kin of IRRE-like 2 isoform c	TRUE
252	HOXC8	cg25917510	8	7	8 (20.5%)	12	homeobox C8	TRUE
253	KCNH7	cg19965810	8	7	8 (20.5%)	2	potassium voltage-gated channel; subfamily H; member 7 isoform 1	TRUE
254	REM1	cg26299767	8	7	8 (20.5%)	20	RAS-like GTP-binding protein REM	TRUE
255	APC	cg20311501	8	7	8 (20.5%)	5	adenomatosis polyposis coli	TRUE
256	HDAC9	cg12081743	8	6	8 (20.5%)	7	histone deacetylase 9 isoform 1	FALSE
257	AKAP2	cg19176447	8	4	8 (20.5%)	9	A kinase (PRKA) anchor protein 2	TRUE
258	GABRG2	cg09220361	8	7	8 (20.5%)	5	gamma-aminobutyric acid A receptor; gamma 2 isoform 2 precursor	FALSE
259	PAX7	cg11428724	8	8	8 (20.5%)	1	paired box gene 7 isoform 1	TRUE
260	KLF3	cg09963123	8	8	8 (20.5%)	4	Kruppel-like factor 3 (basic)	TRUE
261	BTG4	cg22879515	8	7	8 (20.5%)	11	B-cell translocation gene 4	TRUE
262	HTR1B	cg25763788	3	8	8 (20.5%)	6	5-hydroxytryptamine (serotonin) receptor 1B	TRUE
263	HIST1H3J	cg17965019	3	8	8 (20.5%)	6	H3 histone family; member J	TRUE
264	ZNF96	cg07660236	4	8	8 (20.5%)	6	zinc finger protein 96	TRUE
265	GJA5	cg08307963	4	8	8 (20.5%)	1	gap junction protein; alpha 5	FALSE
266	DCC	cg02624705	6	8	8 (20.5%)	18	deleted in colorectal carcinoma	TRUE
267	CA3	cg18674980	7	8	8 (20.5%)	8	carbonic anhydrase III	TRUE
268	PYGM	cg26884581	3	8	8 (20.5%)	11	glycogen phosphorylase	TRUE
269	STAT5A	cg03001305	4	8	8 (20.5%)	17	signal transducer and activator of transcription 5A	FALSE
270	PKIA	cg04689061	7	8	8 (20.5%)	8	cAMP-dependent protein kinase inhibitor alpha isoform 7	TRUE
271	RP11-49G10.8	cg01775265	7	8	8 (20.5%)	20	breast cancer and salivary gland expression gene	FALSE
272	FOXL1	cg06995715	6	8	8 (20.5%)	16	forkhead box L1	TRUE
273	ACADL	cg14795968	0	8	8 (20.5%)	2	acyl-Coenzyme A dehydrogenase; long chain precursor	TRUE
274	TBX21	cg26607785	5	8	8 (20.5%)	17	T-box 21	TRUE
275	DISP2	cg06595693	3	8	8 (20.5%)	15	dispatched B	TRUE
276	HOXD11	cg16632715	1	8	8 (20.5%)	2	homeobox D11	TRUE
277	TLX2	cg00234616	7	8	8 (20.5%)	2	T-cell leukemia; homeobox 2 isoform 2	TRUE
278	WBSCR17	cg01366419	4	8	8 (20.5%)	7	UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase-like 3	TRUE
279	FLJ37478	cg21269934	5	8	8 (20.5%)	4	hypothetical protein LOC339983	TRUE
280	MYADM	cg19717326	3	8	8 (20.5%)	19	myeloid-associated differentiation marker	FALSE
281	COL1A2	cg18511007	6	8	8 (20.5%)	7	alpha 2 type I collagen	TRUE
282	GABBR2	cg07903918	7	8	8 (20.5%)	9	G protein-coupled receptor 51	TRUE
283	PCDHGA12	cg07730329	0	8	8 (20.5%)	5	protocadherin gamma subfamily A; 12 isoform 1 precursor	TRUE
284	CHL1	cg00903242	7	8	8 (20.5%)	3	cell adhesion molecule with homology to L1CAM precursor	TRUE
285	NCAM2	cg13297960	7	8	8 (20.5%)	21	neural cell adhesion molecule 2 precursor	TRUE
286	C1orf188	cg15731815	4	8	8 (20.5%)	1	hypothetical protein LOC148646	TRUE
287	FAM89A	cg16516400	7	8	8 (20.5%)	1	hypothetical protein LOC375061	FALSE
288	ZNF577	cg22472290	3	8	8 (20.5%)	19	zinc finger protein 577	TRUE
289	KLK10	cg06130787	5	8	8 (20.5%)	19	kallikrein 10 precursor	FALSE
290	ERG	cg03127334	7	8	8 (20.5%)	21	v-ets erythroblastosis virus E26 oncogene like isoform 2	TRUE
291	AOC3	cg21602160	2	8	8 (20.5%)	17	amine oxidase; copper containing 3 precursor	FALSE

Appendix C.III_List of 291 methylated probes from the Illumina Infinium study in chapter 4.

The table shows all 291 probes that were classified as hypermethylated from the Illumina Infinium study (chapter 3). In each case, gene symbol, probe name, the number of samples meeting the criteria of having a difference value of 0.4 and a β-values of 0.5 are shown along with the number of samples meeting either of these criteria (Meth+). In addition, the chromosome number is shown and whether the probe is within a true or false CpG island.

GENE	PRIMER	PRIMER SEQUENCE (5'-3')	ANNEALING TEMPERATURE (°C)		PROGRAM	PRODUCT SIZE (bp)	
			PRIMARY REACTION	SECONDARY REACTION		PRIMARY REACTION	SECONDARY REACTION
RASSF10	F	TTGTTTTTGTTGTTTTYGTYGTTTTAGTAGATT	55	56	TOUCHDOWN (PRIMARY) STANDARD (SECONDARY)	638	426
	FN	GTGTGGATTTGTTAGGAAGAGAAGT					
	RN	CTATTTCTCCTAAATCATAACCAAACTAA					
	R	CRATTAAACTTATCCAATTTACAAAAACCTTA					

Appendix D.I_RASSF10 CoBRA primer sequences

The above table shows the primer sequences used for *RASSF10* CoBRA analysis. Annealing temperatures, program details and product sizes are also given.

GENE	PRIMER		PRIMER SEQUENCE (5'-3')	ANNEALING TEMPERATURE (°C)	PROGRAM	PRODUCT SIZE (bp)
RASSF1A	MSP	F	CGAGAGCGCGTTTAGTTTCGTT	58	STANDARD	192
		R	CGATTAAACCCGTACTTCGCTAA			
	USP	F	GGGGGTTTTGTGAGAGTGTGTTT	58		204
		R	CCCAATTAAACCCATACTTCCTAA			
MGMT	MSP	F	TTTCGACGTTCGTAGGTTTTCGC	58	STANDARD	81
		R	GCACTCTTCCGAAAACGAAACG			
	USP	F	TTTGTGTTTTGATGTTTGTAGGTTTTTGT	58		93
		R	AACTCCACACTCTTCCAAAAACAAAACA			

Appendix D.II_*RASSF1A* and *MGMT* MSP primer sequences

The above table shows the primer sequences used for *RASSF1A* and *MGMT* MSP analysis. Pirmer sequences for *RASSF1A* and MGMT were taken from Honorio *et al*, 2003 and Esteller *et al*, 2000a respectively. Annealing temperatures, program details and product sizes are also given.

GENE	PRIMER	SEQUENCE (5'-3')	ANNEALING TEMPERATURE (°C)	PRODUCT SIZE (bp)
RASSF10	F	GTTCAGCAGGAGGAGTTGCT	56	317
	R	CGTCAGCTCCAAAGTGTGCAA		

Appendix D.III_Expression primers used for *RASSF10* expression analysis
Forward (F) and reverse (R) primer sequences are shown for expression analysis of *RASSF10*. Annealing temperature and product size are also shown.

GENE	PRIMER	PRIMER SEQUENCE (5'-3')	ANNEALING TEMPERATURE (°C)		PROGRAM	PRODUCT SIZE (bp)	
			PRIMARY REACTION	SECONDARY REACTION		PRIMARY REACTION	SECONDARY REACTION
KIBRA	F	GTATTTGGYGGAGGTAGAAGTTATAAATT	55	55	TOUCH	638	426
	RN	CATAATCCRAAAAATAACRCCCRCAAATAA					
	R	TATCCCRCRAATCRACCAACTAATAATA					

Appendix E.I_*KIBRA* CoBRA primer sequences

The above table shows the primer sequences used for *KIBRA* CoBRA analysis. Annealing temperatures, program details and product sizes are also given.

GENE	PRIMER	SEQUENCE (5'-3')	ANNEALING TEMPERATURE (°C)	PRODUCT SIZE (bp)
KIBRA	F	AAACAGAGCAGGGAGCTCAA	56	263
	R	CCCATCCATATCAGGTGAGG		

Appendix E.II_Expression primers used for *KIBRA* expression analysis
Forward (F) and reverse (R) primer sequences are shown for expression analysis of *KIBRA*. Annealing temperature and product size are also shown.

Abdel-Wahab O, Mullally A, Hedvat C, Garcia-Manero G, Patel J, Wadleigh M, Malinge S, Yao J, Kilpivaara O, Bhat R, Huberman K, Thomas S, Dolgalev I, Heguy A, Paietta E, Le Beau MM, Beran M, Tallman MS, Ebert BL, Kantarjian HM, Stone RM, Gilliland DG, Crispino JD, Levine RL (2009) Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. *Blood* **114**(1):144-147.

Adams H, Fritzsche FR, Dirnhofer S, Kristiansen G, Tzankov A (2010) Class I histone deacetylases 1, 2 and 3 are highly expressed in classical Hodgkin's lymphoma.*Expert Opinion on Therapeutic Targets* **14**(6):577-584.

Adamson C, Kanu OO, Mehta AI, Di C, Lin N, Mattox AK, Bigner DD (2009) Glioblastoma multiforme: a review of where we have been and where we are going. *Expert Opinion on Investigational Drugs* **18**(8):1061-1083

Agathanggelou A, Bièche I, Ahmed-Choudhury J, Nicke B, Dammann R, Baksh S, Gao B, Minna JD, Downward J, Maher ER, Latif F (2003) Identification of novel gene expression targets for the Ras association domain family 1 (RASSF1A) tumor suppressor gene in non-small cell lung cancer and neuroblastoma. *Cancer Research* **62**(17):5344-5351

Agathanggelou A, Cooper WN, Latif F (2005) Role of the Ras-association domain family 1 tumor suppressor gene in human cancers. *Cancer Research* **65**(9):3497-3508.

Agell L, Hernández S, Salido M, de Muga S, Juanpere N, Arumí-Uria M, Menendez S, Lorenzo M, Lorente JA, Serrano S, Lloreta J (2011) PI3K signaling pathway is activated by PIK3CA mRNA overexpression and copy gain in prostate tumors, but PIK3CA, BRAF, KRAS and AKT1 mutations are infrequent events. *Modern pathology* **24**(3): 443-52.

Ahmed M and Nahman N (2006) *ATM* and breast cancer susceptibility. *Oncogene* **25**:5906-5911.

Akino K, Toyota M, Suzuki H, Mita H, Sasaki Y, Ohe-Toyota M, Issa J-P, Hinoda Y, Imai K, Tokino T (2005) The Ras effector RASSF2 is a novel tumor-suppressor gene in human colorectal cancer. *Gastroenterology* **129**:156-169

Alvarez S, Suela J, Valencia A, Fernández A, Wunderlich M, Agirre X, Prósper F, Martín-Subero JI, Maiques A, Acquadro F, Rodriguez Perales S, Calasanz MJ, Roman-Gómez J, Siebert R, Mulloy JC, Cervera J, Sanz MA, Esteller M, Cigudosa JC (2010) DNA methylation profiles and their relationship with cytogenetic status in adult acute myeloid leukemia. *PLoS One* **5**(8):e12197.

Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genetics* **2**:185-8

Anderton JA, Lindsey JC, Lusher ME, Gilbertstone RJ, Bailey S, Ellison DW and Clifford SC (2008) Global analysis of the medulloblastoma epigenome identifies disease-subgroup-specific inactivation of *COL1A2*. *Neuro-Oncology* **10**(6):981-994.

Antoniou AC, Sinilnikova OM, Simard J, Léoné M, Dumont M, Neuhausen SL, Struewing JP, Stoppa-Lyonnet D, Barjhoux L, Hughes DJ, Coupier I, Belotti M, Lasset C, Bonadona V, Bignon YJ; Genetic Modifiers of Cancer Risk in BRCA1/2 Mutation Carriers Study (GEMO), Rebbeck TR, Wagner T, Lynch HT, Domchek SM, Nathanson KL, Garber JE, Weitzel J, Narod SA, Tomlinson G, Olopade OI, Godwin A, Isaacs C, Jakubowska A, Lubinski J, Gronwald J, Górski B, Byrski T, Huzarski T, Peock S, Cook M, Baynes C, Murray A, Rogers M, Daly PA, Dorkins H; Epidemiological Study of BRCA1 and BRCA2 Mutation Carriers (EMBRACE), Schmutzler RK, Versmold B, Engel C, Meindl A, Arnold N, Niederacher D, Deissler H; German Consortium for Hereditary Breast and Ovarian Cancer (GCHBOC), Spurdle AB, Chen X, Waddell N, Cloonan N; Kathleen Cuningham Consortium for Research into Familial Breast Cancer (kConFab), Kirchhoff T, Offit K, Friedman E, Kaufmann B, Laitman Y, Galore G, Rennert G, Lejbkowicz F, Raskin L, Andrulis IL, Ilyushik E, Ozcelik H, Devilee P, Vreeswijk MP, Greene MH, Prindiville SA, Osorio A, Benitez J, Zikan M, Szabo CI, Kilpivaara O, Nevanlinna H, Hamann U, Durocher F, Arason A, Couch FJ, Easton DF, Chenevix-Trench G; Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA). RAD51 135G-->C modifies breast cancer risk among BRCA2 mutation carriers: results from a combined analysis of 19 studies. *American Journal of Human Genetics* **81**(6):1186-1200.

Argos M, Kibriya MG, Jasmine F, Olopade OI, Su T, Hibshoosh H, Ahsan HV (2008) Genomewide scan for loss of heterozygosity and chromosomal amplification in breast carcinoma using single-nucleotide polymorphism arrays. *Cancer Genetics and Cytogenetics* **182**(2):69-74.

Argraves WS, Greene LM, Cooley MA, Gallagher WM (2003) Fibulins: physiological and disease properties. *EMBO reports* **4**(12):1127-1131.

Armstrong AS and Look AT (2005) Molecular Genetics of Acute Lymphoblastic Leukaemia. *Journal of Clinical Oncology* **23(**26):6306-*6315.*

Arney KL, Fisher AG (2004) Epigenetic aspects of differentiation. *Journal of Cell Science* **117**(Pt 19):4355-4363

Baeza N, Weller M, Yonekawa Y, Kleihues P, Ohgaki H (2003) PTEN methylation and expression in glioblastomas. *Acta neuropathologica* **106**(5):479-485.

Bauerschlag DO, Ammerpohl O, Bräutigam K, Schem C, Lin Q, Weigel MT, Hilpert F, Arnold N, Maass N, Meinhold-Heerlein I, Wagner W (2011) Progression-Free Survival in Ovarian Cancer Is Reflected in Epigenetic DNA Methylation Profiles. *Oncology* **80**(1-2):12-20.

Baumgartner R, Poernbacher I, Busser N, Hafen E and Stocker H (2010) The WW Domain Protein Kibra Acts Upstream of Hippo in Drosophila. *Developmental Cell* **18**:309-316

Basu S, Totty NF, Irwin MS, Sudol M, Downward J (2003) Akt phosphorylates the Yes-associated protein, YAP, to induce interaction with 14-3-3 and attenuation of p73-mediated apoptosis. *Molecular Cell* **11**(1):11-23.

Baylin SB, Höppener JW, de Bustros A, Steenbergh PH, Lips CJ, Nelkin BD (1986) DNA methylation patterns of the calcitonin gene in human lung cancers and lymphomas. *Cancer Research* **46**(6):2917-2922.

Bediaga NG, Acha-Sagredo A, Guerra I, Viguri A, Abaina C, Ruiz Diaz I, Rexola R, Jesus Alberdi M, Dopazo J, Montaner D, de Renobales M, Fernández AF, Field JK, Fraga MF, Liloglou T and de Pancorbo MM (2010). DNA methylation epigenotypes in breast cancer molecular subtypes. *Breast Cancer Research* **12**:R77.

Belinsky MG, Skorobogatko YV, Rink L, Pei J, Cai KQ, Vanderveer LA, Riddell D, Merkel E, Tarn C, Eisenberg BL, von Mehren M, Testa JR, Godwin AK (2009) High density DNA array analysis reveals distinct genomic profiles in a subset of gastrointestinal stromal tumors. *Genes chromosomes and cancer* **48**(10):886-96.

Bell AC, Felsenfeld G (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**(6785):482-5.

Bell DW, Varley JM, Szydlo TE, Kang DH, Wahrer DC, Shannon KE, Lubratovich M, Verselis SJ, Isselbacher KJ, Fraumeni JF, Birch JM, Li FP, Garber JE, Haber DA (1999) Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome. *Science* **286**(5449):2528-2531.

Bell DW (2010) Our changing view of the genomic landscape of cancer. *Journal of Pathology* **220**(2):231-243.

Berger R, Fiegl H, Goebel G, Obexer P, Ausserlechner M, Doppler W, Hauser-Kronberger C, Reitsamer R, Egle D, Reimer D, Müller-Holzner E, Jones A, Widschwendter M (2010) Toll-like receptor 9 expression in breast and ovarian cancer is associated with poorly differentiated tumors. *Cancer Science* **101**(4):1059-1066.

Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko AY, Sboner A, Esgueva R, Pflueger D, Sougnez C, Onofrio R, Carter SL, Park K, Habegger L, Ambrogio L, Fennell T, Parkin M, Saksena G, Voet D, Ramos AH, Pugh TJ, Wilkinson J, Fisher S, Winckler W, Mahan S, Ardlie K, Baldwin J, Simons JW, Kitabayashi N, MacDonald TY, Kantoff PW, Chin L, Gabriel SB, Gerstein MB, Golub TR, Meyerson M, Tewari A, Lander ES, Getz G, Rubin MA, Garraway LA (2011) The genomic complexity of primary human prostate cancer. *Nature* **470**(7333):214-220.

Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, Mc Henry KT, Pinchback RM, Ligon AH, Cho YJ, Haery L, Greulich H, Reich M, Winckler W, Lawrence MS, Weir BA, Tanaka KE, Chiang DY, Bass AJ, Loo A, Hoffman C, Prensner J, Liefeld T, Gao Q, Yecies D, Signoretti S, Maher E, Kaye FJ, Sasaki H, Tepper JE, Fletcher JA, Tabernero J, Baselga J, Tsao MS, Demichelis F, Rubin MA, Janne PA, Daly MJ, Nucera C, Levine RL, Ebert BL, Gabriel S, Rustgi AK, Antonescu CR, Ladanyi M, Letai A, Garraway LA, Loda M, Beer DG, True LD, Okamoto A, Pomeroy SL, Singer S, Golub TR, Lander ES, Getz G, Sellers WR, Meyerson M (2010) The landscape of somatic copy-number alteration across human cancers. *Nature*.**463** (7283):899-905.

Berx G, Nollet F, Strumane K, van Roy F (1997) An efficient and reliable multiplex PCR-SSCP mutation analysis test applied to the human E-cadherin gene. *Human Mutation* **9**(6):567-574.

Bestor TH (1992) Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *EMBO Journal* **11:** 2611-2617.

Bird AP, Wolffe AP (1999) Methylation-induced repression-belts, braces, and chromatin. *Cell* **99**(5):451-454.

Bird A (2002) DNA methylation patterns and epigenetic memory. *Genes and Development* **16**(1):6-21.

Bonifaci N, Górski B, Masojć B, Wokolorczyk D, Jakubowska A, Dębniak T, Berenguer A, Musach JS, Brunet J, Dopazo J, Narod SA, Lubiński J, Lázaro C, Cybulski C and Pujana MA (2010) Exploring the Link between Germline and Somatic Genetic Alterations in Breast Carcinogenesis. *PLoS ONE* **5**(11):e14078

Bourdon JC. (2007) p53 and its isoforms in cancer. *British Journal of Cancer* **97**(3):277-282

Brandt R, Grützmann R, Bauer A, Jesnowski R, Ringel J, Löhr M, Pilarsky C, Hoheisel JD (2004) DNA microarray analysis of pancreatic malignancies. *Pancreatology* **4**:587-597

Brinton LA, Schairer C, Hoover RN, Fraumeni JF Jr.(1988) Menstrual factors and risk of breast cancer. *Cancer Investigation* **6**(3):245-254.

Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, Lopez E, Kyle S, Meuth M, Curtin NJ, Helleday T (1995). Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**(7035):913-917. Erratum in: *Nature* 2007 **447**(7142):346.

Buckingham L, Penfield Faber L, Kim A, Liptay M, Barger C, Basu S, Fidler M, Walters K, Bonomi P, Coon J (2010) PTEN, RASSF1 and DAPK site-specific hypermethylation and outcome in surgically treated stage I and II nonsmall cell lung cancer patients. *International Journal of Cancer* **126**(7):1630-1639.

Burbee DG, Forgacs E, Zöchbauer-Müller S, Shivakumar L, Fong K, Gao B, Randle D, Kondo M, Virmani A, Bader S, Sekido Y, Latif F, Milchgrub S, Toyooka S, Gazdar AF, Lerman MI, Zabarovsky E, White M, Minna JD (2001) Epigenetic inactivation of RASSF1A in lung and breast cancers and malignant phenotype suppression. *Journal of the National Institute of Cancer* **93**(9):691-699.

Burkhart DL, Sage J (2008) Cellular mechanisms of tumour suppression by the retinoblastoma gene. *Nature Reviews Cancer* **8**(9):671-682.

Campbell IG, Russell SE, Choong DY, Montgomery KG, Ciavarella ML, Hooi CS, Cristiano BE, Pearson RB, Phillips WA (2004) Mutation of the PIK3CA gene in ovarian and breast cancer. *Cancer Research* **64**(21):7678-81.

Cantor S, Drapkin R, Zhang F, Lin Y, Han J, Pamidi S, Livingston DM (2004) The BRCA1-associated protein BACH1 is a DNA helicase targeted by clinically relevant inactivating mutations. *PNAS* **101**(8):2357-2363.

Colombatti A, Doliana R, Bot S, Canton A, Mongiat M, Mungiguerra G, Paron-Cilli S, Spessotto P (2000) The EMILIN protein family. *Matirx Biology* **19**:289-301.

Chahrour M, Jung SY, Shaw C, Zhou X, Wong ST, Qin J, Zoghbi HY (2008). MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* **320** (5880)

Chapman PB, Hauschild A, Robert C, Haanen JB, 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 SJ, Sosman JA, Kirkwood JM, Eggermont AM, Dreno B, Nolop K, Li J, Nelson B, Hou J, Lee RJ, Flaherty KT, McArthur GA; the BRIM-3 Study Group (2011) Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation. *New England Journal of Medicine* **Jun 5** [Epub. ahead of print].

Chan SW, Lim CJ, Guo K, Ng CP, Lee I, Hunziker W, Zeng Q, Hong W (2008) A role for TAZ in migration, invasion, and tumorigenesis of breast cancer cells. *Cancer Research* **68**(8):2592-2598.

Chang H-C, Cho C-Y and Hung W-C (2006) Silencing of the Metastasis Suppressor RECK by RAS Oncogene Is Mediated by DNA Methylatransferase 3b-Induced Promoter Methylation. *Cancer Research* **66(**17):8413-8420.

Chang H-C, Cho C-Y, and Hung W-C (2007) Downregulation of RECK by promoter methylation correlates with lymph node metastasis in non-small cell lung cancer. *Cancer Science* **98**(2):169-173.

Chen J, Pee D, Ayyagari R, Graubard B, Schairer C, Byrne C, Benichou J, Gail MH (2006) Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. *Journal of the National Cancer Institute* **98**(17):1215-1226.

Chen T, Hevi S, Gay F, Tsujimoto N, He T, Zhang B, Ueda Y, Li E (2007) Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. *Nature Genetics* **39**(3):391-396

Cheung HH, Lee TL, Davis AJ, Taft DH, Rennert OM, Chan WY (2010) Genome-wide DNA methylation profiling reveals novel epigenetically regulated genes and non-coding RNAs in human testicular cancer. *British Journal of Cancer* **102**(2):419-27

Chie WC, Hsieh C, Newcomb PA, Longnecker MP, Mittendorf R, Greenberg ER, Clapp RW, Burke KP, Titus-Ernstoff L, Trentham-Dietz A, MacMahon B (2000) Age at any full-term pregnancy and breast cancer risk. *American Journal of Epidemiology* **151**(7):715-722.

Chin SF, Wang Y, Thorne NP, Teschendorff AE, Pinder SE, Vias M, Naderi A, Roberts I, Barbosa-Morais NL, Garcia MJ, Iyer NG, Kranjac T, Robertson JF, Aparicio S, Tavaré S, Eliis I, Brenton JD, Caldas C (2007) Using array-comapritive genomic hybridization to define molecular portraits of primary breast cancers. *Oncogene* **26**:1959-1970.

Chissoe SL, Bodenteich A, Wang Y-F, Wang Y-P, Burian D, Clifton SW, Crabtree J, Freeman A, Iyer K, Jian L, Ma Y, McLaury H-J, Pan H-Q, Sarhan O H, Toth S, Wang Z, Zhang G, Heisterkamp N, Groffen J and Roe BA (1995) Sequence and analysis of the human ABL gene, the BCR gene, and regions involved in the Philadelphia chromosomal translocation. *Genomics* 27(1):67-82.

Cho CY, Wang JH, Chang HC, Chang CK, Hung WC (2007) Epigenetic inactivation of the metastasis suppressor RECK enhances invasion of human colon cancer cells. *Journal of Cellular Physiology* **213**(1):65-69

Cho WC, Chow AS, Au JS (2009) Restoration of tumour suppressor hsa-miR-145 inhibits cancer cell growth in lung adenocarcinoma patients with epidermal growth factor receptor mutation. *European Journal of Cancer* **45**(12):2197-2206.

Chompret A, Brugières L, Ronsin M, Gardes M, Dessarps-Freichey F, Abel A, Hua D, Ligot L, Dondon MG, Bressac-de Paillerets B, Frébourg T, Lemerle J, Bonaïti-Pellié C, Feunteun J (2000) P53 germline mutations in childhood cancers and cancer risk for carrier individuals. *British Journal of Cancer* **82**(12):1932-1937

Christensen BC, Kelsey KT, Zheng S, Houseman EA, Marsit CJ, Wrensch MR, Wiemels JL, Nelson HN, Karagas MR, Kushi LH, Kwan ML and Wiencke JK (2010) Breast Cancer DNA Methylation Profiles Are Assocaited with Tumour Size and Alcohol and Folate Intake. *PLoS Genetics* **6**(7):e1001043.

Christensen BC, Smith AA, Zheng S, Koestler DC, Houseman EA, Marsit CJ, Wiemels JL, Nelson HH, Karagas MR, Wrensch MR, Kelsey KT, Wiencke JK (2011) DNA methylation, isocitrate dehydrogenase mutation, and survival in glioma. *Journal of National Cancer Institute* **103**(2):143-153.

Cho C-Y, Wang J-H, Chang H-C, Chang C-K and Hung W-C (2007) Epigenetic Inactivation of the Metastasis Suppressor RECK Enhances Invasion of Human Colon Cancer Cells. *Journal of Cellular Physiology* **213**:65-69.

Clark JCM, Thomas DM, Choong PFM, Dass CR (2007) RECK-a newly discovered inhibitor of metastasis with prognostic significance in multiple forms of cancer. *Cancer Metastasis Reviews* **26**:675-683.

Clouaire T and Stancheva I (2008) Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin? *Cellular and Molecular Life Sciences* **65**:1509-1522

Cobleigh MA, Tabesh B, Bitterman P, Baker J, Cronin M, Liu ML, Borchik R, Mosquera JM, Walker MG, Shak S (2005) Tumor gene expression and prognosis in breast cancer patients with 10 or more positive lymph nodes. *Clinical Cancer Research* **11**:8623-8631.

Colditz GA (1998) Relationship between estrogen levels, use of hormone replacement therapy, and breast cancer. *Journal of the National Cancer Institute* **90**(11):814-823

Cooper DN, Krwczak M, Antonorakis (2000) The Nature and Mechanisms of human genome mutation. *The Metabolic and Molecular Basis of Inherited Disease*. **Vol 1** 8th edition. (eds. Scriver CR, Beudet AL, Sly WS, Valle D) McGraw-Hill, New York.

Crist W, Carroll A, Shuster J, Jackson J, Head D, Borowitz M, Behm F, Link M, Steuber P, Ragab A, Hirt A, Brock B, Land V, Pullen J (1990) Philadelphia chromosome positive childhood acute lymphoblastic leukemia: clinical and cytogenetic characteristics and treatment outcome. A Pediatric Oncology Group study. *Blood* **76**(3):489-494.

Cross SH, Charlton JA, Nan X, Bird AP (1994) Purification of CpG islands using a methylated DNA binding column. *Nature Genetics* **6**:236-244.

CRUK: www.cancerresearchuk.org

Czekierdowski A, Czekierdowska S, Wielgos M, Smolen A, Kaminski P, Kotarski J (2006) The role of CpG islands hypomethylation and abnormal expression of neuronal protein synuclein-gamma (SNCG) in ovarian cancer. *Neuro Endocrinology Letters* **27**(3):381-386.

Dai Z, Zhu WG, Morrison CD, Brena RM, Smiraglia DJ, Raval A, Wu YZ, Rush LJ, Ross P, Molina JR, Otterson GA, Plass C (2003) A comprehensive search for DNA amplification in lung cancer identifies inhibitors of apoptosis cIAP1 and cIAP2 as candidate oncogenes. *Human Molecular Genetics* **12**(7):791-801

Dalgliesh GL, Furge K, Greenman C, Chen L, Bignell G, Butler A, Davies H, Edkins S, Hardy C, Latimer C, Teague J, Andrews J, Barthorpe S, Beare D, Buck G, Campbell PJ, Forbes S, Jia M, Jones D, Knott H, Kok CY, Lau KW, Leroy C, Lin ML, McBride DJ, Maddison M, Maguire S, McLay K, Menzies A, Mironenko T, Mulderrig L, Mudie L, O'Meara S, Pleasance E, Rajasingham A, Shepherd R, Smith R, Stebbings L, Stephens P, Tang G, Tarpey PS, Turrell K, Dykema KJ, Khoo SK, Petillo D, Wondergem B, Anema J, Kahnoski RJ, Teh BT, Stratton MR, Futreal PA (2010) Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* **463**(7279):360-363.

Dammann R, Li C, Yoon JH, Chin PL, Bates S, Pfeifer GP (2000) Epigenetic inactivation of a RAS association domain family protein from the lung tumour suppressor locus 3p21.3. *Nature Genetics* **25**(3):315-319

Dammann R, Yang G, Pfeifer GP (2001) Hypermethylation of the CpG island of Ras association domain family 1A (RASSF1A), a putative tumour suppressor gene from the 3p21.3 locus, occurs in a large percentage of human breast cancers. *Cancer Research* **61**(7):3105-3109.

Dammann R, Schagdarsurengin U, Liu L, Otto N, Gimm O, Dralle H, Boehm BO, Pfeifer GP, Hoang-Vu C (2003) Frequent RASSF1A promoter hypermethylation and K-ras mutations in pancreatic carcinoma. *Oncogene* **22**(24):3806-3812.

Deininger MW, Druker BJ (2003) Specific targeted therapy of chronic myelogenous leukemia with imatinib *Pharmacological Reviews* **55**(3):401-23.

deVos T, Tetzner R, Model F, Weiss G, Schuster M, Distler J, Steiger KV, Grützmann R, Pilarsky C, Habermann JK, Fleshner PR, Oubre BM, Day R, Sledziewski AZ, Lofton-Day C (2009) Circulating methylated SEPT9 DNA in plasma is a biomarker for colorectal cancer. *Clinical Chemistry* **55**(7):1337-1346.

Deng ZH, Wen JF, Li JH, Xiao DS, Zhou JH (2008) Activator protein-1 involved in growth inhibition by RASSF1A gene in the human gastric carcinoma cell line SGC7901. *World Journal of Gastroenterology* **14**(9):1437-1443

Deng D, Lu Z, Du Y (2010) Epigenetic Alterations as Cancer Diagnostic, Prognostic, and Predictive Biomarkers. *Advances in Genetics* **71**(125-176).

Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, Harris CC, McLellan MD, Fulton RS, Fulton LL, Abbott RM, Hoog J, Dooling DJ, Koboldt DC, Schmidt H, Kalicki J, Zhang Q, Chen L, Lin L, Wendl MC, McMichael JF, Magrini VJ, Cook L, McGrath SD, Vickery TL, Appelbaum E, Deschryver K, Davies S, Guintoli T, Lin L, Crowder R, Tao Y, Snider JE, Smith SM, Dukes AF, Sanderson GE, Pohl CS, Delehaunty KD, Fronick CC, Pape KA, Reed JS, Robinson JS, Hodges JS, Schierding W, Dees ND, Shen D, Locke DP, Wiechert ME, Eldred JM, Peck JB, Oberkfell BJ, Lolofie JT, Du F, Hawkins AE, O'Laughlin MD, Bernard KE, Cunningham M, Elliott G, Mason MD, Thompson DM Jr, Ivanovich JL, Goodfellow PJ, Perou CM, Weinstock GM, Aft R, Watson M, Ley TJ, Wilson RK, Mardis ER (2010) Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464**(7291):999-1005.

Dobrovic A, Simpfendorfer D (1997) Methylation of the BRCA1 gene in sporadic breast cancer. *Cancer Research* **57**(16):3347-3350.

Du Y-Y, Dai D-Q and Yang Z (2010) Role of *RECK* methylation in gastric cancer and its clinical significance. *World Journal of Gastrenerology* **16**(7):904-908.

Dunwell TL, Hesson LB, Pavlova T, Zabarovska V, Kashuba V, Catchpoole D, Chiaramonte R, Brini AT, Griffiths M, Maher ER, Zabarovsky E, Latif F (2009) Epigenetic analysis of childhood acute lymphoblastic leukemia. *Epigenetics* **4**(3):185-193.

Dunwell T (2010) Epigenetic analysis of childhood leukaemia and the Hippo pathway. *University of Birmingham thesis.*

Dunwell T, Hesson L, Rauch TA, Wang L, Clark RE, Dallol A, Gentle D, Catchpoole D, Maher ER, Pfeifer GP and Latif F (2010) A genome wide screen identifies frequently methylated genes in haematological and epithelial cancers. *Molecular Cancer* **9**:44

Eads CA, Danenberg KD, Kawakami K, Saltz LB, Blake C, Shibata D, Danenberg PV, Laird PW (2000). MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Research* **28**(8):e32

Eckfeld K, Hesson L, Vos MD, Bieche I, Latif F, Clarke GJ (2004) RASSF4/AD037 is a potential ras effector/tumour suppressor of the RASSF family. *Cancer Research* **64**:8688-8693

Ehrich M (2009) DNA hypomethylation in cancer cells. *Epigenomics* **1**(2):239-259.

Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G, Cantor CR, Field JK, van den Boom D (2005) Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *PNAS* **102**(44):15785-15790.

Endoh M, Tamura G, Honda T, Homma N, Terashima M, Nishizka S, Motoyama T (2005) RASSF2, a potential tumour suppressor, is silenced by CpG island hypermethylation in gastric cancer *British Journal of Cancer* **93**:1395-1399

England R and Pettersson M (2005) Pyro Q-CpG^{TM} : quantitative analysis of methylation in multiple CpG sites by Pyropsequencing. *Nature Methods* doi:10.1038.

Erkko H, Xia B, Nikkilä J, Schleutker J, Syrjäkoski K, Mannermaa A, Kallioniemi A, Pylkäs K, Karppinen SM, Rapakko K, Miron A, Sheng Q, Li G, Mattila H, Bell DW, Haber DA, Grip M, Reiman M, Jukkola-Vuorinen A, Mustonen A, Kere J, Aaltonen LA, Kosma VM, Kataja V, Soini Y, Drapkin RI, Livingston DM, Winqvist R (2007) A recurrent mutation in PALB2 in Finnish cancer families. *Nature* **446**(7133):316-319.

Esteller M, Corn PG, Urena JM, Gabrielson E, Baylin SB, Herman JG (1998) Inactivation of glutathione S-transferase P1 gene by promoter hypermethylation in human neoplasia. *Cancer Research.* **58**(20):4514-8.

Esteller M, Hamilton SR, Burger PC, Baylin SB, Herman JG (1999) Inactivation of the DNA repair gene O6-methylguanine DNA-methyltransferase by promoter hypermethylation is a common event in primary human neoplasia. *Cancer Research* **59**(4):793-797

Esteller M, Garcia-Foncillas J, Andion E, Goodman SN, Hidalgo OF, Vanaclocha V, Baylin SB and Herman JG (2000a) Inactivation of the DNA repair gene *MGMT* and the clinical response of gliomas to alkylating agents. *New England Journal of Medicine* **343**(9):1350-1354.

Esteller M, Silva JM, Dominguez G, Bonilla F, Matias-Guiu X, Lerma E, Bussaglia E, Prat J, Harkes IC, Repasky EA, Gabrielson E, Schutte M, Baylin SB, Herman JG (2000b)

Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *Journal of the National Cancer Institute* **92**(7):564-569.

Esteller M, Sparks A, Toyota M, Sanchez-Cespedes M, Capella G, Peinado MA, Gonzalez S, Tarafa G, Sidransky D, Meltzer SJ, Baylin SB, Herman JG (2000c) Analysis of adenomatous polyposis coli promoter hypermethylation in human cancer. *Cancer Research* **60**(16):4366-4371.

Evron E, Umbricht CB, Korz D, Raman V, Loeb DM, Niranjan B, Buluwela L, Weitzman SA, Marks J, Sukumar S (2001) Loss of cyclin D2 expression in the majority of breast cancers is associated with promoter hypermethylation. *Cancer Research.* **61**(6):2782-7.

Fabbri M, Garzon R, Cimmino A, Liu Z, Zanesi N, Callegari E, Liu S, Alder H, Costinean S, Fernandez-Cymering C, Volinia S, Guler G, Morrison CD, Chan KK, Marcucci G, Calin GA, Huebner K, Croce CM (2007) MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. *PNAS* **104**(40):15805-10.

Fabbro M, Savage K, Hobson K, Deans AJ, Powell SN, McArthur GA, Khanna KK (2004) BRCA1-BARD1 complexes are required for p53Ser-15 phosphorylation and a G1/S arrest following ionizing radiation-induced DNA damage. *Journal of Biological Chemistry* **279**(30):31251-32158

Fackler MJ, McVeigh M, Evron E, Garrett E, Mehrotra J, Polyak K, Sukumar S, Argani P (2003) DNA methylation of RASSF1A, HIN-1, RAR-beta, Cyclin D2 and Twist in in situ and invasive lobular breast carcinoma. *International Journal of Cancer* **107**(6):970-975.

Falvella FS, Manenti G, Spinola M, Pignatiello C, Conti B, Pastorino U, Dragani TA (2006) Identification of RASSF8 as a candidate lung tumour suppressor gene. *Oncogene* **25**:3934-3938.

Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, Richardson TB, Santarosa M, Dillon KJ, Hickson I, Knights C, Martin NM, Jackson SP, Smith GC, Ashworth A. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy (1995) *Nature* **434**(7035):917-921.

Feuer EJ, Wun LM, Boring CC, Flanders WD, Timmel MJ, Tong T (1993) The lifetime risk of developing breast cancer. *Journal of the National Cancer Institute* **85**(11):892-897.

Fears S, Gavin M, Zhang DE, Hetherington C, Ben-David Y, Rowley JD, Nucifora G (1997) Functional characterization of ETV6 and ETV6/CBFA2 in the regulation of the MCSFR proximal promoter. *PNAS* **94**(5):1949-1954.

Feinberg AP, Tycko B (2004) The history of cancer epigenetics. *Nature Reviews Cancer* **4**(2):143-53

Feldman DR, Bosl GJ, Sheinfeld J, Motzer RJ (2008). Medical treatment of advanced testicular cancer. *Journal of the American Medical Association*. **299**(6):672-684.

Feng Q, Hawes SE, Stern JE, Wiens L, Lu H, Dong ZM, Jordan CD, Kiviat NB, Vesselle H (2008) DNA methylation in tumor and matched normal tissues from non-small cell lung cancer patients. *Cancer epidemiology, biomarkers and prevention.* **17**(3):645-654.

Fernandez-L A, Northcott PA, Dalton J, Fraga C, Ellison D, Angers S, Taylor MD, Kenney AM (2009) YAP1 is amplified and up-regulated in hedgehog-associated medulloblastomas and mediates Sonic hedgehog-driven neural precursor proliferation. *Genes and Development* **23**(23):2729-2741.

Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S, Reik W (2011) Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature. Apr 3 [Epub ahead of print]*

Figueira RCS, Gomes LR, Neto JS, Silva FC, Silva IDCG and Sogayar MC (2009) Correlation between MMPs and their inhibitors in breast cancer tumor tissye specimens and in cell lines with different metastatic potential. *BMC Cancer **9** 20*

Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, Li Y, Bhagwat N, Vasanthakumar A, Fernandez HF, Tallman MS, Sun Z, Wolniak K, Peeters JK, Liu W, Choe SE, Fantin VR, Paietta E, Löwenberg B, Licht JD, Godley LA, Delwel R, Valk PJ, Thompson CB, Levine RL, Melnick A (2010) Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**(6):553-567.

Flanagan JM, Cocciardi S, Waddell N, Johnstone CN, Marsh A, Henderson S, Simpson P, da Silva L; kConFab Investigators, Khanna K, Lakhani S, Boshoff C, Chenevix-Trench G (2010) DNA methylome of familial breast cancer identifies distinct profiles defined by mutation status. *American Journal of Human Genetics* **86**(3):420-433.

Friess H, Ding J, Kleeff J, Fenkell L, Rosinski JA, Guweidhi A, Reidhaar-Olson JF, Korc M, Hammer J, Büchler MW (2003) Microarray-based identification of differentially expressed growth- and metastasis-associated genes in pancreatic cancer. *Cellular and Molecular Life Sciences* **60**:1180-1199

Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *PNAS* **89**:1827-1831.

Fu YP, Yu JC, Cheng TC, Lou MA, Hsu GC, Wu CY, Chen ST, Wu HS, Wu PE, Shen CY (2003) Breast cancer risk associated with genotypic polymorphism of the nonhomologous end-joining genes: a multigenic study on cancer susceptibility. *Cancer Research* **63**(10):2440-2446.

Fujii H, Biel MA, Zhou W, Weitzman SA, Baylin SB, Gabrielson E (1998) Methylation of the HIC-1 candidate tumor suppressor gene in human breast cancer. *Oncogene* **16**(16):2159-2164.

Fung YK, Murphree AL, T'Ang A, Qian J, Hinrichs SH, Benedict WF (1987) Structural evidence for the authenticity of the human retinoblastoma gene. *Science* **236**(4809):1657-1661.

Gao S, Worm J, Guldberg P, Eiberg H, Krogdahl A, Sørensen JA, Liu C-J, Reibel J, Dabelsteen E (2004) Loss of heterozygosity at 9q33 and hypermethylation of the DBCCR1 gene in oral squamous cell carcinoma. *British Journal of Cancer* **91**:760-764.

Gaudet MM, Hunter K, Pharoah P, Dunning AM, Driver K, Lissowska J, Sherman M, Peplonska B, Brinton LA, Chanock S, Garcia-Closas M (2009) Genetic variation in SIPA1 in relation to breast cancer risk and survival after breast cancer diagnosis. *International Journal of Cancer* **124**(7):1716-20.

Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. *Journal of Molecular Biology* **196**(2):261-282

Genevet A, Wehr MC, Brain R, Thompson BJ and Tapon N (2010) Kibra is a Regulator of the Salvador/Warts/Hippo Signaling Network. *Developmental Cell* **18**:300-308.

Giles RH, Dauwerse JG, Higgins C, Petrij F, Wessels JW, Beverstock GC, Döhner H, Jotterand-Bellomo M, Falkenburg JH, Slater RM, van Ommen GJ, Hagemeijer A, van der Reijden BA, Breuning MH (1997) Detection of CBP rearrangements in acute myelogenous leukemia with t(8;16). *Leukaemia* **11**(12):2087-2096

Giotopoulos G, McCormick C, Cole C, Zanker A, Jawad M, Brown R, Plumb M (2006) DNA methylation during mouse hemopoietic differentiation and radiation-induced leukemia. *Experimental Hematology* **34**(11):1462-1470.

Glas AM, Floore A, Delahaye LJ, Witteveen AT, Pover RC, Bakx N, Lahti-Domenici JS, Bruinsma TJ, Warmoes MO, Bernards R, Wessels LF, Van't Veer LJ (2006) Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics* **30**(7):278

Globisch D, Münzel M, Müller M, Michalakis S, Wagner M, Koch S, Brückl T, Biel M, Carell T (2011) Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* 5(12):e15367

Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang X, Golic KG, Jacobsen SE, Bestor TH (2006) Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science* **311**(5759):395-398.

Gopalakrishnan S, Sullivan BA, Trazzi S, Della Vale G, Robertson KD (2009) DNMT3B interacts with constitutive centromeric protein CEMP-C to modulate DNA methylation and the histone code at centromeric regions. *Human Molecular Genetics* **18**:3178-3193.

Gorringe KL, Ramakrishna M, Williams LH, Sridhar A, Boyle SE, Bearfoot JL, Li J, Anglesio MS, Campbell IG (2009) Are there any more ovarian tumor suppressor genes? A new perspective using ultra high-resolution copy number and loss of heterozygosity analysis. *Genes chromosomes and cancer* **48**(10):931-942.

Goshima G, Wollman R, Goodwin SS, Zhang N, Scholey JM, Vale RD, Stuurman N (2007) Genes required for mitotic spindle assembly in Drosophila S2 cells. *Science* **316**(5823):417-421

Grady WM, Willis J, Guilford PJ, Dunbier AK, Toro TT, Lynch H, Wiesner G, Ferguson K, Eng C, Park JG, Kim SJ, Markowitz S (2000) Methylation of the CDH1 promoter as the second genetic hit in hereditary diffuse gastric cancer. *Nature Genetics* **26**(1):16-17.

Greaves MF, Wiemels J (2003) Origins of chromosome translocations in childhood leukaemia. *Nature Reviews Cancer* **3**(9):639-649.

Greenberg RA, Sobhian B, Pathania S, Cantor SB, Nakatani Y, Livingston DM (2006) Multifactorial contributions to an acute DNA damage response by BRCA1/BARD1-containing complexes. *Genes and Development* **20**(1):34-46.

Greenman C, STepehens P, Smith R, Dalgleish GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, Edkins S, O'Meara S, Vastrik I, Schmidt EE, Avis T, Barthorpe S, Bhamra G, Buck G, Choudhury B, Clements J, Cole J, Dicks E, Forbes S, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jenkinson A, Jones D, Menzies A, Mironenko T, Perry J, Raine K, Richardson D, Shepherd R, Small A, Tofts C, Varian J, Webb T, West S, Widaa S, Yates A, Cahill DP, Louis DN, Goldstraw P, Nicholson AG, Brasseur F, Looijenga L, Weber BL, Chiew Y-E, deFazio D, Greaves MF, Green AR, Campbell P, Birney E, Easton DF, Chenevix-Trench G, Tan M-H, Khoo SK, Teh BT, Yuen ST, Leung SY, Wooster R, Futreal PA and Stratton MR (2007) Patterns of somatic mutation in human cancer genomes. *Nature* **446**(7132):153-158.

Greger V, Passarge E, Höpping W, Messmer E, Horsthemke B (1989) Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. *Human Genetics* **83**(2):155-158.

Groden J, Thliveris A, Samowitz W, Carlson W, Gelbert L, Albertsen H, Joslyn G, Stevens J, Spirio L, Robertson M, Sargeant L, Krapcho K, Wolff E, Burt R, Hughes JP, Warrington J, McPherson J, Wasmuth J, Le Paslier D, Abderrahim H, Cohen D, Leppert M and White R (1991) Identification and characterization of the familial adenomatous polyposis coli gene *Cell*. **66**(3):589-600

Grusche FA, Richardson HE and Harvey KF (2010) Upstream Regulation of the Hippo Size Control Pathway. *Current Biology* **20:**R574-R582.

Guidez F, Petrie K, Ford AM, Lu H, Bennett CA, MacGregor A, Hannemann J, Ito Y, Ghysdael J, Greaves M, Wiedemann LM, Zelent A (2000) Recruitment of the nuclear receptor corepressor N-CoR by the TEL moiety of the childhood leukemia-associated TEL-AML1 oncoprotein. *Blood* **96**(7):2557-2561.

Guénard F, Labrie Y, Ouellette G, Beauparlant CJ, Bessette P, Chiquette J, Laframboise R, Lépine J, Lespérance B, Pichette R, Plante M, Durocher F; INHERIT BRCAs (2007) Germline mutations in the breast cancer susceptibility gene PTEN are rare in high-risk non-BRCA1/2 French Canadian breast cancer families. *Familial Cancer* **6**(4):483-490.

Guilford PJ, Hopkins JB, Grady WM, Markowitz SD, Willis J, Lynch H, Rajput A, Wiesner GL, Lindor NM, Burgart LJ, Toro TT, Lee D, Limacher JM, Shaw DW, Findlay MP, Reeve AE (1999) E-cadherin germline mutations define an inherited cancer syndrome dominated by diffuse gastric cancer. *Human Mutation* **14**(3):249-255.

Guo C, Tommasi S, Liu L, Yee JK, Dammann R, Pfeifer GP (2007) RASSF1A is part of a complex similar to the Drosophila Hippo/Salvador/Lats tumor-suppressor network. *Current Biology* **17**(8):700-705.

Gupta A, Godwin AK, Vanderveer L, Lu A, Liu J (2003) Hypomethylation of the synuclein gamma gene CpG island promotes its aberrant expression in breast carcinoma and ovarian carcinoma. *Cancer Research* **63**(3):664-673.

Habuchi T, Luscombe M, Elder PA, Knowles MA (1998) Structure and methylation based silencing of a Gene (DBCCR1) within a candidate bladder cancer tumor suppressor region at 9q23-33. *Genomics* **48**:277-288.

Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250**(4988):1684-1689.

Hamajima N, Hirose K, Tajima K, Rohan T, Calle EE, Heath CW Jr, Coates RJ, Liff JM, Talamini R, Chantarakul N, Koetsawang S, Rachawat D, Morabia A, Schuman L, Stewart W, Szklo M, Bain C, Schofield F, Siskind V, Band P, Coldman AJ, Gallagher RP, Hislop TG, Yang P, Kolonel LM, Nomura AM, Hu J, Johnson KC, Mao Y, De Sanjosé S, Lee N, Marchbanks P, Ory HW, Peterson HB, Wilson HG, Wingo PA, Ebeling K, Kunde D, Nishan P, Hopper JL, Colditz G, Gajalanski V, Martin N, Pardthaisong T, Silpisornkosol S, Theetranont C, Boosiri B, Chutivongse S, Jimakorn P, Virutamasen P, Wongsrichanalai C, Ewertz M, Adami HO, Bergkvist L, Magnusson C, Persson I, Chang-Claude J, Paul C, Skegg DC, Spears GF, Boyle P, Evstifeeva T, Daling JR, Hutchinson WB, Malone K, Noonan EA, Stanford JL, Thomas DB, Weiss NS, White E, Andrieu N, Brêmond A, Clavel F, Gairard B, Lansac J, Piana L, Renaud R, Izquierdo A, Viladiu P, Cuevas HR, Ontiveros P, Palet A, Salazar SB, Aristizabel N, Cuadros A, Tryggvadottir L, Tulinius H, Bachelot A, Lê MG, Peto J, Franceschi S, Lubin F, Modan B, Ron E, Wax Y, Friedman GD, Hiatt RA, Levi F, Bishop T, Kosmelj K, Primic-Zakelj M, Ravnihar B, Stare J, Beeson WL, Fraser G, Bullbrook RD, Cuzick J, Duffy SW, Fentiman IS, Hayward JL, Wang DY, McMichael AJ, McPherson K, Hanson RL, Leske MC, Mahoney MC, Nasca PC, Varma AO, Weinstein AL, Moller TR, Olsson H, Ranstam J, Goldbohm RA, van den Brandt PA, Apelo RA, Baens J, de la Cruz JR, Javier B, Lacaya LB, Ngelangel CA, La Vecchia C, Negri E, Marubini E, Ferraroni M, Gerber M, Richardson S, Segala C, Gatei D, Kenya P, Kungu A, Mati JG, Brinton LA, Hoover R, Schairer C, Spirtas R, Lee HP, Rookus MA, van Leeuwen FE, Schoenberg JA, McCredie M, Gammon MD, Clarke EA, Jones L, Neil A, Vessey M, Yeates D, Appleby P, Banks E, Beral V, Bull D, Crossley B, Goodill A, Green J, Hermon C, Key T, Langston N, Lewis C, Reeves G, Collins R, Doll R, Peto R, Mabuchi K, Preston D, Hannaford P, Kay C, Rosero-Bixby L, Gao YT, Jin F, Yuan JM, Wei HY, Yun T, Zhiheng C, Berry G, Cooper Booth J, Jelihovsky T, MacLennan R, Shearman R, Wang QS, Baines CJ, Miller AB, Wall C, Lund E, Stalsberg H, Shu XO, Zheng W, Katsouyanni K, Trichopoulou A, Trichopoulos D, Dabancens A,

Martinez L, Molina R, Salas O, Alexander FE, Anderson K, Folsom AR, Hulka BS, Bernstein L, Enger S, Haile RW, Paganini-Hill A, Pike MC, Ross RK, Ursin G, Yu MC, Longnecker MP, Newcomb P, Bergkvist L, Kalache A, Farley TM, Holck S, Meirik O; Collaborative Group on Hormonal Factors in Breast Cancer (2002) Alcohol, tobacco and breast cancer-collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. *British Journal of Cancer* **87**(11):1234-1245.

Hanahan D and Weinberg RA (2000) The Hallmarks of Cancer *Cell* **100**:57-70.

Harbour JW, Onken MD, Roberson ED, Duan S, Cao L, Worley LA, Council ML, Matatall KA, Helms C, Bowcock AM (2010) Frequent mutation of BAP1 in metastasizing uveal melanomas. *Science* **330**(6009):1410-1413.

Hartmann O, Spyratos F, Harbeck N, Dietrich D, Fassbender A, Schmitt M, Eppenberger-Castori S, Vuaroqueaux V, Lerebours F, Welzel K, Maier S, Plum A, Niemann S, Foekens JA, Lesche R, Martens JW (2009) DNA methylation markers predict outcome in node-positive, estrogen receptor-positive breast cancer with adjuvant anthracycline-based chemotherapy. *Clinical Cancer Research* **15**(1):315-323.

Hartmann EM, Campo E, Wright G, Lenz G, Salaverria I, Jares P, Xiao W, Braziel RM, Rimsza LM, Chan W-C, Weisenburger DD, Delabie J, Jaffe ES, Gascoyne RD, Dave SS, Mueller-Hermelink H-K, Staudt LM, Ott G, Beá S and Rosenwald A (2010) Pathway discovery in mantle cell lymphoma by integrated analysis of high-resolution gene expression and copy number profiling. *Blood* **116:**953-961.

Harvey KF, Pfleger CM, Hariharan IK (2003) The Drosophila Mst ortholog, hippo, restricts growth and cell proliferation and promotes apoptosis. *Cell* **114**(4):457-467.

Harvey K and Tapon N (2007) The Salvador-Warts-Hippo pathway- an emerging tumour suppressor network. *Nature Reviews Cancer* **7:**182-191.

Hayami S, Kelly JD, Cho HS, Yoshimatsu M, Unoki M, Tsunoda T, Field HI, Neal DE, Yamaue H, Ponder BA, Nakamura Y, Hamamoto R (2011) Overexpression of LSD1 contributes to human carcinogenesis through chromatin regulation in various cancers. *International Journal of Cancer* **128**(3):574-586.

Hayatsu H, Wataya Y, Kai K (1970) The addition of sodium bisulphite to uracil and cytosine. *Journal of the American Chemical Society* **92**:724-726.

Hayashi A, Horiuchi A, Kikuchi N, Hayashi T, Fuseya C, Suzuki A, Konishi I, Shiozawa T (2010) Type-specific roles of histone deacetylase (HDAC) overexpression in ovarian carcinoma: HDAC1 enhances cell proliferation and HDAC3 stimulates cell migration with downregulation of E-cadherin. *International Journal of Cancer* **127**(6):1332-1346.

Hayes DF (2000) Atlas of Breast Cancer 2nd Edition. *Mosby*.

Heard E (2004). Recent advances in X-chromosome inactivation. *Current Opinion in Cell Biology* **16**(3):247-55.

Hearle N, Schumacher V, Menko FH, Olschwang S, Boardman LA, Gille JJ, Keller JJ, Westerman AM, Scott RJ, Lim W, Trimbath JD, Giardiello FM, Gruber SB, Offerhaus GJ, de Rooij FW, Wilson JH, Hansmann A, Möslein G, Royer-Pokora B, Vogel T, Phillips RK, Spigelman AD, Houlston RS (2006) Frequency and spectrum of cancers in the Peutz-Jeghers syndrome. *Clinical Cancer Research* **12**(10):3209-3215.

Helzlsouer KJ, Harris EL, Parshad R, Fogel S, Bigbee WL, Sanford KK (1995) Familial clustering of breast cancer: possible interaction between DNA repair proficiency and radiation exposure in the development of breast cancer. *International Journal of Cancer* **64**(1):14-17.

Hendrich B and Bird A (1998) Identification and characrerization of a family of mammalian methyl-CpG binding proteins. *Molecular and Cellular Biology* **18**:6538-6547.

Herman JG, Latif F, Weng Y, Lerman MI, Zbar B, Liu S, Samid D, Duan DS, Gnarra JR, Linehan WM, Baylin SB (1994) Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *PNAS* **91**(21):9700-9704.

Herman JG, Merlo A, Mao L, Lapidus RG, Issa JP, Davidson NE, Sidransky D, Baylin SB (1995) Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. *Cancer Research* **55**(20):4525-4530.

Herman JG, Graff JR, Myöhänen S, Nelkin BD, Baylin SB (1996) Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *PNAS* **93**(18):9821-9826.

Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S, Backlund MG, Yin Y, Khramtsov AI, Bastein R, Quackenbush J, Glazer RI, Brown PH, Green JE, Kopelovich L, Furth PA, Palazzo JP, Olopade OI, Bernard PS, Churchill GA, Van Dyke T, Perou CM (2007) Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biology* **8**(5):R76

Hesson L, Dallol A, Minna JD, Maher ER, Latif F (2003) NORE1A, a homologue of RASSF1A tumour suppressor gene is inactivated in human cancers. *Oncogene* **22**:947-954

Hesson L, Bièche I, Krex D, Criniere E, Hoang-Xuan K, Maher ER, Latif F (2004) Frequent epigenetic inactivation of *RASSF1A* and *BLU* genes located within the critical 3p21.3 region in gliomas. *Oncogene* **23**(13):2408-19

Hesson LB, Wilson R, Morton D, Adams C, Walker M, Maher ER, Latif F (2005) CpG island promoter hypermethylation of a novel Ras-effector gene RASSF2A s an early event in colon carcinogenesis and correlates inversely with K-ras mutations. *Oncogene* **24**:3987-3994

Hesson LB, Cooper WN, Latif F (2007) The role of RASSF1A methylation in cancer. *Disease Markers* **23**:73-87.

Hesson LB, Krex D and Latif F (2008) Epigenetic markers in human gliomas: prospects for therapeutic intervention. *Expert Review of Neurotherapeutics* **8**(10):1475-1496

Hesson LB, Dunwell TL, Cooper WN, Catchpoole D, Brini AT, Chiaramonte R, Griffiths M, Chalmers AD, Maher ER and Latif F (2009) The novel *RASSF6* and *RASSF10* candidate tumour suppressor genes are frequently epigenetically inactivated in childhood leukaemias. *Molecular Cancer* **8**:42

Hiebert SW, Sun W, Davis JN, Golub T, Shurtleff S, Buijs A, Downing JR, Grosveld G, Roussell MF, Gilliland DG, Lenny N, Meyers S (1996) The t(12;21) translocation converts AML-1B from an activator to a repressor of transcription. *Molecular and cellular biology* **16**(4):1349-1355.

Hill VK, Hesson LB, Dansranjavin T, Dallol A, Bieche I, Vacher S, Tommasi S, Dobbins T, Gentle D, Euhus D, Lewis C, Dammann R, Ward RL, Minna J, Maher ER, Pfeifer GP, Latif F (2010) Identification of 5 novel genes methylated in breast and other epithelial cancers. *Molecular Cancer* **5**(9):51

Hill VK, Underhill-Day N, Krex D, Robel K, Sangan CB, Summersgill HR, Morris M, Gentle D, Chalmers AD, Maher ER and Latif F (2011) Epigenetic inactivation of the *RASSF10* candidate tumor suppressor gene is a frequent and an early event in gliomagenesis. *Oncogene* **30**(8):978-989

Hitomi J, Christofferson DE, Ng A, Yao J, Degterev A, Xavier RJ and Yuan J (2008) Identification of a molecular signalling network that regulates a cellular necrotic cell deat pathway. *Cell* **135**:1311-1323.

Holm K, Hegardt C, Staaf J, Vallon-Christersson J, Jonsson G, Olsson H, Borg A and Ringner M (2010) Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns. *Breast Cancer Research* **12**(3):R36.

Hoffman A, Zheng T, Yi C, Learderer D, Weidhaas J, Slack F, Zhang Y, Paranjape T, Zhu Y (2009) microRNA miR-196-a2 and Breast Cancer: A Genetic and Epigenetic Association study and Functional Analysis. *Cancer Research* **69**(14):5970-5977.

Honorio S, Agathanggelou A, Wernert N, Rothe M, Maher ER, Latif F (2003) Frequent epigenetic inactivation of the RASSF1A tumour suppressor gene in testicular tumours and distinct methylation profiles of seminoma and nonseminoma testicular germ cell tumours. *Oncogene* **22**(3):461-466.

Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A(2010) The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One* **26**(5):e8888.

Huang J, Zhang L, Greshock J, Colligon TA, Wang Y, Ward R, Katsaros D, Lassus H, Butzow R, Godwin AK, Testa JR, Nathanson KL, Gimotty PA, Coukos G, Weber BL, Degenhardt Y (2011) Frequent genetic abnormalities of the PI3K/AKT pathway in primary ovarian cancer predict patient outcome. *Genes Chromosomes and Cancer* **50** (8):606-618.

Hudis CA (2007). Trastuzumab--mechanism of action and use in clinical practice. *New England Journal of Medicine* **357**(1):39-51.

Hussussian CJ, Struewing JP, Goldstein AM, Higgins PA, Ally DS, Sheahan MD, Clark WH Jr, Tucker MA, Dracopoli NC (1994) Germline p16 mutations in familial melanoma. *Nature Genetics* **8**(1):15-21.

ICRF (1997) Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Collaborative Group on Hormonal Factors in Breast Cancer. *Lancet* **350**(9084):1047-1059.

Ida K, Kitabayashi I, Taki T, Taniwaki M, Noro K, Yamamoto M, Ohki M, Hayashi Y (1997) Adenoviral E1A-associated protein p300 is involved in acute myeloid leukemia with t(11;22)(q23;q13). *Blood* **90**(12):4699-4704

Ikeda M, Hirabayashi S, Fujiwara N, Mori H, Kawata A, Iida J, Bao Y, Sato Y, Iida T, Sugimura H, Hata Y (2007) Ras-association domain family protein 6 induces apoptosis via both caspase-dependent and caspase-independent pathways. *Experimental Cell Research* **313**(7):1484-1495.

Imoto I, Yang ZQ, Pimkhaokham A, Tsuda H, Shimada Y, Imamura M, Ohki M, Inazawa J (2001) Identification of cIAP1 as a candidate target gene within an amplicon at 11q22 in esophageal squamous cell carcinomas. *Cancer Research* **61**(18):6629-6634.

Inohara N, Koseki T, Chen S, Wu X, Núńez G (1998). CIDE, a novel family of cell death activators with homology to the 45kDa subunit of the DNA fragmentation factor. *The EMBO Journal* **17**(9):2526-2533.

Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, Ji H, Potash JB, Abunciyan S, Feinberg AP (2009) The human colon cancer methylation shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genetics* **41**(2):178-186.

Ishii N, Maier D, Merlo A, Tada M, Sawamura Y, Diserens AC, Van Meir EG (1999) Frequent co-alterations of TP53, p16/CDKN2A, p14ARF, PTEN tumor suppressor genes in human glioma cell lines. *Brain Pathology* **3**:469-479.

Issa JP, Zehnbauer BA, Kaufmann SH, Biel MA, Baylin SB (1997) HIC1 hypermethylation is a late event in hematopoietic neoplasms. *Cancer Research* **57**(9):1678-1681.

Izumi H, Inoue J, Yokoi S, Hosoda H, Shibata T, Sunamori M, Hirohashi S, Inazawa J, Imoto I (2005) Frequent silencing of *DBC1* is by genetic or epigenetic mechanisms in non-small cell lung cancers. *Human Molecular Genetics* **14**(8):997-1007.

Jackson K, Soutto M, Peng D, Hu T, Marshal D and El-Rifai W (2010) Epigenetic silencing of somatostatin in gastric cancer. *Digestive Diseases and Sciences* Oct 7.

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D (2011) Global cancer statistics. *CA: a cancer journal for clinicians* **61**:69-90.

Jenne DE, Reimann H, Nezu J, Friedel W, Loff S, Jeschke R, Müller O, Back W, Zimmer M. Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase (1998) *Nature Genetics* **18**(1):38-43

Jerónimo C, Usadel H, Henrique R, Oliveira J, Lopes C, Nelson WG, Sidransky D (2001) Quantitation of GSTP1 methylation in non-neoplastic prostatic tissue and organ-confined prostate adenocarcinoma. *Journal of the National Cancer Institute* **93**(22):1747-1752

Jiang Z, Li X, Hu J, Zhou W, Jiang Y, Li G, Lu D (2006) Promoter hypermethylation-mediated down-regulation of LATS1 and LATS2 in human astrocytoma. *Neuroscience Research* **56**(4):450-458.

Jicai Z, Zongtao Y, Jun L, Haiping L, Jianmin W, Lihua H (2006) Persistent infection of hepatitis B virus is involved in high rate of p16 methylation in hepatocellular carcinoma. *Molecular Carcinogenesis* **45**(7):530-536.

Jiménez-Velasco A, Román-Gómez J, Aggirre X, Barrios M, Navarro G, Vázquez I, Prósper F, Torres A and Heiniger A (2005) Downregulation of the large tumour suppressor 2 (*LATS2/KPM)* gene is associated with poor prognosis in acute lymphoblastic leukemia. *Leukemia* **19**:2347-2350.

Jin SG, Kadam S, Pfeifer GP (2010) Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydorxymethylcytosine. *Nucleic Acids Research* **38**(11):e125.

Jin KL, Pak JH, Park JY, Choi WH, Lee JY, Kim JH, Nam JH (2008) Expression profile of histone deacetylases 1, 2 and 3 in ovarian cancer tissues. *Journal of Gynecological Oncology* **19**(3):185-190

Jin Z, Mori Y, amilton JP, Olaru A, Sato F, Yang J, Ito T, Kan T, Agarwal R and Meltzer SJ (2007) Hypermethylation of the Somatostatin Promoter Is a Common, Early Event in Human Esophageal Carcinogenesis. *Cancer* **112**(1):43-49.

Jin Z, Tamura G, Tsuchiya T, Sakata K, Kashiwaba M, Osakabe M, Motoyama T (2001) Adenomatous polyposis coli (APC) gene promoter hypermethylation in primary breast cancers. *British Journal of Cancer* **85**(1):69-73.

Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J, Jacobsen SE (2007) The SRA methyl-cytosine-binding domain links DNA and histone methylation. *Current Biology* **17**(4):379-384.

Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Kamiyama H, Jimeno A, Hong SM, Fu B, Lin MT, Calhoun ES, Kamiyama M, Walter K, Nikolskaya T, Nikolsky Y, Hartigan J, Smith DR, Hidalgo M, Leach SD, Klein AP, Jaffee EM, Goggins M, Maitra A, Iacobuzio-Donahue C, Eshleman JR, Kern SE, Hruban RH, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE, Kinzler KW

(2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321**(5897):1801-1806

Jones S, Wang TL, Shih IeM, Mao TL, Nakayama K, Roden R, Glas R, Slamon D, Diaz LA Jr, Vogelstein B, Kinzler KW, Velculescu VE, Papadopoulos N (2010) Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* **330**(6001):228-231.

Jordan VC (1993) Fourteenth Gaddum Memorial Lecture. A current view of tamoxifen for the treatment and prevention of breast cancer. *British Journal of Pharmacology* **110**(2):507-517

Justice RW, Zilian O, Woods DF, Noll M, Bryant PJ (1995) The Drosophila tumor suppressor gene warts encodes a homolog of human myotonic dystrophy kinase and is required for the control of cell shape and proliferation. *Genes and Development* **9**(5):534-546.

Kaiser J (2010) Epigenetic drugs take on cancer. *Science* **330**:576-578.

Kang S, Lee JM, Jeon ES, Lee S, Kim H, Kim HS, Seo SS, Park SY, Sidranksy D, Dong SM (2006) RASSF1A hypermethylation and its inverse correlation with BRAF and/or KRAS mutations in MSI-associated endometrial carcinoma. *International Journal of Cancer* **119**(6):1316-1321.

Karkera JD, Ayache S, Ransome RJ, Jackson MA, Elsayem AF, Sridhar R, Detera-Wadleigh SD, Wadleigh RG (2000) Refinement of Regions with Allelic Loss on Chromosome 18p11.3 and 18q12.2 in Esophageal Squamous Cell Carcinoma. *Clinical Cancer Research* **6**:3565-3569.

Kato M, Yano K, Matsuo F, Saito H, Katagiri T, Kurumizaka H, Yoshimoto M, Kasumi F, Akiyama F, Sakamoto G, Nagawa H, Nakamura Y, Miki Y (2000) Identification of Rad51 alteration in patients with bilateral breast cancer. *Jounral of Human Genetics* **45**(3):133-137.

Kawakami K, Ishida Y, Danenberg KD, Omura K, Watanabe G, Danenberg PV (2002) Functional polymorphism of the thymidylate synthase gene in colorectal cancer accompanied by frequent loss of heterozygosity. *Japanese Journal of Cancer Research* **93**(11):1221-1229.

Keshet I, Schlesinger Y, Farkash S, Rand E, Hecht M, Segal E, Pikarski E, Young RA, Niveleau A, Cedar H, Simon I (2006). Evidence for an instructive mechanism of de novo methylation in cancer cells. *Nature Genetics* **38**(2):149-53

Khokhlatchev A, Rabizadeh S, Xavier R, Nedwidek M, Chen T, Zhang XF, Seed B, Avruch J (2002) Identification of a novel Ras-regulated proapoptotic pathway. *Current Biology* **12**(4):253-265.

Kim SJ, Kang HS, Chang HL, Jung YC, Sim HB, Lee KS, Ro J, Lee ES (2008) Promoter hypomethylation of the N-acetyltransferase 1 gene in breast cancer. *Oncology Reports* **19**(3):663-668.

Kimura H, Shiota K (2003) Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. *Journal of Biological Chemistry* **278**(7):4806-4812

Kitabayashi I, Yokoyama A, Shimizu K, Ohki M (1998) Interaction and functional cooperation of the leukemia-associated factors AML1 and p300 in myeloid cell differentiation. *EMBO Journal* **17**(11):2994-3004.

Kitagawa D, Kajiho H, Negishi T, Ura S, Watanabe T, Wada T, Ichijo H, Katada T, Nishina H (2006) Release of RASSF1C from the nucleus by Daxx degradation links DNA damage and SAPK/JNK activation. *EMBO Journal* **25**(14):3286-3297.

Kittiniyom K, Gorse KM, Dakbegue F, Lichy JH, Taubenberger JK, Newsham IF (2001) Allelic loss on chromosome band 18p11.3 occurs early and reveals heterogeneity in breast cancer progression. *Breast Cancer Research* **3**:192-198.

Knudson AG Jr (1971). Mutation and cancer: statistical study of retinoblastoma. *PNAS* **68**(4):820-3.

Ko M, Huang Y, Jankowska AM, Pape UJ, Tahiliani M, Bandukwala HS, An J, Lamperti ED, Koh KP, Ganetzky R, Liu XS, Aravind L, Agarwal S, Maciejewski JP, Rao A (2010) Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**(7325):839-843.

Kobayashi Y, Absher DM, Gulzar ZG, Young SR, McKenney JK, Peehl DM, Brooks JD, Myers RM, Sherlock G (2011) DNA methylation profiling reveals novel biomarkers and important roles for DNA methyltransferases in prostate cancer. *Genome Research* **Apr 26** [Epub. ahead of print].

Koga Y, Pelizzola M, Cheng E, Krauthammer M, Sznol M, Ariyan S, Narayan D, Molinaro AM, Halaban R, Weissman SM (2009) Genome-wide screen of promoter methylation identifies novel markers in melanoma. *Genome Research* **19**(8):1462-1470

Kok K, Naylor SL, Buys CH (1997) Deletions of the short arm of chromosome 3 in solid tumors and the search for suppressor genes. *Advances in Cancer Research* **71**:27-92

Kosaka Y, Mimori K, Tanaka F, Inoue H, Watanabe M, Mori M (2007) Clinical significance of the loss of MATS1 mRNA expression in colorectal cancer. *International Journal of Oncology* **31**(2):333-338.

Koshiishi N, Chong JM, Fukasawa T, Ikeno R, Hayashi Y, Funata N, Nagai H, Miyaki M, Matsumoto Y, Fukayama M (2004) p300 gene alterations in intestinal and diffuse types of gastric carcinoma. *Gastric Cancer* **7**(2):85-90.

Kriaucionis S, Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**(5929):929-930.

Kuang SQ, Tong WG, Yang H, Lin W, Lee MK, Fang ZH, Wei Y, Jelinek J, Issa JP, Garcia-Manero G (2008) Genome-wide identification of aberrantly methylated promoter associated CpG islands in acute lymphocytic leukemia. *Leukemia* **22**:1529-1538

Kukitsu T, Takayama T, Miyanishi K, Nobuoka A, Katsuki S, Sato Y, Takimoto R, Matsunaga T, Kato J, Sonoda T, Sakamaki S, Niitsu Y (2008) Aberrant crypt foci as precursors of the dysplasia-carcinoma sequence in patients with ulcerative colitis. *Clinical Cancer Research* **14**(1):48-54.

Lai ZC, Wei X, Shimizu T, Ramos E, Rohrbaugh M, Nikolaidis N, Ho LL, Li Y (2005) Control of cell proliferation and apoptosis by mob as tumor suppressor, mats. *Cell* **120**(5):675-685.

Lambe M, Hsieh CC, Chan HW, Ekbom A, Trichopoulos D, Adami HO (1996) Parity, age at first and last birth, and risk of breast cancer: a population-based study in Sweden. *Breast Cancer Research and Treatment* **38**(3):305-311.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-

Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ; International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* **15**(409):860-921

Langer R, Mutze K, Becker K, Feith M, Ott K, Höfler H, Keller G (2010) Expression of class I histone deacetylases (HDAC1 and HDAC2) in oesophageal adenocarcinomas: an immunohistochemical study. *Journal of Clinical Pathology* **63**(11):994-998.

Larsen F, Gundersen G, Lopez R, Prydz H (1992) CpG islands as gene markers in the human genome. *Genomics* **13**:1095–1107

Latif F., Tory K., Gnarra J., Yao M., Duh F-M., Orcutt M. L., Stackhouse T., Kuzmin I., Modi W., Geil L., Schmidt L., Zhou F., Li H., Wei M. H., Chen F., Glenn G., Choyke P., Walther M. M., Weng Y., Duan D. R., Dean M., Glavac D., Richards F. M., Crossey P. A., Ferguson-Smith M. A., Le Paslier D., Chumakov I., Cohen D., Chinault A. C., Maher E. R., Linehan W. M., Zbar B., Lerman M. I (1993) Identification of the von Hippel-Lindau disease tumor suppressor gene. *Science* **260**:1317-1320

Laurent E, Talpaz M, Kantarjian H, Kurzrock R (2001) The BCR gene and Philadelphia chromosome-positive leukemogenesis. *Cancer Research* **61**(6):2343-55.

Le LP, Nielsen GP, Rosenberg AE, Thomas D, Batten JM, Deshpande V, Schwab J, Duan Z, Xavier RJ, Hornicek FJ, Iafrate AJ (2011) Recurrent chromosomal copy number alterations in sporadic chordomas. *PLoS One* **6**(5):e18846.

Le Guezennec X, Vermeulen M, Brinkman AB, Hoeijmakers WA, Cohen A, Lasonder E, Stunnenberg HG (2006) MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Molecular and Cellular Biology* **26**(3):843-851.

Ledford H (2010) *Nature* **464**:972-974.

Lee WH, Shew JY, Hong FD, Sery TW, Donoso LA, Young LJ, Bookstein R, Lee EY (1987) The retinoblastoma susceptibility gene encodes a nuclear phosphoprotein associated with DNA binding activity. *Nature* **329**(6140):524-625.

Lee MG, Kim HY, Byun DS, Lee SJ, Lee CH, Kim JI, Chang SG, Chi SG (2001) Frequent epigenetic inactivation of RASSF1A in human bladder carcinoma. *Cancer Research* 61(18):6688-6692.

Levy-Lahad E, Lahad A, Eisenberg S, Dagan E, Paperna T, Kasinetz L, Catane R, Kaufman B, Beller U, Renbaum P, Gershoni-Baruch R (2001) A single nucleotide polymorphism in the RAD51 gene modifies cancer risk in BRCA2 but not BRCA1 carriers. *PNAS* **98**(6):3232-3236.

Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, Kandoth C, Payton JE, Baty J, Welch J, Harris CC, Lichti CF, Townsend RR, Fulton RS, Dooling DJ, Koboldt DC, Schmidt H, Zhang Q, Osborne JR, Lin L, O'Laughlin M, McMichael JF, Delehaunty KD, McGrath SD, Fulton LA, Magrini VJ, Vickery TL, Hundal J, Cook LL, Conyers JJ, Swift GW, Reed JP, Alldredge PA, Wylie T, Walker J, Kalicki J, Watson MA, Heath S, Shannon WD, Varghese N, Nagarajan R, Westervelt P, Tomasson MH, Link DC, Graubert TA, DiPersio JF, Mardis ER, Wilson RK (2010) DNMT3A mutations in acute myeloid leukemia. *New England Journal of Medicine* **363**:2424-2433.

Li CI, Uribe DJ, Daling JR (2005) Clinical characteristics of different histologic types of breast cancer. *British Journal of Cancer* **93**:1046-52.

Li D, Liang D, Tang H, Li T, Zhao M (2008) CpG methylation plays a vital role in determining tissue- and cell-specific expression of the human cell-death inducing DFF45-like effector A gene through the regulation of Sp1/Sp3 binding. *Nucleic Acids Research* **36**:330-341.

Li E, Bestor TH and Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69:** 915-926.

Li L, Lee K-M, Hna M, Choi J-Y, Lee JY, Kang GH, Park SK, Noh D-Y , Yoo K-Y and Kang D (2010) Estrogen and progesterone receptor status affect genome-wide DNA methylation profile in breast cancer. *Human Molecular Genetics* **19**(21):4273-7.

Li S, Ting NSY, Zheng L, Chen P-L, Ziv Y, Shiloh Y, Lee EY-HP, Lee W-H (2000) Functional link of BRCA1 and ataxia telangiectasia gene product in DNA damage response. *Nature* **406**:210-215.

Li Q, Jedlicka A, Ahuja N, Gibbons MC, Baylin SB, Burger PC, Issa JP (1998) Concordant methylation of the ER and N33 genes in glioblastoma multiforme. *Oncogene* **16**(24):3197-3202.

Liaw D, Marsh DJ, Li J, Dahia PL, Wang SI, Zheng Z, Bose S, Call KM, Tsou HC, Peacocke M, Eng C, Parsons R (1997). Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. *Nature Genetics* **16**(1):64-67.

Licchesi JD, Westra WH, Hooker CM, Herman JG (2008) Promoter hypermethylation of hallmark cancer genes in atypical adenomatous hyperplasia of the lung. *Clinical Cancer Research* **14**(9):2570-2578.

Lim DHK and Maher ER (2010) Genomic Inprinting Syndromes and Cancer. *Advances in Genetics* **70**:145-175.

Lim S, Janzer A, Becker A, Zimmer A, Schüle R, Buettner R, Kirfel J (2010) Lysine-specific demethylase 1 (LSD1) is highly expressed in ER-negative breast cancers and a biomarker predicting aggressive biology. *Carcinogenesis* **31**(3):512-520.

Lin Y, Jiang X, Shen Y, Li M, Ma H, Xing M, Lu Y (2009) Frequent mutations and amplifications of the PIK3CA gene in pituitary tumors. *Endocrine related cancer* **16**(1):301-310.

Liu L, Vo A, McKeehan WL (2005) Specificity of the methylation-suppressed A isoform of candidate tumor suppressor RASSF1 for microtubule hyperstabilization is determined by cell death inducer C19ORF5. *Cancer Research* **65**(5):1830-1838.

Liu S, Ren S, Howell P, Fodstad O and Riker AI (2008) Identification of novel epigenetically modified genes in human melanoma via promoter methylation gene profiling. *Pigment Cell Melanoma Research* **21**:545-558

Lock FE, Underhill-Day N, Dunwell T, Matallanas D, Cooper W, Hesson L, Recino A, Ward A, Pavlova T, Zabarovsky E, Grant MM, Maher ER, Chalmers AD, Kolch W and Latif F (2010) The RASSF8 candidate tumour suppressor inhibits cell growth and regulates the Wnt and NF-kappaB signalling pathways. *Oncogene* **29**(30):4307-4316.

Lofton-Day C, Model F, Devos T, Tetzner R, Distler J, Schuster M, Song X, Lesche R, Liebenberg V, Ebert M, Molnar B, Grützmann R, Pilarsky C, Sledziewski A (2008) DNA methylation biomarkers for blood-based colorectal cancer screening. *Clinical chemistry* **54**(2):414-423.

Logsdon CD, Simeone DM, Binkley C, Arumugam T, Greenson JK, Giordano TJ, Misek DE, Kuick R, Hanash S (2003) Molecular profiling of pancreatic adenocarcinoma and chronic pancreatitis identifies multiple genes differentially regulated in pancreatic cancer. *Cancer Research* **63**:2649-2657

Long NK, Kato K, Ymashita T, Makita H, Toida M, Hatakeyama D, Hara A, Mori H, Shibata T (2008) Hypermethylation of the *RECK* gene predicts poor prognosis in oral squamous cell carcinomas. *Oral Oncology* **44**:1052-1058.

Lowe AW, Olsen M, Hao Y, Lee SP, Taek Lee K, Chen X, van de Rijn M, Brown PO (2007) Gene expression patterns in pancreatic tumours, cells and tissues. *PLoS One* **2**:e323

Lu C, Soria JC, Tang X, Xu XC, Wang L, Mao L, Lotan R, Kemp B, Bekele BN, Feng L, Hong WK, Khuri FR (2004) Prognostic factors in resected stage I non-small-cell lung cancer: a multivariate analysis of six molecular markers. *Journal of Clinical Oncology* **22**(22):4575-4583.

Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, Louis DN, Christiani DC, Settleman J, Haber DA (2004) Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *New England Journal of Medicine* **350**(21):2129-2139

Mahalingam D, Szegezdi E, Keane M, deJong S and Samali A (2009) TRAIL receptor signalling and modulation: Are we on the right TRAIL? *Cancer Treatment Reviews* **35**:280-288.

Malhotra GK, Zhao X, Band H, Band V (2010) Histological, molecular and functional subtypes of breast cancers. *Cancer Biology and Therapy* **10**:955-960.

Malkin D, Li FP, Strong LC, Fraumeni JF Jr, Nelson CE, Kim DH, Kassel J, Gryka MA, Bischoff FZ, Tainsky MA, Friend SH (1990). Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* **250**(4985):1233-8. Erratum in: *Science* (1993) **259**(5097):878

Malumbres M and Barbacid M (2003) RAS oncogenes: the first 30 years. *Nature Reviews Cancer* **10**(3):7-13

Martinez R, Martin-Subero JI, Rohde V, Kirsch M, Alaminos M, Fernandez AF, Ropero S, Shackert G and Esteller M (2009) A microarray-based DNA methylation study of glioblastoma multiforme. *Epigenetics* **4**:255-264.

Martín-Subero JI, Kreuz M, Bibikova M, Bentink S, Ammerpohl O, Wickham-Garcia E, Roslowski M, Richter J, Lopez-Serra L, Ballestar E, Berger H, Agirre X, Bernd H-W, Calvanese V, Cogliatti SB, Drexler HG, Fan J-B, Fraga MF, Hansmann ML, Hummel M, Klapper W, Korn B, Küppers R, MacLeod RAF, Möller P, Ott G, Pott C, Proser F, Rosenwald A, Schwaenen C, Schübeler D, Seifet M, Stürzenhofecker B, Weber m, Wassendorf S, Loeffler M, Trümper L, Stein H, Spang R, Esteller M, Barker D, Hasenclever D and Siebert R (2009a) New insight into the biology and origin of mature aggressive B-cell lymphomas bycombine epigenomic, genomic and transcriptional profiling. *Blood* **113**(11):2488-2497.

Martin-Subero JI, Ammerpohl O, Bibikova M, Wickham-Garcia E, Agirre X, Alvarez S, Brüggemann M, Bug S, Calasanz MJ, Deckert M, Dreyling M, Du MQ, Dürig J, Dyer MJ, Fan JB, Gesk S, Hansmann ML, Harder L, Hartmann S, Klapper W, Küppers R, Montesinos-Rongen M, Nagel I, Pott C, Richter J, Román-Gómez J, Seifert M, Stein H, Suela J, Trümper L, Vater I, Prosper F, Haferlach C, Cruz Cigudosa J and Siebert R (2009b) A comprehensive microarray-based DNA methylation study of 367 hematological neoplasms. *PLoS One* **4**:e6986

Masciari S, Larsson N, Senz J, Boyd N, Kaurah P, Kandel MJ, Harris LN, Pinheiro HC, Troussard A, Miron P, Tung N, Oliveira C, Collins L, Schnitt S, Garber JE, Huntsman D (2007) Germline E-cadherin mutations in familial lobular breast cancer. *Journal of Medical Genetics* **44**(11):726-731.

Matsuda M, Miyagawa K, Takahashi M, Fukuda T, Kataoka T, Asahara T, Inui H, Watatani M, Yasutomi M, Kamada N, Dohi K, Kamiya K (1999) Mutations in the RAD54 recombination gene in primary cancers. *Oncogene* **18**(22):3427-3430

McRonald FE, Morris MR, Gentle D, Winchester L, Baban D, Ragoussis J, Clarke NW, Brown MD, Kishida T, Yao M, Latif F, Maher ER (2009) CpG methylation profiling in VHL related and VHL unrelated renal cell carcinoma. *Molecular Cancer* **8**(31).

Meijers-Heijboer H, van den Ouweland A, Klijn J, Wasielewski M, de Snoo A, Oldenburg R, Hollestelle A, Houben M, Crepin E, van Veghel-Plandsoen M, Elstrodt F, van Duijn C, Bartels C, Meijers C, Schutte M, McGuffog L, Thompson D, Easton D, Sodha N, Seal S,

Barfoot R, Mangion J, Chang-Claude J, Eccles D, Eeles R, Evans DG, Houlston R, Murday V, Narod S, Peretz T, Peto J, Phelan C, Zhang HX, Szabo C, Devilee P, Goldgar D, Futreal PA, Nathanson KL, Weber B, Rahman N, Stratton MR; CHEK2-Breast Cancer Consortium (2002) Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nature Genetics* **31**(1):55-59.

Melki JR, Vincent PC, Clark SJ (1999) Cancer-specific region of hypermethylation identified within the HIC1 putative tumour suppressor gene in acute myeloid leukaemia. *Leukemia* **13**(6):877-883.

Ménard S, Pupa SM, Campiglio M, Tagliabue E (2003) Biologic and therapeutic role of HER2 in cancer. *Oncogene* **22**(42):6570-6578.

Miki Y, Swensen J, Shattuck-Eidens D, Futreal AP, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W, Bell R, Rosenthal J, Hussey C, Tran T, McClure M, Frye C, Hattier T, Phelps R, Haugen-Strano A, Katcher H, Yakumo K, Gholami Z, Shaffer D, Stone S, Bayer S, Wray C, Bogden R, Dayananth R, Ward J, Tonin P, Narod S, Bristow PK, Norris FH, Helvering L, Morrison P, Rosteck P, Lai M, Barrett JC, Lewis C, Neuhausen S,Cannon-Albright L, Goldgar D, Wiseman R, Kamb A, Skolnick MH (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**(5182):66-71.

Milani L, Lundmark A, Kiialainen A, Nordlund J, Flaegstad T, Forestier E, Heyman M, Jonmundsson G, Kanerva J, Schmiegelow K, Söderhäll S, Gustafsson MG, Lönnerholm G, Syvänen AC (2010) DNA methylation for subtype classification and prediction of treatment outcome in patients with childhood acute lymphoblastic leukemia. *Blood* **115**(6):1214-1225.

Milne RL, Knight JA, John EM, Dite GS, Balbuena R, Ziogas A, Andrulis IL, West DW, Li FP, Southey MC, Giles GG, McCredie MR, Hopper JL, Whittemore AS (2005) Oral contraceptive use and risk of early-onset breast cancer in carriers and noncarriers of BRCA1 and BRCA2 mutations. *Cancer Epidemiology Biomarkers and Prevention* **14**(2):350-356.

Mithani SK, Smith IM, Califano JA (2011) Use of integrative epigenetic and cytogenetic analyses to identify novel tumor-suppressor genes in malignant melanoma. *Melanoma Research* **May 20** [epub. ahead of print]

Mongiat M, Ligresti G, Marastoni S, Lorenzon E, Doliana R, Colombatti A (2007) Regulation of the Extrinsic Apoptotic Pathway by the Extracellular Matrix Glycoprotein EMILIN2. *Molecular and Cellular Biology* **27**(20):7176-7187.

Mongiat M, Marastoni S, Ligresti G, Lorenzon E, Schiappacassi M, Perris R, Fustaci S, Colombatti A (2010) The Extracellular Matrix Glycoprotein Elastin Microfibril Interface Located Protein 2: A Dual Role in the Tumour Microenvironment. *Neoplasia* **12**(4):294-304.

Moreno DA, Scrideli CA, Cortez MA, de Paula Queiroz R, Valera ET, da Silva Silveira V, Yunes JA, Brandalise SR, Tone LG (2010) Differential expression of HDAC3, HDAC7 and HDAC9 is associated with prognosis and survival in childhood acute lymphoblastic leukaemia. *British Journal of Haematology* **150**(6):665-673.

Möricke A, Zimmermann M, Reiter A, Gadner H, Odenwald E, Harbott J, Ludwig WD, Riehm H, Schrappe M (2005) Prognostic impact of age in children and adolescents with acute lymphoblastic leukemia: data from the trials ALL-BFM 86, 90, and 95. *Klinische Pädiatrie* **217**(6):310-320.

Morris MR, Ricketts C, Gentle D, Abdulrahman M, Clarke N, Brown M, Kishida T, Yao M, Latif F, Maher ER (2010) Identification of candidate tumour suppressor genes frequently methylated in renal cell carcinoma. *Oncogene* **29**(14):2104-2117.

Morris MR, Ricketts CJ, Gentle D, McRonald F, Carli N, Khalili H, Brown M, Kishida T, Yao M, Banks RE, Clarke N, Latif F and Maher ER (2011) Genome-wide methylation analysis identifies epigenetically inactivated candidate tumour suppressor genes in renal cell carcinoma. *Oncogene* **30**(12):1390-1401.

Morrissey C, Martinez A, Zatyka M, Agathanggelou A, Honorio S, Astuti D, Morgan NV, Moch H, Richards FM, Kishida T, Yao M, Schraml P, Latif F, Maher ER (2001) Epigenetic inactivation of the RASSF1A 3p21.3 tumor suppressor gene in both clear clel and papillary renal cell carcinoma. *Cancer Research* **61**:7277-7281

Mori Y, Cai K, Cheng Y, Wang S, Paun B, Hamilton JP, Jin Z, Sato F, Berki AT, Kan T, Ito T, Mantzur C, Abraham JM and Meltzer SJ (2006) A Genome Wide Search Identifies Epigenetic Silencing of *Somatostatin*, *Tachykinin-1*, and 5 Other Genes in Colon Cancer. *Gastroenterology* **131**:797-808.

Mori K, Enokida H, Kagara I, Kawakami K, Chiyamaru, Tatarano S, Kawahara K, Nishiyama K, Seki N and Nakagawa M (2009) CpG hypermethylation of *collagen type I α 2* contributes to proliferation and migration activity of human bladder cancer. *International Journal of Oncology* **34**:1593-1602.

Mosquera JM, Mehra R, Regan MM, Perner S, Genega EM, Bueti G, Shah RB, Gaston S, Tomlins SA, Wei JT, Kearney MC, Johnson LA, Tang JM, Chinnaiyan AM, Rubin MA, Sanda MG (2009) Prevalence of TMPRSS2-ERG fusion prostate cancer among men undergoing prostate biopsy in the United States. *Clinical Cancer Research* **15**(14):4706-4711.

Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Fewell GA, Flicek P,

Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigó R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, Karolchik D, Kasprzyk A, Kawai J, Keibler E, Kells C, Kent WJ, Kirby A, Kolbe DL, Korf I, Kucherlapati RS, Kulbokas EJ, Kulp D, Landers T, Leger JP, Leonard S, Letunic I, Levine R, Li J, Li M, Lloyd C, Lucas S, Ma B, Maglott DR, Mardis ER, Matthews L, Mauceli E, Mayer JH, McCarthy M, McCombie WR, McLaren S, McLay K, McPherson JD, Meldrim J, Meredith B, Mesirov JP, Miller W, Miner TL, Mongin E, Montgomery KT, Morgan M, Mott R, Mullikin JC, Muzny DM, Nash WE, Nelson JO, Nhan MN, Nicol R, Ning Z, Nusbaum C, O'Connor MJ, Okazaki Y, Oliver K, Overton-Larty E, Pachter L, Parra G, Pepin KH, Peterson J, Pevzner P, Plumb R, Pohl CS, Poliakov A, Ponce TC, Ponting CP, Potter S, Quail M, Reymond A, Roe BA, Roskin KM, Rubin EM, Rust AG, Santos R, Sapojnikov V, Schultz B, Schultz J, Schwartz MS, Schwartz S, Scott C, Seaman S, Searle S, Sharpe T, Sheridan A, Shownkeen R, Sims S, Singer JB, Slater G, Smit A, Smith DR, Spencer B, Stabenau A, Stange-Thomann N, Sugnet C, Suyama M, Tesler G, Thompson J, Torrents D, Trevaskis E, Tromp J, Ucla C, Ureta-Vidal A, Vinson JP, Von Niederhausern AC, Wade CM, Wall M, Weber RJ, Weiss RB, Wendl MC, West AP, Wetterstrand K, Wheeler R, Whelan S, Wierzbowski J, Willey D, Williams S, Wilson RK, Winter E, Worley KC, Wyman D, Yang S, Yang SP, Zdobnov EM, Zody MC, Lander ES (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**(6915):520-562.

Muggerud AA, Rønneberg JA, Wärnberg F, Botling J, Busato F, Jovanovic J, Solvang H, Bukholm I, Børresen-Dale AL, Kristensen VN, Sørlie T, Tost J (2010) Frequent aberrant DNA methylation of ABCB1, FOXC1, PPP2R2B and PTEN in ductal carcinoma in situ and early invasive breast cancer. *Breast Cancer Research* **12**(1):R3.

Müller HM, Oberwalder M, Fiegl H, Morandell M, Goebel G, Zitt M, Mühlthaler M, Ofner D, Margreiter R, Widschwendter M (2004) Methylation changes in faecal DNA: a marker for colorectal cancer screening? *Lancet* **363**(9417):1283-1285.

Mutter GL, Baak JP, Fitzgerald JT, Gray R, Neuberg D, Kust GA, Gentleman R, Gullans SR, Wei LJ, Wilcox M (2001) Global expression changes of constitutive and hormonally regulated genes during endometrial neoplastic transformation. *Gynecologic Oncology* **83**:177-185

Mutze K, Langer R, Becker K, Ott K, Novotny A, Luber B, Hapfelmeier A, Göttlicher M, Höfler H, Keller G (2010) Histone deacetylase (HDAC) 1 and 2 expression and chemotherapy in gastric cancer. *Annals of Surgical Oncology* **17**(12):3336-3343

Nachman JB, Heerema NA, Sather H, Camitta B, Forestier E, Harrison CJ, Dastugue N, Schrappe M, Pui CH, Basso G, Silverman LB, Janka-Schaub GE (2007) Outcome of treatment in children with hypodiploid acute lymphoblastic leukemia. *Blood* **110**(4):1112-1115.

Nair SS, Coolen MW, Strizaker C, Song JZ, Statham AL, Strbenac D, Robinson MD, Clark SJ (2011) Comparison of methyl-DNA immunoprecipitation (MeDIP) and methyl-

CpG binding domain (MBD) protein capture for genome-wide DNA methylation analysis reveal CpG sequence coverage bias. *Epigenetics* **6**(1):34-44.

Nishiyama H, Gill JH, Pitt E, Kennedy W, Knowles M (2001) Negative regulation of G_1/S transition by the candidate bladder tumour suppressor gene *DBCCR1*. *Oncogene* **20**:2956-2964.

Nakamura M, Yonekawa Y, Kleihues P, Ohgaki H (2001) Promoter hypermethylation of the RB1 gene in glioblastomas. *Laboratory investigation: a journal of technical methods and pathology* **81**(7):77-82.

Nojima M, Suzuki H, Toyota M, Watanabe Y, Maruyama R, Sasaki S, Sasaki Y, Mita H, Nishikawa N, Yamaguchi K, Hirata K, Itoh F, Tokino T, Mori M, Imai K, Shinomura Y (2007) Frequent epigenetic inactivation of SFRP genes and constitutive activation of Wnt signaling in gastric cancer. *Oncogene* **26**(32):4699-4713.

Noushmehr H, Weisnberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, Pan F, Pelloski CE, Sulman EP, Bhat KP, Verhaak RGW, Hoadley KA, Hayes DN, Perou CM, Schmidt HK, Ding L, Wilson RK, Van Den Berg D, Shen H, Bengtsson H, Neuvial, Cope LM, Buckley J, Herman JG, Baylin SB, Laird PW, Aldape K and The Cancer Genome Atlas Research Network (2010) Identification of a CpG Island Methylator Phenotype that Defines a Distinct Subgroup of Glioma. *Cancer Cell* **17**:510-522.

Nowee ME, Snijders AM, Rockx DA, de Wit RM, Kosma VM, Hämäläinen K, Schouten JP, Verheijen RH, van Diest PJ, Albertson DG, Dorsman JC (2007) DNA profiling of primary serous ovarian and fallopian tube carcinomas with array comparative genomic hybridization and multiplex ligation-dependent probe amplification. *The Journal of Pathology* **213**(1):46-55.

Nowell P and Hungerford D (1960) A minute chromosome in human chronic granulocytic leukaemia. *Science* **132**:1497.

Oh J, Takahashi R, Kondo S, Mizoguchi A, Adachi E, Sasahara RM, Nishimura S, Imamura Y, Kitayama A, Alexander DB, Ide C, Horan TP, Arakawa T, Yoshida H, Nishikawa S, Itoh Y, Seiki M, Itohara S, Tkahashi C and Noda M (2001) The membrane-anchored MMP inhibitor RECK is a key regulator of extracellular matrix integrity and angiogenesis. *Cell* **107**:789-800.

Okano M, Bell DW, Haber DA and Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell* **99**: 247-257.

Ongenaert M, Wisman GBA, Volders HH, Koning AJ, van der Zee A, van Criekinge W and Schuuring E (2008) Discovery of DNA methylation markers in cervical cancer using relaxation ranking. *BMC Medical Genomics* **1**:57

O'Riain C, O'Shea DM, Yang Y, Le Dieu R, Gribben JG, Summers K, Yeboah-Afari J, Bhaw-Rosun L, Fleischmann C, Mein CA, Crook T, Smith P, Kelly G, Rosenwald A, Ott

G, Campo E, Rimsza LM, Smeland EB, Chan WC, Johnson N, Gascoyne RD, Reimer S, Braziel RM, Wright GW, Staudt LM, Lister TA, Fitzgibbon J (2009) Array-based DNA methylation profiling in follicular lymphoma. *Leukaemia* **23**(10):1858-1866.

Ortiz-Vega S, Khokhlatchev A, Nedwidek M, Zhang XF, Dammann R, Pfeifer GP, Avruch J (2002) The putative tumor suppressor RASSF1A homodimerizes and heterodimerizes with the Ras-GTP binding protein Nore1. *Oncogene* **21**(9):1381-1390.

Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, Naoki K, Sasaki H, Fujii Y, Eck MJ, Sellers WR, Johnson BE, Meyerson M (2004) EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304**(5676):1497-1500.

Pan D (2010) The Hippo Signalling Pathway in Development and Cancer. *Developmental Cell* **19**:491-505.

Pandis N, Bardi G, Mitelman F, Heim S (1997) Deletion of the short arm of chromosome 3 in breast tumors. *Genes, chromosomes and cancer* **18**(4):241-245.

Paredes J, Albergaria A, Oliveira JT, Jerónimo C, Milanezi F, Schmitt FC (2005) P-cadherin overexpression is an indicator of clinical outcome in invasive breast carcinomas and is associated with CDH3 promoter hypomethylation. *Clinical Cancer Research* **11**(15):5869-5877.

Park HW, Kang HC, Kim IJ, Jang SG, Kim K, Yoon HJ, Jeong SY, Park JG (2007) Correlation between hyperemthylation of the RASSF2A promoter and k-ras/BRAF mutations in microsatellite-stable colorectal cancers. *International Journal of Cancer* **120**:7-12

Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, Palazzo J, Marron JS, Nobel AB, Mardis E, Nielsen TO, Ellis MJ, Perou CM, Bernard PS (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology* **27**(8):1160-1167.

Parsons DW, Li M, Zhang X, Jones S, Leary RJ, Lin JC, Boca SM, Carter H, Samayoa J, Bettegowda C, Gallia GL, Jallo GI, Binder ZA, Nikolsky Y, Hartigan J, Smith DR, Gerhard DS, Fults DW, VandenBerg S, Berger MS, Marie SK, Shinjo SM, Clara C, Phillips PC, Minturn JE, Biegel JA, Judkins AR, Resnick AC, Storm PB, Curran T, He Y, Rasheed BA, Friedman HS, Keir ST, McLendon R, Northcott PA, Taylor MD, Burger PC, Riggins GJ, Karchin R, Parmigiani G, Bigner DD, Yan H, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE (2011) The genetic landscape of the childhood cancer medulloblastoma. *Science* **331**(6016):435-439.

Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL, Olivi A, McLendon R, Rasheed BA, Keir S, Nikolskaya T, Nikolsky Y, Busam DA, Tekleab H, Diaz LA Jr, Hartigan J, Smith DR, Strausberg RL, Marie SK, Shinjo SM, Yan H, Riggins GJ, Bigner DD, Karchin R, Papadopoulos N, Parmigiani G,

Vogelstein B, Velculescu VE, Kinzler KW. (2008) An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**:1807-1812

Paulsson K, Johansson B (2009) High hyperdiploid childhood acute lymphoblastic laukemia. *Genes chromosomes and cancer* **48**(8):637-660.

Pehlivan S, Artac M, Sever T, Bozcuk H, Kilincarslan C, Pehlivan M (2010) Gene methylation of SFRP2, P16, DAPK1, HIC1, and MGMT and KRAS mutations in sporadic colorectal cancer. *Cancer genetics and cytogenetics* **201**(2):128-132

Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO, Botstein D (2000) Molecular portraits of human breast tumours. *Nature* **406**(6797):747-752.

Pfeifer GP, Dammann R and Tommasi S (2010) RASSF proteins. *Current Biology* **20**(8):R344-R345

Phelan CM, Lancaster JM, Tonin P, Gumbs C, Cochran C, Carter R, Ghadirian P, Perret C, Moslehi R, Dion F, Faucher MC, Dole K, Karimi S, Foulkes W, Lounis H, Warner E, Goss P, Anderson D, Larsson C, Narod SA, Futreal PA (1996) Mutation analysis of the BRCA2 gene in 49 site-specific breast cancer families. *Nature Genetics* **13**(1):120-122.

Polesello C, Huelsmann S, Brown NH, Tapon N (2006) The Drosophila RASSF homolog antagonizes the hippo pathway. *Current Biology* **16**(24):2459-2465.

Ponting CP, Benjamin DR (1996) A novel family of Ras-binding domains. *Trends in Biochemical Sciences* **21**(11):422-425.

Pui CH, Chessells JM, Camitta B, Baruchel A, Biondi A, Boyett JM, Carroll A, Eden OB, Evans WE, Gadner H, Harbott J, Harms DO, Harrison CJ, Harrison PL, Heerema N, Janka-Schaub G, Kamps W, Masera G, Pullen J, Raimondi SC, Richards S, Riehm H, Sallan S, Sather H, Shuster J, Silverman LB, Valsecchi MG, Vilmer E, Zhou Y, Gaynon PS, Schrappe M (2003) Clinical heterogeneity in childhood acute lymphoblastic leukemia with 11q23 rearrangements. *Leukemia* **17**(4):700-706.

Pui C-H, Robison LL and Look AT (2008) Acute lymphoblastic leukaemia. *Lancet* **371**:1030-1043.

Prabhu JS, Korlimala A, Banerjee A, Wani S, K P, Sahoo R (2009**)** Gene-specific methylation: potential markers for colorectal cancer. *International Journal of Biological Markers* **24**(1):57-62.

Praskova M, Khoklatchev A, Ortiz-Vega S, Avruch J (2004) Regulation of the MST1 kinase by autophosphorylation, by the growth inhibitory proteins, RASSF1 and NORE1, and by Ras. *The Biochemical Journal* **381**:453-462.

Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, He X, Perou CM (2010) Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Research* **12**(5):R68

Prosperi JR, Goss KH (2010) A Wnt-ow of opportunity: targeting the Wnt/beta-catenin pathway in breast cancer. *Current Drug Targets* **11**(9):1074-1088.

Quina AS, Buschbeck M, Di Croce L (2006) Chromatin structure and epigenetics. *Biochemical Pharmacology* **72**(11):1563-1569

Rabizadeh S, Xavier RJ, Ishiguro K, Bernabeortiz J, Lopez-Ilasaca M, Khokhlatchev A, Mollahan P, Pfeifer GP, Avruch J, Seed B (2004) The scaffold protein CNK1 interacts with the tumor suppressor RASSF1A and augments RASSF1A-induced cell death. *Journal of Biological Chemistry* **279**(28):29247-29254.

Radisky ES, Radisky DC (2010) Matrix metalloproteinase-induced epithelial-mesenchymal transition in breast cancer. *Mammary Gland Biology and Neoplasia* **15**(2):201-212.

Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, Reid S, Spanova K, Barfoot R, Chagtai T, Jayatilake H, McGuffog L, Hanks S, Evans DG, Eccles D; Breast Cancer Susceptibility Collaboration (UK), Easton DF, Stratton MR (2007) PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nature Genetics* **39**(2):165-167.

Ramaswamy B, Majumder S, Roy S, Ghoshal K, Kutay H, Datta J, Younes M, Shapiro CL, Motiwala T, Jacob ST (2009) Estrogen-mediated suppression of the gene encoding protein tyrosine phosphatase PTPRO in human breast cancer: mechanism and role in tamoxifen sensitivity.*Molecular Endocrinology* **Feb 23**(2):176-87.

Rasheed BK, McLendon RE, Friedman HS, Friedman AH, Fuchs HE, Bigner DD, Bigner SH (1995) Chromosome 10 deletion mapping in human gliomas: a common deletion region in 10q25. *Oncogene* **10**(11):2243-2246.

Rauch T and Pfeifer GP (2005) Methylated CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer. *Laboratory Investigation* **85**:1172-1180.

Rauch T, Li H, Wu X, Pfeifer GP (2006) MIRA assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain containing genes in lung cancer cells. *Cancer Research* **66**:7939-7947.

Rauch T, Wang Z, Zhang X, Zhong X, Wu X, Lau SK, Kemstine KH, Riggs AD, Pfeifer GP (2007) Homeobox gene methylation in lung cancer studied by genome wide analysis with a microarray based methylated CpG island recovery assay. *PNAS* **104**:5527-5532.

Recino A, Sherwood V, Flaxman A, Cooper WN, Latif F, Ward A, Chalmers AD (2010) Human RASSF7 regulates the microtubule cytoskeleton and is required for spindle

formation, Aurora B activation and chromosomal congression during mitosis. *The Biochemical Journal* **430**(2):207-213.

Reik W, Santos F, Dean W (2003) Mammalian epigenomics: reprogramming the genome for development and therapy. *Theriogenology* **59**(1):21-32

Rhee I, Bachman KE, Park BH, Jair K-W, Yen R-W C, Schuebel KE, Cui H, Feinberg AP, Lengauer C, Kinzler KW, Baylin SB, Volgelstein B (2002) DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature* **416**:552-556.

Richter AM, Pfeifer GP, Dammann RH (2009) The RASSF proteins in cancer; from epigenetic silencing to functional characterization. *Biochimica at Biophysica Acta* **1796**:114-128

Robert MF, Morin S, Beaulieu N, Gauthier F, Chute IC, Barsalou A, MacLeod AR (2003) DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells. *Nature Genetics* **33**(1):61-65.

Rocco JW, Sidransky D (2001) p16(MTS-1/CDKN2/INK4a) in cancer progression. *Experimental Cell Research*. 2001 **264**(1):42-55.

Rodriguez-Viciana P, Sabatier C and McCormick (2004) Signalling specificity by Ras family GTPases is determined by the full spectrum of effectors they regulate. *Molecular and Cellular Biology* **24**:4943-4954.

Rodríguez-Paredes M and Esteller M (2011) Cancer Epigenetics reaches mainstream oncology. *Nature Medicine* **17**(3):330-339.

Roman-Gomez J, Castillejo JA, Jimenez A, Gonzalez MG, Moreno F, Rodriguez Mdel C, Barrios M, Maldonado J, Torres A (2002) 5' CpG island hypermethylation is associated with transcriptional silencing of the p21(CIP1/WAF1/SDI1) gene and confers poor prognosis in acute lymphoblastic leukemia. *Blood* **99**(7):2291-2296.

Roman-Gomez J, Jimenez-Velasco A, Castillejo JA, Agirre X, Barrios M, Navarro G, Molina FJ, Calasanz MJ, Prosper F, Heiniger A and Torres A (2004) Promoter hypermethylation of cancer-related genes: a strong independent prognostic factpr in acute lymphoblastic leukemia. *Blood* **104**:2429-2498.

Roman-Gomez J, Jimenez-Velasco A, Agirre X, Prosper F, Heiniger A, Torres A (2005) Lack of CpG island methylator phenotype defines a clinical subtype of T-cell acute lymphoblastic leukemia associated with good prognosis. *Journal of Clinical Oncology* **23**(28):7043-7049.

Roman-Gomez J, Jimenez-Velasco A, Agirre X, Castillejo JA, Navarro G, Calasanz MJ, Garate L, San Jose-Eneriz E, Cordeu L, Prosper F, Heiniger A and Torres A, (2006) CpG island Methylator Phenotype Redefines the Prognostic Effect of t(12;21) in Childhood Acute Lymphoblastic Leukemia. *Clinical Cancer Research* **12**(16):4845-4850.

Rouleau GA, Merel P, Lutchman M, Sanson M, Zucman J, Marineau C, Hoang-Xuan K, Demczuk S, Desmaze C, Plougastel B, Pulstsm SM, Lenoir G, Bijlsma E, Fashold R, Dumanski J, De Jong P, Parry D, Eldrige R, Aurias A, Delattre O, Thomas G (1993) Alteration in a new gene encoding a putative membrane-organizing protein causes neuro-fibromatosis type 2. *Nature* **363**(6429):515-521.

Ruike Y, Imanaka Y, Sato F, Shimizu K, Tsujimoto G (2010) Genome-wide analysis of aberrant methylation in human breast cancer cells using methyl-DNA immunoprecipitation combined with high-throughput sequencing. *BMC Genomics* **11**:137.

Rutter JL, Smith AM, Dávila MR, Sigurdson AJ, Giusti RM, Pineda MA, Doody MM, Tucker MA, Greene MH, Zhang J, Struewing JP (2003) Mutational analysis of the BRCA1-interacting genes ZNF350/ZBRK1 and BRIP1/BACH1 among BRCA1 and BRCA2-negative probands from breast-ovarian cancer families and among early-onset breast cancer cases and reference individuals. *Human Mutation* **22**(2):121-128

Sadr-Nabavi A, Ramser J, Volkmann J, Naehrig J, Wiesmann F, Betz B, Hellebrand H, Engert S, Seitz S, Kreutzfeld R, Sasaki T, Arnold N, Schmutzler R, Kiechle M, Niederacher D, Harbeck N, Dahl E, Meindl A (2009) Decreased expression of angiogenesis antagonist EFEMP1 in sporadic breast cancer is caused by aberrant promoter methylation and points to an impact of EFEMP1 as molecular biomarker. *International Journal of Cancer* **124**:1727-1735.

Sakai T, Toguchida J, Ohtani N, Yandell DW, Rapaport JM, Dryja TP (1991) Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene. *American Journal of Genetics* **48**(5):880-888.

Sakuma T, Uzawa K, Onda T, Shiiba M, Yokoe H, Shibahara T, Tanzawa H (2006) Aberrant expression of histone deacetylase 6 in oral squamous cell carcinoma. *International Journal of Oncology* **29**(1):117-124.

Samuels Y, Wang Z, Bardelli A, Silliman N, Ptak J, Szabo S, Yan H, Gazdar A, Powell SM, Riggins GJ, Willson JK, Markowitz S, Kinzler KW, Vogelstein B, Velculescu VE (2004) High frequency of mutations of the PIK3CA gene in human cancers. *Science* **304**(5670):554

San José-Enériz E, Agirre X, Román-Gómez J, Cordeu L, Garate L, Jiménez-Velasco A, Vázquez I, Calasanz MJ, Heiniger A, Torres A, Prósper F (2006) Downregulation of *DBC1* expression in acute lymphoblastic leukaemia is mediated by aberrant methylation of its promoter. *British Journal of Haematology* **134**:137-144.

Sandoval J, Heyn HA, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M (2011) Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* **6**(6) [epub. ahead of print].

Sarraf SA, Stancheva I (2004) Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Molecular Cell* **15**(4):595-605.

Sasaki H, Kawano O, Endo K, Suzuki E, Yukiue H, Kobayashi Y, Yano M, Fujii Y (2007) Human MOB1 expression in non-small-cell lung cancer. *Clinical Lung Cancer* **8**(4):273-276.

Savitsky K, Bar-Shira A, Gilad S, Rotman G, Ziv Y, Vanagaite L, Tagle DA, Smith S, Uziel T, Sfez S, Ashkenazi M, Pecker I, Frydman M, Harnik R, Patanjali SR, Simmons A, Clines GA, Sartiel A, Gatti RA, Chessa L, Sanal O, Lavin MF, Jaspers NG, Taylor AM, Arlett CF, Miki T, Weissman SM, Lovett M, Collins FS, Shiloh Y (1995) A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* **268**(5218):1749-53.

Sawan C and Herceg Z (2010) Histone Modifications and Cancer. *Advances in Genetics* **70**:57-85.

Schagdarsurengin U, Richter AM, Hornung J, Lange C, Steinmann K and Dammann RH (2010) Frequent epigenetic inactivation of RASSF2 in thyroid cancer and functional consequences. *Molecular Cancer* **9**:264

Scheel H, Hofmann K (2003) A novel interaction motif, SARAH, connects three classes of tumor suppressor. *Current Biology* **13**(23):R899-900.

Schiffman JD, Hodgson JG, VandenBerg SR, Flaherty P, Polley MY, Yu M, Fisher PG, Rowitch DH, Ford JM, Berger MS, Ji H, Gutmann DH, James CD (2010) Oncogenic BRAF mutation with CDKN2A inactivation is characteristic of a subset of pediatric malignant astrocytomas. *Cancer research* **70**(1):512-519.

Schrager CA, Schneider D, Gruener AC, Tsou HC, Peacocke M (1998) Clinical and pathological features of breast disease in Cowden's syndrome: an underrecognized syndrome with an increased risk of breast cancer. *Human Pathology* **29**(1):47-53.

Schulmann K, Sterian A, Berki A, Yin J, Sato F, Xu Y, Olaru A, Wang S, Mori Y, Deacu E, Hamilton J, Kan T, Krasna MJ, Beer DG, Pepe MS, Abraham JM, Feng Z, Schmiegel W, Greenwald BD, Meltzer SJ (2005) Inactivation of p16, RUNX3, and HPP1 occurs early in Barrett's-associated neoplastic progression and predicts progression risk. *Oncogene* **24**(25):4138-4148.

Schulte JH, Lim S, Schramm A, Friedrichs N, Koster J, Versteeg R, Ora I, Pajtler K, Klein-Hitpass L, Kuhfittig-Kulle S, Metzger E, Schüle R, Eggert A, Buettner R, Kirfel J (2009) Lysine-specific demethylase 1 is strongly expressed in poorly differentiated neuroblastoma: implications for therapy. *Cancer Research* **69**(5):2065-2071.

Scott SP, Bendix R, Chen P, Clark R, Dork T, Lavin M F (2002) Missense mutations but not allelic variants alter the function of ATM by dominant interference in patients with breast cancer. *PNAS* **99**:925-930.

Seal S, Thompson D, Renwick A, Elliott A, Kelly P, Barfoot R, Chagtai T, Jayatilake H, Ahmed M, Spanova K, North B, McGuffog L, Evans DG, Eccles D; Breast Cancer Susceptibility Collaboration (UK), Easton DF, Stratton MR, Rahman N (2006) Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nature Genetics* **38**(11):1239-1241.

Seidel C, Schagdarsurengin U, Blümke K, Würl P, Pfeifer GP, Hauptmann S, Taubert H, Dammann R (2007) Frequent hypermethylation of MST1 and MST2 in soft tissue sarcoma. *Molecular Carcinogenesis* **46**(10):865-871.

Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A, Delaney A, Gelmon K, Guliany R, Senz J, Steidl C, Holt RA, Jones S, Sun M, Leung G, Moore R, Severson T, Taylor GA, Teschendorff AE, Tse K, Turashvili G, Varhol R, Warren RL, Watson P, Zhao Y, Caldas C, Huntsman D, Hirst M, Marra MA, Aparicio S (2010) Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**(7265):809-813.

Shames DS, Girard L, Gao B, Sato M, Lewis CM, Shivapurkar N, Jiang A, Perou CM, Kim YH, Pollack JR, Fong KM, Lam CL, Wong M, Shyr Y, Nanda R, Olopade OI, Gerald W, Euhus DM, Shay JW, Gazdar AF, Minna JD (2006) A genome-wide screen for promoter methylation in lung cancer identifies novel methylation markers for multiple malignancies. *PloS Medicine* **3**(12):e486.

Shao C, Sun W, Tan M, Glazer CA, Bhan S, Zhong X, Fakhry C, Sharma R, Westra WH, Hoque MO, Moskaluk CA, Sidransky D, Califano JA, Ha PK (2011) Integrated, genome-wide screening for hypomethylated oncogenes in salivary gland adenoid cystic carcinoma. *Clinical Cancer Research* **May 6** [Epub ahead of print]

Shapiro R, Servis RE, Welcher M (1970) Reactions of uracil and cytosine derivatives with sodium bisulphite. A chemical deamination method. *Journal of the Americal Chemical Society* **92:**422-424.

Sharma G, Mirza S, Parshad R, Srivastava A, Datta Gupta S, Pandya P, Ralhan R (2010) CpG hypomethylation of MDR1 gene in tumor and serum of invasive ductal breast carcinoma patients. *Clinical Biochemistry* **43**(4-5):373-379.

Sherwood V, Manbodh R, Sheppard C, Chalmers AD (2008) RASSF7 is a member of a new family of RAS association domain-containing proteins and is required for completing mitosis. *Molecular Biology of the Cell* **19**:1772-1782

Sherwood V, Recino A, Jeffires A, Ward A, Chlamers AD (2010) The N-terminal RASSF family: a new group of Ras-association-domain-containing proteins, with emerging links to cancer formation. *The Biochemical Journal* **425**:303-311.

Shi L, Cheng Z, Zhang J, Li R, Zhao P, Fu Z, Yu Y (2008) hsa-mir-181a and hsa-mir-181b function as tumour suppressors in human glioma cells. *Brain Research* **1236**:185-193.

Shimozaki K, Namihira M, Nakashima K, Taga T (2005) Stage- and site-specific DNA demethylation during neural cell development from embryonic stem cells. *Journal of Neurochemistry* **93**(2):432-439.

Shivapurkar N, Toyooka S Toyooka KO, Reddy J, Miyajima K, Suzuki M, Shigematsu H, Takahashi T, Parikh G, Pass HI, Chaudhary PR and Gazdar AK (2004) Aberrant methylation of TRAIL decoy receptor genes is frequent in multiple tumour types. *International Journal of Cancer* **109**:786-792.

Shivakumar L, Minna J, Sakamaki T, Pestell R, White MA (2002) The RASSF1A tumor suppressor blocks cell cycle progression and inhibits cyclin D1 accumulation. *Molecular and Cellular Biology* **22**(12):4309-4318.

Shurtleff SA, Buijs A, Behm FG, Rubnitz JE, Raimondi SC, Hancock ML, Chan GC, Pui CH, Grosveld G, Downing JR (1995) TEL/AML1 fusion resulting from a cryptic t(12;21) is the most common genetic lesion in pediatric ALL and defines a subgroup of patients with an excellent prognosis. *Leukemia* **9**(15):1985-1989.

Singer-Sam J, LeBon JM, Tanguay RL, Riggs AD (1990) A quantitative HpaII-PCR assay to measure methylation of DNA from a small number of cells. *Nucleic Acid Research* **18**(3):687.

Singh P, Yang M, Dai H, Yu D, Huang Q, Tan W, Kernstine KH, Lin D, Shen B (2008) Overexpression and hypomethylation of flap endonuclease 1 gene in breast and other cancers. *Molecular Cancer Research* **6**(11):1710-1717.

Sjöblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J, Dawson D, Willson JK, Gazdar AF, Hartigan J, Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE (2006) The consensus coding sequences of human breast and colorectal cancers. *Science* **314**(5797):268-274.

Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A and McGuire WL (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235**:177-182.

Smiraglia DJ, Rush LJ, Frühwald MC, Dai Z, Held WA, Costello MA, Lang JC, Eng C, Li B, Wright FA, Caligiuri MA and Plass C (2001) Excessive CpG island hypermethylation in cancer cell lines versus primary human malignancies. *Human Molecular Genetics* **10**(13):1413-1419.

Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, Bando M, Ohno S, Ishikawa Y, Aburatani H, Niki T, Sohara Y, Sugiyama Y, Mano H (2007) Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* **448**(7153):561-566.

Son KS, Kang HS, Kim SJ, Jung SY, Min SY, Lee SY, Kim SW, Kwon Y, Lee KS, Shin KH, Ro J (2010) Hypomethylation of the interleukin-10 gene in breast cancer tissues. *Breast* **19**(6):484-488.

Song CX, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen CH, Zhang W, Jian X, Wang J, Zhang L, Looney TJ, Zhang B, Godley LA, Hicks LM, Lahn BT, Jin P, He C (2011) Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature Biotechnology* **29**(1):68-72.

Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lønning P, Børresen-Dale AL (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *PNAS* **98**(19):10869-10874.

Sørlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lønning PE, Brown PO, Børresen-Dale AL, Botstein D (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. *Nature* **100**(8):8418-8423.

Span PN, Sweep CGJ, Manders P, Beex LVAM, Leppert D and Lindberg RLP (2003) Matrix Metalloproteinase Inhibitor Reversion-Inducing Cysteine-Rich Protein with Kazal Motifs A Prognositc Marker for Good Clinical Outcome in Human Breast Carcinoma. *Cancer* **97(**11):2710-2715.

Stehelin D, Varmus HE, Bishop JM, Vogt PK (1976) DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature* **260**(5547):170-173.

Stephens P, Edkons A, Davies H, Greenman C, Cox C, Hunter C, Bignell G, Teague J, Smith R, Stevens C, O'Meara S, Parker A, Tarpey P, Avis T, Barthorpe A, Brackenbury L, Buck G, Butler A, Clements J, Cole J, Dicks E, Edwards K, Forbes S, Gorton M, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jones D, Kosmidou V, Laman R, Lugg R, Menzies A, Perry J, Pretty R, Raine K, Shepherd R, Small Solomon H, Stephens Y, Tofts C, Varian J, Webb A, West S, Widaa S, Yates A, Brasseur F, Cooper CS, Flanagan AM, Green A, Knowles M, Leung SY, Looijenga LHJ, Malkowicz B, Pierotti MA, Teh B, Yuen ST, Nicholson AG, Lakhani S, Easton DF, Weber BL, Stratton MR, Futreal PA and Wooster R (2005) A screen of the complete protein kinase family identifies diverse patterns of somatic mutations in human breast cancer. *Nature Genetics* **37**(6):590-592.

Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT, Stebbings LA, Leroy C, Edkins S, Mudie LJ, Greenman CD, Jia M, Latimer C, Teague JW, Lau KW, Burton J, Quail MA, Swerdlow H, Churcher C, Natrajan R, Sieuwerts AM, Martens JW, Silver DP, Langerød A, Russnes HE, Foekens JA, Reis-Filho JS, van 't Veer L, Richardson AL, Børresen-Dale AL, Campbell PJ, Futreal PA, Stratton MR (2009) Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462**(7276):1005-1010.

Stanbery L, D'Silva NJ, Lee JS, Bradford CR, Carey TE, Prince ME, Wolf GT, Worden FP, Cordell KG, Petty EM (2010) High SEPT9_v1 Expression Is Associated with Poor Clinical Outcomes in Head and Neck Squamous Cell Carcinoma. *Translational Oncology* **3**(4):239-245

Strachan T and Read AP (2004) Human Molecular Genetics. *Garland Science*

Strathdee G, Appleton K, Illand M, Millan DW, Sargent J, Paul J, Brown R (2001) Primary ovarian carcinomas display multiple methylator phenotypes involving known tumor suppressor genes. *American Journal of Pathology* **158**(3):1121-1127.

Suganuma T, Kawabata M, Ohshima T, Ikeda MA (2002) Growth suppression of human carcinoma cells by reintroduction of the p300 coactivator. *PNAS* **99**(20):13073-13078

Swanberg SE, Nagarajan RP, Peddada S, Yasui DH, LaSalle JM (2009). Reciprocal co-regulation of EGR2 and MECP2 is disrupted in Rett syndrome and autism. *Human Molecular Genetics* **18**(3):525-34.

Sy SM, Huen MS, Chen J (2009) PALB2 is an integral component of the BRCA complex required for homologous recombination repair. *PNAS* **106**(17):7155-7160.

Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H (2010) Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucelic Acids Research* **38**(19):e181.

Takahashi C, Sheng Z, Horan TP, Kitayama H, Maki M, Hitomi K, Kitaura Y, Takai S, Sasahara RM, Horimoto A, Ikawa Y, Ratzkin BJ, Arakawa T, Noda M, (1998) Regulation of matrix metalloproteinase-9 and inhibition of tumour invasion by the membrane anchored glycoprotein RECK. *PNAS* **95**:13221-13226.

Takahashi Y, Miyoshi Y, Takahata C, Irahara N, Taguchi T, Tamaki Y, Noguchi S (2005) Down-regulation of LATS1 and LATS2 mRNA expression by promoter hypermethylation and its association with biologically aggressive phenotype in human breast cancers. *Clinical Cancer Research* **11**(4):1380-1385.

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Lui DR, Aravind L, Rao A (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**(5929):930-935.

Tan DS, Lambros MB, Rayter S, Natrajan R, Vatcheva R, Gao Q, Marchiò C, Geyer FC, Savage K, Parry S, Fenwick K, Tamber N, Mackay A, Dexter T, Jameson C, McCluggage WG, Williams A, Graham A, Faratian D, El-Bahrawy M, Paige AJ, Gabra H, Gore ME, Zvelebil M, Lord CJ, Kaye SB, Ashworth A, Reis-Filho JS (2009) PPM1D is a potential therapeutic target in ovarian clear cell carcinomas. *Clinical Cancer Research* **15**:2269-2280.

Tao W, Zhang S, Turenchalk GS, Stewart RA, St John MA, Chen W, Xu T (1999) Human homologue of the Drosophila melanogaster lats tumour suppressor modulates CDC2 activity. *Nature Genetics* **21**(2):177-181.

Tapon N, Harvey KF, Bell DW, Wahrer DC, Schiripo TA, Haber DA, Hariharan IK (2002) salvador Promotes both cell cycle exit and apoptosis in Drosophila and is mutated in human cancer cell lines. *Cell* **110**(4):467-478.

The CHEK2 Breast Cancer Case–Control Consortium (2004) *American Journal of Human Genetics* **74**:1175–1182.

Tillinghast GW, Partee J, Albert P, Kelley JM, Burtow KH, Kelly K (2003) Analysis of genetic stability at the EP300 and CREBBP loci in a panel of cancer cell lines. *Genes Chromosomes and Cancer* **37**(2):121-131.

Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM (2005) Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**(5748):644-648.

Tommasi S, Dammann R, Jin SG, Zhang XF, Avruch J, Pfeifer GP (2002) RASSF3 and NORE1: identification and cloning of two human homologues of the putative tumor suppressor gene RASSF1. *Oncogene* **21**(17):2713-2720.

Tommasi S, Karm DL, Wu X, Yen Y, Pfeifer GP (2009) Methylation of homeobox genes is a frequent and early epigenetic event in breast cancer. *Breast Cancer Research* **11**(1):R14

Tong W-G, Wierda WG, Lin E, Kuang S-Q, Bekele BN, Estrov Z, Wei Y, Yang H, Keating MJ, Garcia-Manero G (2010). Genome-wide DNA methylation profiling of chronic lymphocytic leukaemia allows identification of epigenetically represses molecular pathways with clinical impact. *Epigenetics* **5**(6): 499-508.

Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP (1999a) CpG island methylator phenotype in colorectal cancer. *PNAS* **96**(15):8681-8686.

Toyota M, Ahuja N, Suzuki H, Itoh F, Ohe-Toyota M, Imai K, Baylin SB, Issa JP (1999b) Aberrant methylation in gastric cancer associated with the CpG island methylator phenotype. *Cancer Research* **59**(21):5438-5442.

Toyota M, Ohe-Toyota M, Ahuja N, Issa JP (2000) Distinct genetic profiles in colorectal tumors with or without the CpG island methylator phenotype. *PNAS* **97**(2):710-705.

Tozlu-Kara S, Roux V, Andrieu C, Vendrell J, Vacher S, Lazar V, Spyratos F, Tubiana-Hulin, Cohen P, Dessen P, Lidereau R and Bièche I (2007) Oligonucleotide microarray analysis of estrogen receptor α-positive postmenopausal breast carcinomas: identification of *HRPAP20* and *TIMELESS* as outstanding candidate markers to predict the response to tamoxifen. *Journal of Molecular Endocrinology **39**:305-318.*

Tsunoda S, Smith E, De Young NJ, Wang X, Tian ZQ, Liu JF, Jamieson GG, Drew PA (2009) Methylation of CLDN6, FBN2, RBP1, RBP4, TFPI2, and TMEFF2 in esophageal squamous cell carcinoma. *Oncol. Rep.* **Apr 21**(4):1067-73

Turnbull C and Rahman N (2008) Genetic Predisposition to Breast Cancer: Past, Present and Future. *Annual Review of Genomics and Human Genetics* **9**:321-345.

Uchida H, Zhang J, Nimer SD (1997) AML1A and AML1B can transactivate the human IL-3 promoter. *Journal of Immunology* **158**(5):2251-2258.

Uckun FM, Sensel MG, Sather HN, Gaynon PS, Arthur DC, Lange BJ, Steinherz PG, Kraft P, Hutchinson R, Nachman JB, Reaman GH, Heerema NA (1998) Clinical significance of translocation t(1;19) in childhood acute lymphoblastic leukemia in the context of contemporary therapies: a report from the Children's Cancer Group. *Journal of Clinical Oncology* **16**(2):527-535.

Uhlman K, Brinckmann A, Toliat MR, Ritter H, Nürnberg P (2002) Evaluation of a potential epigenetic biomarker by quantitative methyl-single nucleotide polymorphism analysis. *Electrophoresis* **23**:4072-4079.

Underhill-Day N, Hill V and Latif F (2011) N-terminal RASSF family (RASSF7-RASSF10) A mini review. *Epigenetics* **6**(3):284-292.

Urakami S, Shiina H, Enokida H, Hirata H, Kawamoto K, Kawakami T, Kikuno N, Tanaka Y, Majid S, Nakagawa M, Igawa M, Dahiya R (2006) Wnt antagonist family genes as biomarkers for diagnosis, staging, and prognosis of renal cell carcinoma using tumor and serum DNA. *Cinical Cancer Research* **12**(23):6989-6997.

Vanaja DK, Ehrich M, Van den Boom D, Cheville JC, Karnes RJ, Tindall DJ, Cantor CR, Young CY (2000) Hypermethylation of genes for diagnosis and risk stratification of prostate cancer. *Cancer Investigation* **27**(5):549-560.

Van der Auwera I, Yu W, Suo L, Van Neste L, van Dam P, Van Marck EA, Pauwels P, Vermeulen PB, Dirix LY and Van Laere SJ (2010) Array-Based DNA Methylation Profiling for Breast Cancer Subtype Discrimination. *PloS One* **5**(9):e12616

van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R (2002) A gene-expression signature as a predictor of survival in breast cancer. *New England Journal of Medicine* **347**(25):1999-2009.

van Engerland M, Roeman GM, Brink M, Pachen MM, Weijenberg MP, de Bruïne AP, Arends JW, van den Bradnt PA, de Goeij AF, Herman JG (2002) K-ras mutations and RASSF1A promoter methylation in colorectal cancer. *Oncogene* **21**(23):3792-3795.

van Haaften G, Dalgliesh GL, Davies H, Chen L, Bignell G, Greenman C, Edkins S, Hardy C, O'Meara S, Teague J, Butler A, Hinton J, Latimer C, Andrews J, Barthorpe S, Beare D, Buck G, Campbell PJ, Cole J, Forbes S, Jia M, Jones D, Kok CY, Leroy C, Lin ML, McBride DJ, Maddison M, Maquire S, McLay K, Menzies A, Mironenko T, Mulderrig L, Mudie L, Pleasance E, Shepherd R, Smith R, Stebbings L, Stephens P, Tang G, Tarpey PS, Turner R, Turrell K, Varian J, West S, Widaa S, Wray P, Collins VP, Ichimura K, Law S, Wong J, Yuen ST, Leung SY, Tonon G, DePinho RA, Tai YT, Anderson KC, Kahnoski RJ, Massie A, Khoo SK, Teh BT, Stratton MR, Futreal PA (2009) Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. *Nature Genetics* **41**(5):521-523.

van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**(6871):530-536.

Varela I, Tarpey P, Raine K, Huang D, Ong CK, Stephens P, Davies H, Jones D, Lin ML, Teague J, Bignell G, Butler A, Cho J, Dalgliesh GL, Galappaththige D, Greenman C, Hardy C, Jia M, Latimer C, Lau KW, Marshall J, McLaren S, Menzies A, Mudie L, Stebbings L, Largaespada DA, Wessels LF, Richard S, Kahnoski RJ, Anema J, Tuveson DA, Perez-Mancera PA, Mustonen V, Fischer A, Adams DJ, Rust A, Chan-on W, Subimerb C, Dykema K, Furge K, Campbell PJ, Teh BT, Stratton MR, Futreal PA (2011) Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature* **469**(7331):539-542.

Veeck J, Geisler C, Noetzel E, Alkaya S, Hartmann A, Knüchel R, Dahl E (2008a) Epigenetic inactivation of the secreted frizzled-related protein-5 (SFRP5) gene in human breast cancer is associated with unfavorable prognosis. *Carcinogenesis* **May 25**(5):991-8

Veeck J, Noetzel E, Bektas N, Jost E, Hartmann A, Knüchel R and Dahl E (2008b) Promoter hypermethylation of the *SFRP2* gene is a high-frequent alteration and tumour-specific epigenetic marker in human breast cancer. *Molecular Cancer 7(83)*

Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, Alexe G, Lawrence M, O'Kelly M, Tamayo P, Weir BA, Gabriel S, Winckler W, Gupta S, Jakkula L, Feiler HS, Hodgson JG, James CD, Sarkaria JN, Brennan C, Kahn A, Spellman PT, Wilson RK, Speed TP, Gray JW, Meyerson M, Getz G, Perou CM, Hayes DN; Cancer Genome Atlas Research Network (2010) Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR and NF1. *Cancer Cell* **17:**98-110.

Virmani AK, Rathi A, Sathyanarayana UG, Padar A, Huang CX, Cunnigham HT, Farinas AJ, Milchgrub S, Euhus DM, Gilcrease M, Herman J, Minna JD, Gazdar AF (2001) Aberrant methylation of the adenomatous polyposis coli (APC) gene promoter 1A in breast and lung carcinomas. *Clinical Cancer Research* **Jul 7**(7):1998-2004

Vos MD, Ellis CA, Bell A, Birrer MJ, Clark GJ (2000) Ras uses the novel tumor suppressor RASSF1 as an effector to mediate apoptosis. *Journal of Biological Chemistry* **275**(46):35669-35672.

Vos MD, Martinez A, Ellis CA, Vallecorsa T, Clark GJ (2003a) The pro-apoptotic Ras effector Nore1 may serve as a Ras-regulated tumor suppressor in the lung. *Journal of Biological Chemistry* **278**(24):21938-21943

Vos MD, Ellis CA, Elam C, Ulku AS, Taylor BJ, Clark GJ (2003b) RASSF2 is a novel K-Ras-specific effector and potential tumor suppressor. *Journal of Biological Chemistry* **278**(30):28045-28051.

Vos MD, Dallol A, Eckfeld K, Allen NP, Donninger H, Hesson LB, Calvisi D, Latif F, Clark GJ (2006) The RASSF1A tumor suppressor activates Bax via MOAP-1. *Journal of Biological Chemistry* **281**(8):4557-4563.

Vousden KH, Lane DP (2007). p53 in health and disease. *Nature Reviews Molecular Cell Biology* **8**(4):275-83.

Walter MJ, Ding L, Shen D, Shao J, Grillot M, McLellan M, Fulton R, Schmidt H, Kalicki-Veizer J, O'Laughlin M, Kandoth C, Baty J, Westervelt P, Dipersio JF, Mardis ER, Wilson RK, Ley TJ, Graubert TA (2011) Recurrent DNMT3A mutations in patients with myelodysplastic syndromes. *Leukemia* March 18: Epub. ahead of print

Wang WW, Spurdle AB, Kolachana P, Bove B, Modan B, Ebbers SM, Suthers G, Tucker MA, Kaufman DJ, Doody MM, Tarone RE, Daly M, Levavi H, Pierce H, Chetrit A, Yechezkel GH, Chenevix-Trench G, Offit K, Godwin AK, Struewing JP (2001) A single nucleotide polymorphism in the 5' untranslated region of RAD51 and risk of cancer among BRCA1/2 mutation carriers. *Cancer epidemiology, biomarkers and prevention* **10**(9):955-960.

Wang Y, Leung FC (2004) An evaluation of new criteria for CpG islands in the human genome as gene markers. *Bioinformatics* **20**: 1170–1177.

Wang JS, Guo M, Montgomery EA, Thompson RE, Cosby H, Hicks L, Wang S, Herman JG, Canto MI (2009) DNA promoter hypermethylation of p16 and APC predicts neoplastic progression in Barrett's esophagus. *Amercian Journal of Gastroenterology* **104**(9):2153-2160.

Watanabe K, Tachibana O, Sata K, Yonekawa Y, Kleihues P, Ohgaki H (1996) Overexpression of the EGF receptor and p53 mutations are mutually exclusive in the evolution of primary and secondary glioblastomas *Brain Pathology* **6**(3):217-223.

Watt HL, Kharmate G and Kumar U (2008) Biology of somatostatin in breast cancer. *Molecular and Cellular Endocrinology* **286**:251-261.

Wazlinski A, Engers R, Hoffmann MJ, Hader C, Jung V, Müller M, Schulz WA (2007) Downregulation of Several Fibulin Genes in Prostate Cancer. *The Prostate* **67**:1770-1780.

Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schübeler D (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genetics* **37**(8):853-62.

Wei X, Walia V, Lin JC, Teer JK, Prickett TD, Gartner J, Davis S; NISC Comparative Sequencing Program, Stemke-Hale K, Davies MA, Gershenwald JE, Robinson W, Robinson S, Rosenberg SA, Samuels Y (2011) Exome sequencing identifies GRIN2A as frequently mutated in melanoma. *Nature Genetics* **43**(5):442-446.

Weinberg RA (2007) The Biology of Cancer. *Garland Scince Publishing.*

Wiegand KC, Shah SP, Al-Agha OM, Zhao Y, Tse K, Zeng T, Senz J, McConechy MK, Anglesio MS, Kalloger SE, Yang W, Heravi-Moussavi A, Giuliany R, Chow C, Fee J, Zayed A, Prentice L, Melnyk N, Turashvili G, Delaney AD, Madore J, Yip S, McPherson AW, Ha G, Bell L, Fereday S, Tam A, Galletta L, Tonin PN, Provencher D, Miller D, Jones SJ, Moore RA, Morin GB, Oloumi A, Boyd N, Aparicio SA, Shih IeM, Mes-Masson AM, Bowtell DD, Hirst M, Gilks B, Marra MA, Huntsman DG (2010) ARID1A mutations in endometriosis-associated ovarian carcinomas. *New England Journal of Medicine* **363**(16):1532-1543.

Wood LD, Parsons DW, Jones S, Lin J, Sjöblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Willson JK, Sukumar S, Polyak K, Park BH, Pethiyagoda CL, Pant PV, Ballinger DG, Sparks AB, Hartigan J, Smith DR, Suh E, Papadopoulos N, Buckhaults P, Markowitz SD, Parmigiani G, Kinzler KW, Velculescu VE, Vogelstein B (2007) The genomic landscapes of human breast and colorectal cancers. *Science* **318**(5853):1108-1113.

Wolff EM, Chihara Y, Pan F, Weisenberger DJ, Siegmund KD, Sugano K, Kawashima K, Laird PW, Jones PA, Liang G (2010) Unique DNA methylation patterns distinguish noninvasive and invasive urothelial cancers and establish an epigenetic field defect in premalignant tissue. *Cancer Research* **70**(20):8169-8178.

Wolin KY, Carson K, Colditz GA (2010) Obesity and cancer. *Oncologist* **15**(6):556-565.

Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G (1995) Identification of the breast cancer susceptibility gene BRCA2 *Nature.* **378**(6559):789-92. Erratum in: *Nature* 1996 **379**(6567):749.

Wu S, Liu Y, Zheng Y, Dong J, Pan D (2008) The TEAD/TEF family protein Scalloped mediates transcriptional output of the Hippo growth-regulatory pathway. *Developmental Cell* **14**(3):388-398.

Wu S, Huang J, Dong J, Pan D (2003) hippo encodes a Ste-20 family protein kinase that restricts cell proliferation and promotes apoptosis in conjunction with salvador and warts. *Cell* **114**(4):445-456.

Wu X, Rauch TA, Zhong X, Bennett WP, Latif F, Krex D and Pfeifer GP (2010) CpG island hypermethylation in human astrocytomas. *Cancer Research* **70**(7):2718-2727

Xong Z and Laird P (1997) COBRA: a sensitive and quantitative DNA methylation assay. *Nucelic Acids Research* **25**(12):2532-2534.

Xu B, Kim St, Kastan MB (2001) Involvement of Brca1 in S-phase and G(2)-phase checkpoints after ionizing irradiation. *Molecular and Cellular Biology* **21**(10):3445-3450.

Yabuta N, Fujii T, Copeland NG, Gilbert DJ, Jenkins NA, Nishiguchi H, Endo Y, Toji S, Tanaka H, Nishimune Y, Nojima H (2000) Structure, expression, and chromosome

mapping of LATS2, a mammalian homologue of the Drosophila tumor suppressor gene lats/warts. *Genomics* **63**(2):263-270.

Yamamoto C, Fukuda N, Matsumoto T, Higuchi T, Ueno T, Matsumoto K (2010) Zinc-finger transcriptional factor Sall1 induces angiogenesis by activation of the gene for VEGF-A. *Hypertension Research* **33**(2):143-148.

Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, Kos I, Batinic-Haberle I, Jones S, Riggins GJ, Friedman H, Friedman A, Reardon D, Herndon J, Kinzler KW, Velculescu VE, Vogelstein B, Bigner DD (2009) IDH1 and IDH2 mutations in gliomas. *New England Journal of Medicine* **360**:765-773.

Yan XJ, Xu J, Gu ZH, Pan CM, Lu G, Shen Y, Shi JY, Zhu YM, Tang L, Zhang XW, Liang WX, Mi JQ, Song HD, Li KQ, Chen Z, Chen SJ (2011) Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nature Genetics* 3 March: Epub. ahead of print

Yanagitani A, Yamada S, Yasui S, Shimomura T, Murai R, Murawaki Y, Hashiguchi K, Kanbe T, Saeki T, Ichiba M, Tanabe Y, Yoshida Y, Morino S, Kurimasa A, Usuda N, Yamazaki H, Kunisada T, Ito H, Murawaki Y, Shiota G (2004) Retinoic acid receptor alpha dominant negative form causes steatohepatitis and liver tumors in transgenic mice. *Hepatology* **40**(2):366-375.

Yi CH, Smith DJ, West WW, Hollingsworth MA (2007) Loss of fibulin-2 expression is associated with breast cancer progression. *American Journal of Pathology* **170**(5):1535-45.

Yoga K, Pelizzola M, Cheng E, Krauthammer M, Sznol M, Ariyan S, Narayan D, Molinaro AM, Halaban R and Weissman SM (2009) Genome-wide screen of promoter methylation identifies novel markers in melanoma. *Genome Research* **19**:1462-1470.

Yonezawa T, Haga S, Kobayashi Y, Katoh K, Obara Y (2009) Saturated fatty acids stimulate and insulin suppresses CIDE-A expression in bovine mammary epithelial cells. *Biochemical and Biophysical Research Communication* **384**:535-539.

Yoon HG, Chan DW, Reynolds AB, Qin J, Wong J (2003) N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. *Molecular Cell* **12**:723-734.

Yu J, Zheng Y, Dong J, Klusza S, Deng W-M and Pan D (2010) Kibra Functions as a Tumour Suppressor Protein that regulates Hippo Signaling in Conjunction with Merlin and Expanded. *Developmental Cell* **18**:288-299

Yuan M, Tomlinson V, Lara R, Holliday D, Chelala C, Harada T, Gangeswaran R, Manson-Bishop C, Smith P, Danovi SA, Pardo O, Crook T, Mein CA, Lemoine NR, Jones LJ, Basu S (2008) Yes-associated protein (YAP) functions as a tumor suppressor in breast. *Cell Death and Differentiation* **15**(11):1752-1759.

Zhang F, Fan Q, Ren K, Andreassen PR (2009a) PALB2 functionally connects the breast cancer susceptibility proteins BRCA1 and BRCA2. *Molecular Cancer Research* **7**(7):1110-1118.

Zhang F, Ma J, Wu J, Ye L, Cai H, Xia B, Yu X (2009b) PALB2 links BRCA1 and BRCA2 in the DNA-damage response. *Current Biology* **19**(6):524-529.

Zhang Z, Sun D, Van do N, Tang A, Hu L, G Huang (2007) Inactivation of RASSF2A by promoter methylation correlates with lymph node metastasis in nasopharyngeal carcinoma. *International Journal of Cancer* **120**:32-38

Zhou Z, Yon Toh S, Chen Z, Guo K, Ng CP, Ponniah S, Lin SC, Hong W, Li P (2003) Cidea-deficient mice have lean phenotype and are resistant to obesity. *Nature Genetics* **35**:49-56.